CONFERENCE OF PHD STUDENTS IN COMPUTER SCIENCE

Volume of extended abstracts

CS^2

Organized by the Institute of Informatics of the University of Szeged

June 27–30, 2006
Szeged, Hungary

Scientific Committee:

Mátyás Arató (KLTE)
Miklós Bartha (SZTE)
András Benczúr (ELTE)
Tibor Csendes (SZTE)
János Csirik (SZTE)
János Demetrovics (SZTAKI)
Sarolta Dibuz (Ericsson)
József Dombi (SZTE)
Zoltán Ésik (SZTE)
Ferenc Friedler (VE)
Zoltán Fülöp (SZTE)
Ferenc Gécseg (chair, SZTE)
Tibor Gyimóthy (SZTE)
Balázs Imreh (SZTE)
János Kormos (KLTE)
László Kozma (ELTE)
Attila Kuba (SZTE)
Eörs Máté (SZTE)
Gyula Pap (KLTE)
András Recski (BME)
Endre Selényi (BME)
Katalin Tarnay (NOKIA)
György Turán (SZTE)
László Varga (ELTE)

Organizing Committee:

Tibor Csendes, Péter Gábor Szabó, Mariann Sebő, Balázs Bánhelyi, Judit Jász, and Gabriella Nagyné Hecskó

Address of the Organizing Committee

c/o. Tibor Csendes
University of Szeged, Institute of Informatics
H-6701 Szeged, P.O. Box 652, Hungary
Phone: +36 62 544 305, Fax: +36 62 546 397
E-mail: `cscs@inf.u-szeged.hu`
URL: `http://www.inf.u-szeged.hu/~cscs/`

Main sponsor
SIEMENS Sysdata

Sponsors

City Major's Office, Szeged, Novadat Bt., Polygon Publisher, the Szeged Region Committee of the Hungarian Academy of Sciences, University of Szeged, Institute of Informatics.

Preface

This conference is the fifth in a series. The organizers have tried to get together those PhD students who work on any fields of computer science and its applications to help them possibly in writing their first abstract and paper, and may be to give their first scientific talk. As far as we know, this is one of the few such conferences. The aims of the scientific meeting were determined on the council meeting of the Hungarian PhD Schools in Informatics: it should

- provide a forum for PhD students in computer science to discuss their ideas and research results,

- give a possibility to have constructive criticism before they present the results in professional conferences,

- promote the publication of their results in the form of fully refereed journal articles, and finally

- promote hopefully fruitful research collaboration among the participants.

The best talks will be awarded with the help of our sponsors. The papers emerging from the presented talks will be forwarded to the journals of Acta Cybernetica (Szeged), and Periodica Polytechnica (Budapest); and the mathematics oriented papers to Publicationes Mathematicae (Debrecen). The deadline for the submission of the papers is the end of August 2006. The manuscripts will be forwarded to the proper journals. To get acquainted with the style of the journals please study earlier issues of them. One sample paper is available at http://www.inf.u-szeged.hu/~cscs/csallner.tex.

Although we did not advertise it on the web, a high number of good quality abstracts have been submitted. If you encounter any problems during the meeting, please do not hesitate to contact one of the Organizing Committee members. The organizers hope that the conference will be a valuable contribution to the research of the participants, and wish a pleasant stay in Szeged.

Szeged, June 2006

Tibor Csendes

Contents

Preliminary Program

Overview

Tuesday, June 27

- 8:00 - 10:00 Registration
- 10:00 - 10:15 Opening
- 10:15 - 11:00 Plenary talk
- 11:00 - 11:15 Break
- 11:15 - 12:45 Talks in 2 streams (3x30 minutes)
- 12:45 - 14:00 Lunch
- 14:00 - 15:30 Talks in 2 streams (3x30 minutes)
- 15:30 - 15:45 Break
- 15:45 - 17:45 Talks in 2 streams (4x30 minutes)
- 18:15   Reception at the Town Hall

Wednesday, June 28

- 08:30 - 10:00 Talks in 2 streams (3x30 minutes)
- 10:00 - 10:15 Break
- 10:15 - 11:00 Plenary talk
- 11:00 - 11:15 Break
- 11:15 - 12:45 Talks in 2 streams (3x30 minutes)
- 12:45 - 14:00 Lunch
- 14:00 - 16:00 Talks in 2 streams (4x30 minutes)
- 16:00 - 16:15 Break
- 16:15 - 18:00 Munkácsy exhibition
- 18:15   Supper

Thursday, June 29

- 08:30 - 10:00 Talks in 2 streams (3x30 minutes)
- 10:00 - 10:15 Break
- 10:15 - 11:00 Plenary talk
- 11:00 - 11:15 Break
- 11:15 - 12:45 Talks in 2 streams (3x30 minutes)
- 12:45 - 14:00 Lunch
- 15:00　　　　Excursion, supper

Friday, June 30

- 08:30 - 10:00 Talks in 2 streams (3x30 minutes)
- 10:00 - 10:15 Break
- 10:15 - 11:00 Plenary talk
- 11:00 - 11:15 Break
- 11:15 - 12:45 Talks in 2 streams (3x30 minutes)
- 12:45 - 14:00 Lunch
- 14:00 - 14:30 Talk in 1 stream (1x30 minutes)
- 14:30　　　　Closing session, awards

Detailed program

Tuesday, June 27

8:00	Registration	
10:00	Opening session	
10:15	**Plenary talk** Tibor Gyimóthy (Szeged, Hungary): *Program slicing: theory and practice*	
11:00	Break	
Sections	**Fuzzy and Control**	**Model-Based Software Development**
11:15	Zsolt Gera: *Extensions and Applicability of the Membership Driven Reasoning Scheme*	László Lengyel, Tihamér Levendovszky, and Hassan Charaf: *Aspect-UML-Driven Model-Based Software Development*
11:45	József Dombi and Norbert Győrbíró: *Pliant Ranking*	Gergely Mezei, Tihamér Levendovszky, and Hassan Charaf: *Optimization algorithms for constraint handling*
12:15	Zoltán Dávid and István Vajk: *Model-Free Control Based on Reinforcement Learning*	Dániel Tóth and Eszter Jósvai: *Model-based Development of Graphical User Interfaces*
12:45	Lunch	
Sections	**Bin Packing**	**Image Processing**
14:00	János Balogh, József Békési, Gábor Galambos, and Gerhard Reinelt: *On-line Bin Packing with Restricted Repacking: Lower Bounds*	Péter Balázs: *On the Ambiguity of Reconstructing Decomposable hv-Convex Binary Matrices*
14:30	Máté Labádi: *A linear programming background for the R^{∞}_{HFF} upper bound proof*	Krisztina Dombi: *Calibration of CCD cameras computer aided surgery*
15:00	János Balogh, József Békési, Gábor Galambos, and Gerhard Reinelt: *One dimensional semi-on-line bin packing algorithms*	Dénes Paczolay: *2D Pattern Repetition Based Lossless Image Compression*

Tuesday, June 27 (continued)

15:30	Break	
Sections	**Speech Recognition**	**Grid Systems**
15:45	András Bánhalmi, Dénes Paczolay, and András Kocsor: *An On-line Speaker Adaptation Method for HMM-based ASRs*	Gábor Gombás: *Modeling of grid monitoring systems*
16:15	Dénes Paczolay, András Bánhalmi, and András Kocsor: *Robust Recognition of Vowels in Speech Impediment Therapy Systems*	Attila Kertész and Péter Kacsuk: *Grid Meta-Broker Architecture: Requirements of an Interoperable Grid Resource Brokering Service*
16:45	György Szaszák and Zsolt Németh: *Word Boundary Detection Based on Phoneme Sequence Constraints*	József Kovács: *Formal analysis of existing checkpointing systems and introduction of a novel approach*
17:15	Csaba Teleki, Szabolcs L. Tóth, and Klára Vicsi: *Testing and optimization of a continuous speech recognizer with a middle sized vocabulary*	Ágnes Lipovits, Előd Kovács, and Zoltán Juhász: *Fitting the statistical module of the adaptive grid scheduler to the data of the NIIF*
18:15	Reception at the Town Hall	

Wednesday, June 28

Sections	Numerical methods	Programming Languages
08:30	Jose A. Egea: *An Optimization Approach for Integrated Design and Parameter Estimation in Process Engineering*	Ádám Balogh and Zoltán Csörnyei: *New Method for Designing Polymorphic System Programming Languages*
09:00	Gábor Horváth: *Design and reengineer metadata with the help of business process modelling*	László Lövei, Máté Tejfel, Mónika Mészáros, Zoltán Horváth, and Tamás Kozsik: *Comparing Specification with Proved Properties of Clean Dynamics*
09:30	Ferenc Domes: *Global optimization: rigorous solutions of constraint satisfaction problems*	Zoltán Horváth, László Lövei, Tamás Kozsik, Anikó Víg, and Tamás Nagy: *Refactoring Erlang Programs*
10:00	Break	
10:15	**Plenary talk** Hermann Schichl (Vienna, Austria): *Global Optimization, Constraint Satisfaction, and the COCONUT Environment*	
11:00	Break	

Sections	File Systems, Documents	Software Engineering
11:15	Róbert Dávid and Gábor Vincze: *Improving DHT-Based File System Performance with B-trees*	Norbert Pataki, Ádám Sipos, and Zoltán Porkoláb: *Structural Complexity Metrics on SDL Programs*
11:45	Kristóf Csorba and István Vajk: *Probabilistic confidence prediction in document clustering*	Ákos Dávid, Tamás Pozsgai, and László Kozma: *Extending a system with verified components*
12:15	Hajnalka Hegedűs and László Csaba Lőrincz: *Towards A Unified Stylesheet Format for Paper-Based and Electronic Documents*	Gergely Dévai: *Programming Language Elements for Correctness Proofs*
12:45	Lunch	

Wednesday, June 28 (continued)

Sections	Model Transformations	Web Applications
14:00	Ákos Horváth: *Automatic Generation of Compiled Model Transformations*	Ágnes Bogárdi-Mészöly, Tihamér Levendovszky, and Hassan Charaf: *Models for Predicting the Performance of ASP.NET Web Applicaitons*
14:30	Máté Kovács: *Simulation and Formal Analysis of Workflow Models Using Model Transformations*	Péter Cserkúti and Hassan Charaf: *A Rule-based Transformation Engine for Web Page Re-authoring*
15:00	Dávid Vágó: *Transformation and simulation of domain specific languages*	Ágnes Bogárdi-Mészöly, Tihamér Levendovszky, and Hassan Charaf: *Methods for Retrieving and Investigating Performance Factors in ASP.NET Web Applications*
15:30	László Vidács: *Model transformations on the Preprocessor Metamodel - Graph Transformation approach*	Gábor Imre and Hassan Charaf: *A Novel Performance Model of J2EE Web Applications Considering Application Server Settings*
16:00	Break	
16:15	Munkácsy exhibiton	
18:15	Supper	

Thursday, June 29

Sections	Databases	Artificial Intelligence
08:30	Róbert Busa-Fekete, Kornél Kovács, and András Kocsor: *Extracting Human Protein Information from MEDLINE Using a Full-Sentence Parser*	Kristóf Aczél: *Instrument separation in poliphonic recordings using instrument prints*
09:00	István Szépkúti: *Caching in Multidimensional Databases*	Richárd Szabó: *Occupancy Grid Based Robot Navigation with Sonar and Camera*
09:30	Ovidiu Vasutiu and Florina Vasutiu: *Database Design Patterns*	Krisztina Tóth, Richárd Farkas, and András Kocsor: *Hybrid algorithm for sentence alignment of Hungarian-English parallel corpora*
10:00	Break	
10:15	**Plenary talk** Marius Minea (Timisoara, Roumania): *Assume-guarantee compositional reasoning*	
11:00	Break	

Sections	Automata, models, temporal logic	Program Analysis, C++
11:15	Loránd Muzamel: *The Power of Deterministic Alternating Tree-Walking Automata*	Szabolcs Payrits and István Zólyomi: *Concept-based C++ template overload compilation with XML transformations*
11:45	Gáspár Erdélyi: *Towards a unified model of computation with correctness proofs*	Endre Horváth and Judit Jász: *Comparison of static-and dynamic call graphs*
12:15	Mónika Mészáros: *Proving Quality of Service Constraints of Multimedia Systems*	Mihály Biczó, Krisztián Pócza, and Zoltán Porkoláb: *Test Generation with Dynamic Impact Analysis for C++ Programs*
12:45	Lunch	
15:00	Excursion and supper	

Friday, June 30

Sections	Software Environments, MDA, DSM, SOA	Telecommunication, TTCN-3
08:30	István Ráth: *Declarative mapping between concrete and abstract syntax of domain-specific visual languages*	Máté J. Csorba and Sándor Palugyai: *A Performance Model for Load Test Components*
09:00	László Gönczy: *Verification of Reconfiguration Mechanisms in Service-Oriented*	Sándor Palugyai and János Miskolczi: *Performance Modeling of Hierarchical List Structures*
09:30	Gábor Bátori, Zoltán Theisz, and Domokos Asztalos: *Model driven testing of component based systems*	Antal Wu-Hen Chang, Dung Le Viet, and Gyula Csopaki: *High-level Restructuring of TTCN-3 Test Suites*
10:00	Break	
10:15	**Plenary talk** Lajos Rónyai (Budapest, Hungary): *Finding sense in large sets of data - some powerful applications* *of Math and CS* (with András Benczúr jr.)	
11:00	Break	
Sections	**Optimization and Interpolation**	**Networks, Frameworks**
11:15	Balázs Bánhelyi: *Investigation of a Delayed Differential Equation with Verified Computing Technique*	Pedone Gianfranco: *An Agent-based Framework for Supporing e-Health*
11:45	Miklós Kozlovszky and T. Berceli: *Optical delay buffer optimization in packet switched optical network*	Vilmos Bilicki: *Network topology discovery*
12:15	Miklós Póth: *Comparison of convolution based interpolation techniques in digital image processing*	Rita Móga, Tamás Polyák, and István Oláh: *DRM systems in wireless environment*
12:45	Lunch	

Friday, June 30 (continued)

Sections	Algorithms	Distributed Systems, Metadata Exchanges
14:00	Szabolcs Faragó: *Implementing and comparing impact algorithms*	Vilmos Bilicki and József Dombi Dániel: *Paxos with multiple leaders*
14:30	Julia Marchis: *Simulations on a fractal growth model*	
15:00	Closing session, announcing the Best Talk Awards	

Instrument separation in poliphonic recordings using instrument prints

Kristóf Aczél

Decomposing a polyphonic musical recording to separate voice and instrument tracks has always been a challenge. Even extracting one of the many instruments in a complex recording is currently unsolved. The importance of this issue shows when we want to apply different filters to single instruments in a recording that have already been mixed into stereo channels. The paper shows a new way of sound separation that works even on mono-aural digital recordings.

There are already achievements on sound source separation. Although these algorithms may do a good job in separating instruments or groups of instruments from each other, they are not capable of separating one single note from the remaining part of the recording. It means that they are e.g. unable to filter out one and only one single misplayed note from a polyphonic piece.

The paper deals with one approach for separating single notes in the recording. The goal is using this approach later for doing pitch shifting on the single separated notes in a polyphonic recording when needed. Therefore we will concentrate on separating only those parts that determine the base frequency of a musical note. This means that in case of e.g. a piano recording we do not deal with the sound of the hammer hitting the strings - which remains more or less the same regardless of which key we press - but only with the sound of the strings which vary as we play different notes on the instrument.

The separation algorithm is based on the discreet Fourier transform which is used for converting the sound data from time domain to frequency domain. We compute the spectogram of two recordings. One is the original recording, which has to be processed. The second one is a recording that contains sample notes of the instrument of interest.
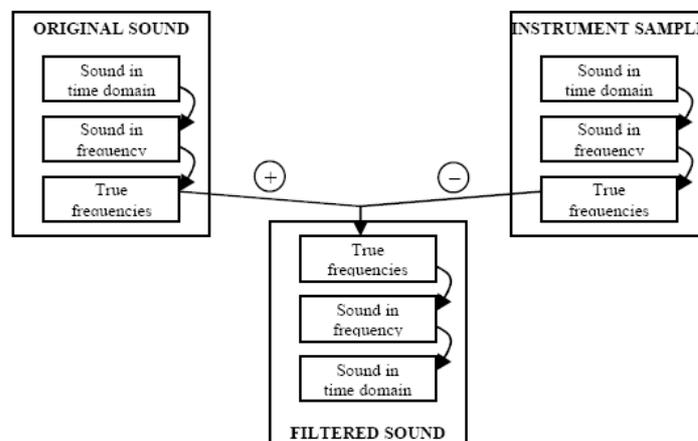
Figure 1: Block diagram of sound separation

We can generate an instrument model of the specific instrument from the second recording, storing its characteristics. We will call this model "instrument print" model. After storing the right instrument print, we can analyse the first recording. We find the spot where we want to separate one or more notes from the remainder of the recording, then subtract the right frequencies from the recording in frequency domain, with the help of the instrument print. The subtracted component can be kept if needed, or thrown away, if not. After this step, the sound data can be converted back to time-domain.

On the Ambiguity of Reconstructing Decomposable hv-Convex Binary Matrices

Péter Balázs

The reconstruction of binary matrices from their projections is a basic problem in discrete tomography. Reconstruction algorithms have a wide area of applications in graph theory, non-destructive testing, statistical data security, biplane angiography, crystallography, radiology, image processing, and so on. For practical reasons the projections can be taken only from few (usually at most four) directions. This often leads to ambiguous or/and NP-hard reconstruction which is inappropriate for applications. One commonly used technique to reduce ambiguity and to avoid intractability is to suppose having some a priori information of the matrix to be reconstructed. In this paper we assume that the matrix to be reconstructed is hv-convex and decomposable. First, we give a construction to prove that the use of only two projections is not sufficient to eliminate ambiguity, that is, for some inputs there can be exponentially large number of hv-convex decomposable binary matrices having the same horizontal and vertical projections. In the case of four projections we are faced the following problem. Although using four projections all the binary matrices of this class with the given projections can be reconstructed in polynomial time [1] the class of decomposable hv-convex matrices is not explicitly defined. In more detail, one criteria of decomposability is that the components of the binary matrix are uniquely reconstructible from their horizontal and vertical projections. However, when reconstructing hv-convex matrices it cannot be decided in advance whether this criteria is satisfied. This sometimes does not lead the reconstruction algorithm to fail, i.e., the reconstruction algorithm possibly gives a solution even if one of the components is not uniquely determined. We do experiments to investigate the possibility that a component (which is an hv-convex polyomino) is uniquely determined by its horizontal and vertical projections. We also study how the knowledge of a component's third projection effects on the ambiguity.

References

[1] P. Balázs. A decomposition technique for reconstructing discrete sets from four projections, *Image and Vision Computing*, submitted.

New Method for Designing Polymorphic System Programming Languages

Ádám Balogh and Zoltán Csörnyei

There are a lot of high level programming languages for application programming which make use of advanced techniques like classes, inheritance, polymorphism and virtual methods. However, the use of these languages for low level programming, like operating system or embedded system software development is quite limited because of their high overhead (e.g. explicit data tags of identifying dynamic type, virtual method tables etc.). This is probably the main cause that software for these systems are developed still in C, which neither does have the above mentioned techniques, nor is type safe. In our presentain we show language constructs which enable the system programmer to use many features of the object-oriented languages without significant extra overhead. Our solution is based on type invariants which identify the dynamic type of a variable instead of explicit data tags. Virtual method tables are replaced by small dispatcher routines generated at link time. The technique discussed in the presentation can be used either to extend already existing languages or to design a new language.

References

[1] K. B. Bruce, A. Schuett, R. van Gent, and A. Fiech. PolyTOIL: A type-safe polymorphic object-oriented language, *ACM Trans. Program. Lang. Syst.*, 25, 2, 2003, 0164-0925, 225–290, `http://doi.acm.org/10.1145/641888.641891`, ACM Press, New York, NY, USA.

[2] J. C. B. Mattos, E. Specht, B. Neves, and Luigi Carro. Making object oriented efficient for embedded system applications, *SBCCI '05: Proceedings of the 18th annual symposium on Integrated circuits and system design*, 2005, 1-59593-174-0, 104–109, Florianolpolis, Brazil, `http://doi.acm.org/10.1145/1081081.1081111`, ACM Press, New York, NY, USA.

[3] A. Chatzigeorgiou and G. Stephanides. Evaluating Performance and Power of Object-Oriented Vs. Procedural Programming in Embedded Processors, *Reliable Software Technologies - Ada-Europe 2002, LNCS2361*, 2002, 0302-9743, 65–75, J. Blieberger and A. Strohmeier, Springer-Verlag, Berlin/Heidelberg.

[4] Y. Harada, K. Yamazaki and R. Potter. CCC: User-Defined Object Structure in C, *ECOOP 2001 - Object-Oriented Programming: 15th European Conference, Budapest, Hungary, LNCS2072*, 2001, 0302-9743, 118-129, J. Lindskov Knudsen, Springer-Verlag, Berlin/Heidelberg.

On-line Bin Packing with Restricted Repacking: Lower Bounds[1]

János Balogh, József Békési, Gábor Galambos, and Gerhard Reinelt

In 1996 Ivkovič and Lloyd [1] gave a lower bound of $\frac{4}{3}$ for those one-dimensional fully dynamic bin packing algorithms, where the number of repackable items is restricted to a constant in each step. In this paper we improve this result to about 1.3871. We present our proof for a semi-on-line case of the classical bin packing, but it works for the same case of fully dynamic bin packing as well. We prove the lower bound by analyzing and solving a specific optimization problem with non-linear constraints. The bound can be expressed exactly using the Lambert W function.

References

[1] Z. Ivkovič and E.L. Lloyd. A Fundamental Restriction on Fully Dynamic Maintenance of Bin Packing, *Information Processing Letters*, 59(4): 229-232, 1996.

One dimensional semi-on-line bin packing algorithms[1]

János Balogh, József Békési, Gábor Galambos, and Gerhard Reinelt

In the talk a family of semi-online bin packing algorithms is defined and analyzed. In case of semi-online algorithms in each step it is allowed to proceed one of certain operations as repacking, reordering or buffering some elements before they having packed. It is clear that such type of algorithms are acceptable if they perform better than the on-line ones. (We require that the more information we need the better algorithm.)

Here we deal with semi-on-line algorithms which allow the repacking. We suppose that in these algorithms any repacking has unit cost independently from the size of an item. Similarly, we assume that the maximum number of elements to be repacked in each step are bounded by a foregiven constant k. We call these algorithms as *k-repacking semi-on-line algorithms*.

For every positive integer k we define a $k - repacking$ algorithm, called $HFR - k$. We prove that the asymptotic competitive ratio (ACR) for a given k is not larger than $\frac{3}{2} + \frac{b_k}{1-b_k}$ where b_k is the root of the equation $\frac{3}{2} + \frac{b_k}{1-b_k} = \frac{1}{\frac{1}{2}+kb_k}$ for which $b_k \in [0, \frac{1}{6k}]$ holds. One can see that for enough large – but finite – values of k the ACR of the algorithm tends to quick to $1, 5$. So, in some special cases – already for small k-s – we can improve some on-line results.

The first interesting particular case is the $k = 2$. Then $ACR(HFR\text{-}2) = 1, 5728\ldots$, which is better than the ACR of the best known on-line algorithm $(1, 58889\ldots)$ [1]. The case $k = 4$ is also remarkable: Here $ACR(HFR\text{-}4) = 1, 5389\ldots$, and so it is smaller than the best known lower bound $(1, 5401\ldots)$ for the on-line algorithms pubished in [2]. This means that the 4-repacking semi-on-line algorithms are more competitive than the on-line ones.

These are those results which show that our algorithm uses well the extra informations about the elements of the list and the relaxation of the on-line rule. Two open questions: Is there such a semi-on-line algorithm which may repack only one item in each step and its ACR is smaller than $1, 583\ldots$? Is there such a semi-on-line algorithm which has a better ACR than $1, 5401\ldots$ and it repacks less than 4 items in each step?

References

[1] S.S. Seiden. On the Online Bin Packing Problem. *Journal of the ACM*, 49(5): 640-671, 2002.

[2] A. van Vliet. An Improved Lower Bound for Online Bin Packing Algorithms. *Information Processing Letters*, 43(5): 277-284, 1992.

An On-line Speaker Adaptation Method for HMM-based ASRs

András Bánhalmi, Dénes Paczolay, and András Kocsor

When building a robust continous speech recognition system, one good way of improving the recognition accuracy [4] is via speaker adaptation [2]. All the state-of-the-art dictation systems use techniques to adapt the initial speaker independent model to a new speaker. These techniques require a large amount of sentences for utterance by the user, and the computation of the adaptation begins after recording these predefined sentences. The dictation system allows speech recognition only after this long-time adaptation procedure has been completed.

Our goal is to avoid this time-consuming procedure by creating a method that is embedded into a continous speech recogniser to retrieve the data for the adaptation process. The method proposed by us can be used with HMM-based speech recognisers [1] where multistack structure is used to store the hypotheses. We call a "hypothesis" a phoneme series, and it has a probability at a specific time. The hypotheses are stored in a stack ordered according to their probability. When a hypothesis is being extended (the next sound frame is being evaluated), the grammar module is asked for the possible phoneme continuatations and probabilities for them. The acoustic probablilty given by the HMM and the probability given by the grammar module will change the order of the hypotheses in the next stack. If two or more hypotheses have the same phoneme series, they will be fused into one hypothesis. This can be done without losing information in HMM-based systems. Then, using Viterbi and N-best cutting we drop the least likely hypotheses from the next stack. More details can be found about systems like ours in [3].

Of course embedding an automatic adaptation data retrieval method into a system like our, will introduce some problems. Not only the phoneme series and the data of the last phoneme HMMare needed to be stored by the system, but the tables of the most probable previous state, the tables of the gaussian components, and the tables containing the information of the jumps to the next phoneme HMM's start state are also required. Moreover, to create fast recognition, it is important to decrease the numbers of the hypotheses by fusing them. If only recognition is performed, all the hypotheses with the same phoneme series can be fused, but when adaptation information is stored, not all these hypotheses can be. In this paper we propose a method for storing the adaptation data efficiently in a graph structure, then we can examine which hypotheses can be fused and which can not.

Using an adaptation data retrieval method like this, the sentences for utterances need not be predefined, as data retrieval could be run during the recognition phase. Using this technique in practice gives rise to certain other problems. The retrieved data can be used for adaptation only if they are accurate, so the user should approve the correctness of the recognition result. If the recognition result is approved, then the accuracy of the retrieved data will be high only if the segmentation borders between phonemes are well determined. Here we will compare the accuracy of our continous automatic segmentation with that automatic segmentation which is based on a predefined sequence and dynamic programming.We will use the manually segmented data as a baseline. The above mentioned automatic segmenting algorithm using a predefined phoneme series builds one large HMM model, which is a chain of the HMMs labeled by the phonemes of the sentence. When segmenting phonetically the given utterance, this large HMM model will be evaluated via the Viterbi algorithm, and the most probable path of HMM states will be traced back to give us the phonetic segmentation. The effect of these segmentation methods will be compared by the accuracy of the recognition after adaptation. Among the widely-used adaptation methods we use the well-known MLLR adaptation algorithm for the evaluation.

References

[1] C. Becchetti and L. P. Ricotti. Speech Recognition, *John Wiley and Son*, England, 2000.

[2] X. Huang, A. Acero, and H. Hon. Spoken Language Processing, *Englewood Cliffs, NJ, Prentice Hall, 2001.*

[3] A. Kocsor, A. Bánhalmi, D. Paczolay, J. Csirik, and L. Pávics. The Oasis Speech Recognition System for Dictating Medical Reports, *Annual Congress of the European Association of Nuclear Medicine, 05, 2005.*

[4] L. Neumeyer, A. Sankar, and V. Digalakis. A Comparative Study of Speaker Adaptation Techniques, *Proceedings Eurospeech, pages 1127-1130, 1995.*

Investigation of a Delayed Differential Equation with Verified Computing Technique

Balázs Bánhelyi

Consider the following delayed differential equation:

$$y' = -\alpha \left(e^{y(t-1)} - 1 \right),$$

where $\alpha \in \mathbb{R}^+$ is a parameter.

When $\alpha \leq 1.5$, it is known, that the trajectory converges to zero, and when $\alpha \geq \pi/2$, the trajectory converges to different periodic solutions. Thus, the question is the $[1.5, \pi/2]$ interval. We want to prove that there do not exist periodic solutions of this delayed differential equation with any $\alpha \in [1.5, \pi/2]$ parameter as conjectured in [6].

The analysis of this problem is very hard with numerical methods, hence in the first part we consider an easier problem. We are interested in checking whether for all $\alpha \in \left[\frac{3}{2}, \frac{\pi}{2}\right]$, there exists a unit length time segment where the absolute value of the solution is less than 0.075. Let the initial function be

$$\phi(s) \equiv -11,$$

where $s \in [-1, 0]$.

Most verified techniques for solving ordinary differential equations apply a Taylor series. Our technique is based on the same idea too. The general form of the Taylor-series is:

$$y(x) = \sum_{k=0}^{n-1} \frac{(x - x_0)^k y^{(k)}(x_0)}{k!} + r_n. \tag{1}$$

Using the mean-value theorem, r_n can be bounded by

$$r_n = \frac{(x - x_0)^n}{n!} y^{(n)}(x^*), \tag{2}$$

for some $x^* \in [x_0, x]$ ($x_0 \leq x$).

If we want a better approximation of the solution, we have to use higher derivatives. We can characterize the higher derivatives with this formula:

$$y^{(k)}(t) = -\alpha y^{(k-1)}(t-1) + \sum_{i=1}^{k-1} \binom{k-2}{i-1} y^{(i)}(t-1) y^{(k-i)}(t).$$

In this case the verification means verification in the mathematical sense, hence rounding and other errors were considered and bounded. Instead of real numbers, we can also calculate with intervals. In case the bounds of the result interval are not representable, then they are rounded outward. In this problem we used the multiple precision interval arithmetic libraries (C-XSC, PROFIL/BIAS) [5, 3].

To provide a mathematical proof, it is not enough to use interval arithmetic, we have to use the formula in a correct, suitable form. We can use the Taylor-series to bound the results:

$$Y(t_1) = \sum_{i=0}^{n-1} Y^i(t_0) \frac{(t_1 - t_0)^i}{i!} + Y([t_0, t_1])^n \frac{(t_1 - t_0)^n}{n!},$$

$$y([t_0, t_1]) = \sum_{i=0}^{n-1} Y^i(t_0) \frac{([0, t_1 - t_0])^i}{i!} + Y([t_0, t_1])^n \frac{([0, t_1 - t_0])^n}{n!}.$$

We use two fix length lists to store the solution bounds. The first list contains the solution and the derivatives on time intervals, which cover the unit length time segment. The other list stores the solution and the derivatives in concrete time points. We calculate the new elements of the lists with the earlier discussed formula. The oldest elements are deleted from the lists, and the new ones are inserted. This technique has three parameters: step length, maximum derivate rank, and a precision of the interval arithmetic. We combine our method and an optimization technique to determine the optimal values for these parameters.

We proved the above original statement with this technique for some tiny intervals around certain computer representable numbers. But we were not able to prove it for all points of the α parameter interval, due to the large amount of necessary CPU time. We show the details of the newest program which is based on the earlier technique and the idea some theoretical results. This compound method is able to prove the original conjecture for $\alpha \in [1.50, 1.568]$.

References

[1] N. S. Nedialkov, K. R. Jackson, and G. F. Corliss, Validated Solutions of Initial Value Problems for Ordinary Differential Equations, *Applied Mathematics and Computation*, 105, 21-68, 1999.

[2] M. Berz, K. Makino, K. Shamseddine, G. Hoffstätter and W. Wan. Computational Differentiation: Techniques, Applications, and Tools, *COSY INFINITY and its Applications to Nonlinear Dynamics*, 365-367, SIAM, 1996.

[3] O. Knüpel. PROFIL – Programmers Runtime Optimized Fast Interval Library, *Berichte des Forschungsschwerpunktes Informations- und Kommunikationstechnik, TU Hamburg-Harburg*, 93.4, 1993.

[4] O. Knüpel. A Multiple Precision Arithmetic for PROFIL, *Berichte des Forschungsschwerpunktes Informations- und Kommunikationstechnik, TU Hamburg-Harburg*, 93.6, 1993.

[5] R. Klatte, U. Kulisch, C. Lawo, M. Rauch, and A. Wiethoff. C-XSC: A C++ Class Library for Extended Scientific Computing, *Springer-Verlag, Berlin*, 1993

[6] E. M. Wright. A non-linear difference-differential equation, *Journal für die Reine und Angewandte Mathematik*, 194, 66-87, 1955.

[7] T. Csendes, B. Bánhelyi, and L. Hatvani. Towards a computer-assisted proof for chaos in a forced damped pendulum equation, *Journal of Computational and Applied Mathematics*, 2005, submitted.

[8] B. Bánhelyi and T. Csendes. A verified computational technique to locate chaotic regions of Hénon systems, *Proceedings of the 6th International Conference on Applied Informatics*, 297-304, 2004.

[9] T. Csendes, B. M. Garay, and Bánhelyi, B.. A verified optimization technique to locate chaotic regions of a Hénon system, *Journal of Global Optimization*, 2005, submitted.

[10] B. Bánhelyi, T. Csendes, and B. M. Garay. Optimization and the Miranda approach in detecting horseshoe-type chaos by computer, 2005,
`http://www.inf.u-szeged.hu/~csendes/publ.html`

[11] T. Krisztin. Periodic orbits and the global attractor for delayed monotone negative feedback, *Electronic Journal of Qualitative Theory of Differential Equations*, 2000, 15 1-12.

[12] T. Csendes. Nonlinear parameter estimation by global optimization - efficiency and reliability, *Acta Cybernetica*, 1988, 8, 361-370.

Model driven testing of component based systems

Gábor Bátori, Zoltán Theisz, and Domonkos Asztalos

The growing demand of the telecommunication market for complex systems cannot be easily satisfied without new development paradigms. Model Driven Architecture (MDA) [1] offers a good way to achieve the desired complexity management. However, MDA uses UML as a notation and in many cases UML is too complicated to use because its philosophy does not match to the one of the modelled system. In that case a well defined modeling language, that is, a Domain Specific Language for the problem is more effective. Domain specific modelling requires at least three ingredients to be well-defined:

- Metamodel that defines which concepts have relevance in the problem domain, how they are related to each other and in which manner they can be put together correctly to provide consistency.

- Model that uses the concepts defined in the metamodel and establishes an instance set of these concepts which describes the problem.

- Model translators that add semantical meanings to the metamodel concepts. A translator could be e.g. a model-to-code translator which generates source code from the model or a model-to-model translator which translate a model describing one aspect of the system to another model describing different aspect. An example model translator is the one which creates the architecture model of the modelled system from the functional model.

The modelling tool we use that enables the domain specific modelling of systems is the Generic Modeling Environment (GME) [2]. GME supports metamodelling and metamodel aware modelling of the application domain. Furthermore, it provides two ways of defining model transformation; an interpreter based and a more formalised graph transformation based method. Increased customer requirements for building highly reconfigurable and reusable system can be handled with an appropriate system architecture. To control these demands component system architecture is necessary that enables reusable modularized services to be composed, interconnected, configured and deployed to create applications rapidly and robustly in dynamically changing distributed environments. To fulfil all requirement of the telecommunication software development we created the ErlCOM system [3], that is, an innovative combination of the beneficial aspects of component-based programming and model based development. There is a strong need for effective testing of model based applications but there is no commonly accepted method. Our main goal is to create a generic framework for testing generated from the metamodel and the model of the application. The generated framework can be used for writing test cases manually but the main goal is to serve as a basis for different automatic test generation algorithm plug-ins. Moreover, since all plug-in test generation algorithms rely on the services of the underlying test framework, therefore, they provide comparable results.

References

[1] R. Soley. Model Driven Architecture: An introduction, `http://www.omg.org/mda`.

[2] G. Karsai, J. Sztipanovits, A. Ledeczi, and T. Bapty. Model-integrated development of embedded software, *Proceedings of the IEEE*, vol. 91, pp. 145-164, January 2003.

[3] G. Batori, Z. Theisz, and D. Asztalos. Robust Reconfigurable Erlang Component System, *Erlang User Conference*, Stockholm, Sweden, November 2005.

Test Generation with Dynamic Impact Analysis for C++ Programs

Mihály Biczó, Krisztián Pócza, and Zoltán Porkoláb

During regression testing tests from an existing test suite are run against a modified version of a program in order to assure that the underlying modifications do not cause any side effects that would demolish the integrity and consistency of the system. Although the test set can grow uncomfortably large, we can safely focus on a smaller subset of modification traversing tests, since only these tests might reveal an error. Unfortunately, in many cases the original and modified system generates identical output for a modification traversing test, which means that the test in question is not effective. In this paper we present an efficient C++ specific approach to automatically repair ineffective test cases. Our solution is a two stage process, both stages performing a forward algorithm, which guarantees results can be obtained on-the-fly. First, the exact cause of ineffectiveness and the affected statement is identified; while in the second stage input variables whose values should be altered to repair the test are selected. Applying our algorithm requires precise instrumentation of the source code. Still this method can be successfully applied to large-scale applications, therefore it not only helps system maintainers to automatically manage the test suite, but it also assures that the system would be run against a harsh test set. Under these circumstances we get a very reliable and safe solution, which minimizes the possibility of an unexpected system failure.

References

[1] B. Korel and A. M. Al-Yami. Automated Regression Test Generation, *ISSTA*, 1998: 143-152

[2] B. Korel. Automated Test Data Generation for Programs with Procedures, *ISSTA* 1996: pp. 209–215

[3] W. Wong, J. Horgan, S. London, and H. Agrawal. A study of effective regression testing in practice, In *Proc. of the Eighth Intl. Symp. on Softw. Rel. Engr.*, pp. 230–238, Nov. 1997. 10

[4] D. Kung, J. Gao, P. Hsia, F. Wen, Y. Toyoshima, and C. Chen. Change impact identification in object-oriented software maintenance, In *Proceedings of the IEEE International Conference on Software Maintenance*, pp. 202–211, September 1994.

[5] B. G. Ryder and F. Tip. Change impact analysis for object oriented programs, In *Proceedings of PASTE '01*, pp. 46–53, Snowbird, Utah, USA, June 2001.

[6] G. Rothermel and M.J. Harrold. Selecting tests and identifying test coverage requirements for modified software, *Proceedings of the International Symposium on SoftwareTesting and Analysis*, pp. 169–184, August 1994.

Network topology discovery

Vilmos Bilicki

Because of the cheap price and high capacity of Ethernet switches, Ethernet technology is conquering the backbone and access networks. It is now an accepted design pattern in network engineering to use VLAN-s not only in access and distribution but in the backbone layer too. As the Ethernet started its career as a LAN technology, it does not have OAM (Operation, Administration, and Maintenance) capabilities comparable to SDH, ATM or MPLS. So it is not surprising that it does not have support for error handling and signalling comparable to ICMP. For the network providers and system administrators knowledge of the actual and the past topology of a network is the most important thing. As current Ethernet technology does not provide special tools for topology maintenance and discovery there are several special methods available in the literature that can be used for discovering the actual Layer 2 topology of a network. In this article we will summarize the best known Layer 2 topology discovery methods and we will compare them on the basis of the complexity, the amount of the information needed and the resolution provided. A novel CAM table-based approach will be shown and compared with existing technologies.

Paxos with multiple leaders

Vilmos Bilicki and József Dombi Dániel

Distributed computing is becoming evermore important as the e-* services become part of our everyday life. It is not an easy task to fulfil the demand for 7x24 availability. One possible way of creating reliable services is to use multiple servers with the same set of services. An issue here is that these servers should have the same knowledge so that they can work in a consistent manner. This problem is frequently solved with the help of different group communication services. These services can provide different levels of fault tolerant services from FIFO multicast to virtual synchrony. One well known group communication service is the Paxos algorithm which is a 2000 year old algorithm first used by ancient Greeks as a parliamentary protocol. The traditional Paxos algorithm uses, among other abstractions, a single leader functionality to serialise parallel events which in turn provides a global ordering property for the system. However, this single dedicated point weakens the reliability of a distributed system. Here we will present several new solutions for distributed global ordering based on the Paxos algorithm and we will compare them on the basis of message complexity and the safety level provided.

Models for Predicting the Performance of ASP.NET Web Applications

Ágnes Bogárdi-Mészöly, Tihamér Levendovszky, and Hassan Charaf

Web applications play an important role in computer science nowadays. The most common consideration is performance, because these systems must provide services with low response time, high availability, and certain throughput level. The performance-related problems emerge very often only at the end of the software project. With the help of properly designed performance models, the performance metrics of a system can be determined at the earlier stages of the development process [1]. Today one of the most prominent technologies of web applications is Microsoft .NET. The goal of our work is to predict the response time, the throughput and the tier utilization of ASP.NET web applications, based on a queueing model [2, 3] handling multiple session classes, with MVA (Mean-Value Analysis) evaluation algorithm and approximate MVA [4, 5], in addition with balanced job bounds calculation [6]. Handling one session class, for large values of customers or if the performance for smaller values is not required, MVA can be too expensive computationally. Handling multiple session classes, the time and space complexities of MVA are proportional to the number of feasible populations, and this number rapidly grows for relatively few classes and jobs per class. Thus, it can be worth using an approximate MVA algorithm or a set of two-sided bounds.

We have estimated the model parameters (maximum number of customers, number of tiers, average user think time, visit number, average service time) based on one measurement. With the help of MATLAB, we have implemented the MVA and approximate MVA algorithms for closed queueing networks along with the calculation of the balanced job bounds. The scripts compute the response times, the throughputs and the tier utilizations up to a maximum number of customers. MVA provides a recursive way, approximate MVA computes these in a few steps, while balanced job bounds method completes in one step. We have tested a web application with concurrent user sessions to validate the models in ASP.NET environment. Our results have shown that the models predict the response time and the throughput acceptably, with MVA, approximate MVA, and calculation of balanced job bounds as well. Furthermore, the presentation tier becomes congested firstly. The utilization of the database tier is the second, and the utilization of the business logic queue is the last one.

References

[1] C.U. Smith and L.G. Williams. Building responsive and scalable web applications, *Computer Measurement Group Conference*, Orlando, FL, USA, 2000, pp. 127-138.

[2] D.A. Manescé and V.A.F. Almeida. Capacity Planning for Web Services, *Prentice Hall*, 2002.

[3] B. Urgaonkar, G. Pacifici, P. Shenoy, M. Speitzer, and A. Tantawi. An Analytical Model for Multi-tier Internet Services and its Applications, *Journal of ACM SIGMETRICS Performance Evaluation Review*, 33(1), 2005, pp. 291-302.

[4] R. Jain. The Art of Computer Systems Performance Analysis, *John Wiley and Sons*, 1991.

[5] M. Reiser and S.S. Lavenberg. Mean-Value Analysis of Closed Multichain Queuing Networks, *Journal of Association for Computing Machinery*, 27, 1980, pp. 313-322.

[6] J. Zahorjan, K.C. Sevcik, D.L. Eager, and B. Galler. Balanced Job Bound Analysis of Queueing Networks, *Journal of Communications of the ACM*, 25(2), 1982, pp. 134-141.

Methods for Retrieving and Investigating Performance Factors in ASP.NET Web Applications

Ágnes Bogárdi-Mészöly, Tihamér Levendovszky, and Hassan Charaf

New frameworks and programming environments were released to aid the development of complex web applications. These new languages, programming models and techniques are proliferated nowadays, thus, developing such applications is not the only issue anymore: operating, maintenance and performance questions have become of key importance. One of the most important factors is performance, because network systems face a large number of users, they must provide high-availability services with low response time, while they guarantee a certain level of throughput. These performance metrics depend on many factors. Several papers have investigated various configurable parameters, how they affect the performance of a web application. Statistical methods and hypothesis tests are used in order to retrieve factors influencing the performance. An approach [1] applies analysis of variance.

Today one of the most prominent technologies of distributed systems and web applications is Microsoft .NET. An ASP.NET application server has several settings which can affect the performance [2]. Our primary goal was to investigate factors influencing the response time, because it is the only performance metric to which the users are directly exposed. We tested a web application with concurrent user sessions [3], focusing on the effect of the different thread pool properties, the global queue limit and the application queue limit on performance. The results are analyzed in a qualitative manner which is followed by using statistical methods with the help of MATLAB: independence tests [4] to investigate which factors influence principally the performance. Our experiments have shown that the *maxWorkerThreads, maxIOThreads, minFreeThreads, minLocalRequestFreeThreads, requestQueueLimit*, and *appRequest-QueueLimit* properties are performance factors. In addition, we have determined the distribution of the response time as a function of the thread pool attributes settings. The normality has been intuitively founded by graphically methods, and has been proven with hypothesis tests [5]. Finally, optimal settings according to the performance-related requirements are determined as a function of client workload.

References

[1] M. Sopitkamol and D.A. Menascé. A Method for Evaluating the Impact of Software Configuration Parameters on E-Commerce Sites, In *Proceedings of the ACM 5th International Workshop on Software and Performance*, Spain, 2005, pp. 53-64.

[2] J.D. Meier, S. Vasireddy, A. Babbar, and A. Mackman. Improving .NET Application Performance and Scalability, *Patters & Practices*, Microsoft Corporation, 2004.

[3] J. Aldous and L. Finnel. Performance Testing Microsoft .NET Web Applications, *Microsoft Press*, 2003.

[4] C. H. Brase and C. P. Brase. Understandable Statistics, *D. C. Heath and Company*, 1987.

[5] R. Jain. The Art of Computer Systems Performance Analysis, *John Wiley and Sons*, 1991.

Extracting Human Protein Information from MEDLINE Using a Full-Sentence Parser

Róbert Busa-Fekete, Kornél Kovács, and András Kocsor

Nowadays the MEDLINE [1] database is becoming the most comprehensive biomedical abstract repository among the life sciences literature. Due to its easy access and availability, it is one of the most widely-used sources of scientific data that uses several information retrieval systems. The NLM (U.S. National Library of Medicine) maintained MEDLINE database contains over 13 million references from about 4900 journals dating from 1965 to the present, and it is updated weekly. Obviously it is a crucial task in bioinformatics text mining to develop an automatic system that extracts information about genes and their interactions. That is why we were motivated in building an information extraction (IE) system that makes use of natural language processing (NLP) techniques.

In biological studies, the researchers are mostly interested in the interactions of the genes, so in this area of science biologists require an IE system that can search for relationships among human proteins [2]. That is why we will focus here only on genes that occur in living human cells. The National Center for Biotechnology Information (NCBI) [3] has many databases about gene interactions distributed taxonomically. Using these data sets we can easily obtain a subset of MEDLINE containing information about human gene interactions. We assume that in each observed abstract the gene names occur in it. We used a thesaurus containing about 58,000 gene names and their synonyms in order to annotate the gene names in the abstracts. The thesaurus was built up using three sources: Unified Medical Language System (UMLS) Metathesaurus [4], UMLS SPECIALIST Lexicon[4] and the Agilent Technologies[5] database.

With our approach we would like to learn more about the interactions of genes using full-sentence parsing [6, 7, 8]. Given a sentence, the syntactic parser assigns to it a syntactic structure, which consists of a set of labelled links connecting pairs of words. The parser also produces a constituent representation of a sentence (showing noun phrases, verb phrases, and so on). Using the syntactic information of each abstract, the biological interactions of genes can be predicted. Our IE system can handle certain types of gene interactions with the help of machine learning (ML) [9] methodologies (Artificial Neural Network [10], Support Vector Machines [11, 12]). Actually, many features of a syntactic tree can be represented as a multidimensional vector (i.e. depth and frequencies of different labels).

The performance of the IE process is influenced mostly by the quality of the syntactic parsing, which is why we chose to examine several methods to see how well they perform. A traditional approach, namely the Link Parser [13], will be applied as the baseline system. Because the LINK Parser is a general purpose syntax analyzer, its special biomedical extension will be also investigated. In addition, we propose a novel ML-based syntax parser for English. The algorithm interprets the words in a sentence as individual subtrees, and it concatenates the most suitable adjoining subtrees according to the ML model used.

When designing our system we had to take into account the fact that MEDLINE is a rapidly growing system and that the data is stored in compressed XML file format. So we created a framework which can handle the abstracts and their updates in their raw form, and incorporate them into our IE system. We evaluated our system on the well-known annotated Human Protein Reference Database (HPRD) [14] corpus and obtained some useful results.

References

[1] http://www.pubmedcentral.nih.gov/

[2] T. Sekimizu, H.S. Park and Jun'ichi. Identifying the Interaction between Genes and Gene Products Based on Frequently Seen Verbs in Medline Abstracts, *Genome Informatics*, 9:62-71, 1998.

[3] `http://www.ncbi.nlm.nih.gov/`

[4] `http://www.nlm.nih.gov/research/umls/`

[5] `http://www.home.agilent.com/`

[6] D. Sleator and D. Temperley. Parsing English with a Link Grammar, *Carnegie Mellon University Computer Science technical report* CMU-CS-91-196, October 1991.

[7] J. Lafferty, D. Sleator, and D. Temperley. Grammatical Trigrams: A Probabilistic Model of Link Grammar, *Proceedings of the AAAI Conference on Probabilistic Approaches to Natural Language*, October, 1992.

[8] D. Grinberg, J. Lafferty, and D. Sleator. A robust parsing algorithm for link grammars, *Carnegie Mellon University Computer Science technical report* CMU-CS-95-125, and *Proceedings of the Fourth International Workshop on Parsing Technologies*, Prague, September, 1995.

[9] V. N. Vapnik. Statistical Learning Theory, *John Wiley and Son*, 1998.

[10] C.M. Bishop. Neural Networks for Pattern Recognition, *Oxford University Press*, 1995.

[11] N. Cristianini and J. Shawe-Taylor. Support Vector Machines and other kernel-based learning methods, *Cambridge University Press*, ISBN 0-521-78019-5, 2000.

[12] B. Schölkopf, C.J.C. Burges, and A.J. Smola. Advances in Kernel Methods: Support Vector Learning, *MIT PRess*, Cambridge, MA, 1999.

[13] `http://www.link.cs.cmu.edu/link/`

[14] `http://www.hprd.org/`

A Rule-based Transformation Engine for Web Page Re-authoring

Péter Cserkúti and Hassan Charaf

This paper introduces SmartWeb, a system for web page re-authoring. It reveals its main algorithms and highly extensible architecture. The purpose of web page re-authoring (a.k.a. web content adaptation) is to transform the original web content in a way that it can adequately be displayed on a client device, let it be a mobile, a PDA or a desktop computer. The adaptation process should take the capabilities of the rendering client into consideration such as screen resolution, network bandwidth, type and speed of processor, amount of memory or software configuration and produce the best version of the original content for that. This results a renderer independent web browsing capability.

SmartWeb is a proxy-based solution. It is always situated between the client and the web server that contains the requested web page. It catches all the queries of the client and before returning the requested content it fulfills an adaptation process on the original page.

The essence of SmartWeb is the adaptation process itself. Its input is the original html page and its output is a modified version of it. SmartWeb handles html pages as trees that we call transformation tree. The adaptation process is realized by the help of graph transformation. One can define special graph transformation rules that the engine can execute against the transformation tree. A rule is made up by two parts: a left hand side (LHS) pattern and a right hand side (RHS) tree with witch the LHS pattern must be replaced. The paper introduces the method how SmartWeb facilitates defining LHS patterns representing common block types in web pages, and exhibits its technique for describing RHS trees for them. A major point of the graph transformation process is matching the specified patterns in the transformation tree. There are existing algorithms for pattern matching such as Ullmann's algorithm or the VF2 algorithm. However because of the complexity of our pattern definition method they can not be used here or must be extended. This paper will also introduce an algorithm for matching these patterns in the transformation tree and an other algorithm for transforming them according to the RHS tree. There are two things that make these algorithms powerful and SmartWeb easy and convenient to configure: first when defining an LHS pattern one can freely refer to an other previously matched pattern as a node of the graph and second when defining the RHS tree referring to subtrees of the LHS pattern is allowed.

SmartWeb is an extensible framework for web content adaptation. Its extensibility has got two levels. First, there are pipelines defined in the system and all the algorithms implemented are attached to an adequate one. New algorithms that analyze or transform the transformation tree can easily be implemented and attached to the system. Second, the set of transformation rules can be extended. New patterns and rules can be defined in XML. The paper also introduces the architecture of SmartWeb.

Probabilistic confidence prediction in document clustering

Kristóf Csorba and István Vajk

In this paper we propose a novel technique to predict the confidence of the clustering of a document before performing the clustering procedure itself. If an application has time limitations for the document clustering and it has to cluster as much documents as possible, it can use the predicted confidence value to sort the documents and process the ones with high expected confidence first. This means that the system begins with the documents for which the probability of a certain result is high. If it has enough time, it can process the remaining documents as well, but only after the most beneficial ones. The base method consists of three key techniques of the feature-space based document clustering: term frequency quantizing, singular value decomposition and double clustering.

Term frequency quantizing removes the unnecessary variation of the term frequencies. This step was introduced through the idea, that the exact occurrence number of a frequent term in a document does not provide more information than a "frequent" status.

Singular value decomposition was already shown to be capable to capture semantic relationships between terms based on occurrence behavior similarities. SVD is used to reduce the feature space dimensionality before the clustering itself.

Double clustering means, that instead of clustering the feature vectors of the documents, the feature vectors of the individual terms are clustered to create termclusters first. Document clustering is then performed in a feature-space based on these term-clusters.

After these steps the document clustering is performed in the resulting feature-space by employing the k-means algorithm using the cosine distance measure. During the clustering beside the assigned cluster an additional confidence value is provided for every document to mark ambiguous cases.

Before the procedure above a *probabilistic model* of it can be employed to calculate the expected confidence value. Before employing, this *prediction method* uses an initial training phase: after executing the original clustering method on a training set, the internal state is analyzed to retrieve probability of internal decisions. After the training the predictions for the further documents are calculated with a single matrix multiplication. This prediction is faster, but therefore less accurate. The predicted results can be used to *reorder the documents* to process the easier cases first which helps timecritical applications to get more sure results in the available time.

In this paper we provide a description of the prediction method and experimental results in connection with absolute performance and about the correlation between the predicted and measured results.

A Performance Model for Load Test Components

Máté J. Csorba and Sándor Palugyai

Performance issues of telecommunication system testing, load testing with Testing and Test Control Notation version 3 (TTCN-3) specifically, is investigated in this work. To be able to design test components that simulate behavior of real nodes in a telecom test network and that are even interchangeable with a real node in terms of performance, modeling is desirable. The aim of successful traffic mix composition is to simulate behavior of real nodes, or even to produce an equivalent counterpart of the real node in TTCN-3.

In TTCN-3, configuration of the test system can be set dynamically. This means that a configuration may consist of several components participating in the test, either as a component that communicates directly with the System Under Test (SUT), or as a component having only registration or internal purposes, meaning that it communicates only with parts of the test system itself. The various components used by a test system are called Parallel Test Components (PTCs). PTCs communicate with each other via test ports. Similarly, communication towards the SUT is established via test ports too.

Each test port that connects either two test components (internal port) or a test component and the interface towards the SUT is modeled as a FIFO queue for the incoming/outgoing messages. Properties of the FIFO queues assigned to a test port are dependent on the actual implementation of the TTCN-3 compiler. The queues can be infinite in principle, as long as the system memory lasts, but might overflow indeed. More importantly, in a load test system response time must be considerably short. This means that it is inexpedient to implement a virtually infinite buffer for a PCO and forget about message loss at all. Although a sufficiently long buffer might eliminate message loss, but response time increases significantly at the same time.

The actual behavior of a test case is defined by dynamic behavioral statements in a test component that communicates over certain test ports. Usually, sequences of statements can be expressed as a tree of execution paths called alternatives. In TTCN-3, the alt statement is used to handle events possible in a particular state of the test component. These events include reception of messages, timer events and termination of other PTCs. The alt statement uses a so-called snapshot logic. However, execution does not stop after a snapshot was taken, so the state of the test component and the queues assigned to it might change in between. So, generally the two most significant factors we consider, while performance evaluating a test component are the matching mechanism and the queuing at the test ports. In our modeling approach, we build an analytic model of test components that use alternatives to handle internal messages and messages coming from the SUT. We describe concurrent queues underlying the test ports of a component with a stochastic process and calculate the steady-state solution. After solving the analytical model, we predict the probability that one of the queues contain a message that is postponed indefinitely, because of the fact that race conditions arise between the queues. The probability of an indefinite postponement is calculated as a function of arrival intensities at the corresponding queues and of other parameters relevant to the implementation of the actual test component.

Extending a system with verified components[1]

Ákos Dávid, Tamás Pozsgai, and László Kozma

In [1] an outline of an educational framework was presented that is able to help students of Information Technology to build applications from a set of verified components. The testing of component-based systems can be extremely complicated because it is usually not possible for system developers to pre-check the compatibility of the individual parts before the actual integration takes place. The situation may be even worse when prefabricated components, so-called components off-the-shelf (COTS) are used as the environment of the development and the deployment are rarely the same. Even if they are the same or relatively similar, the use of these components may differ considerably from the original purpose they were developed for. Testing of a component-based solution consists of two distinct activities: testing of the components and an integration testing of the assembled application. A system cannot be considered correct if its components do not work properly. On the other hand, can the proper functioning of a system be guaranteed if its components are verified in some ways? Unfortunately, all the information on the correctness of the individual components become irrelevant and out-of-date from the moment they are used anywhere but the original environment. The solution to this problem can be based on the idea of building correct programs in which reliability is built-in. This means the correctness properties of a program may be transferable or at least checkable in other environments as well. The concept behind that is the method of "design-by-contract" introduced by Bertrand Meyer in Object-Oriented Programming (OOP) [1]. The idea of using contracts in combination with built-in testing technologies according to Hans Gerhard Gross would offer a long-waited solution to increase our confidence in third-party components [3]. Considering other alternatives model-checking seems to be applicable as an automatic technique for verifying finite state concurrent systems [4]. The main challenge in model checking is dealing with the state space explosion problem. This problem occurs in systems with many components that can make transitions in parallel. It means that extending an existing system with one or more components may cause difficulties for model checking tools by increasing the state space exponentially. In [5] open incremental model checking is introduced - and compared to traditional modular model checking methods - to address the changes to a system rather than re-checking the entire system model including the new extensions. In our paper we study the practical aspects of using open model checking by working out a sample system consisting of verified components and prove the efficiency of this new method within an educational framework.

References

[1] Á. Dávid, T. Pozsgai, and L. Kozma. Educational framework for developing applications built from verified components *Proceedings of the Ninth Symposium on Programming Languages and Software Tools*, pp. 7-18, Aug 13-14, 2005, Tartu, Estonia, published by Tartu University Press.

[2] B. Meyer. Object-Oriented Software Construction Second Edition, *Prentice Hall*, 1997.

[3] H.G. Gross. Component-Based Software Testing with UML, *Springer*, 2005.

[4] E. Clarke, O. Grumberg, and D. Peled. Model Checking, *MIT Press*, 2000.

[5] N.T. Thang and T. Katayama. Open Incremental Model Checking (Extended Abstract) *Proceedings of ACM SIGSOFT Symposium on the Foundations of Software Engineering*, Oct 31-Nov 1, 2004, Newport Beach, California, USA.

Improving DHT-Based File System Performance with B-trees

Róbert Dávid and Gábor Vincze

In this paper, we present a way of improving the performance of network file systems based on Distributed Hash Tables (DHT) by incorporating B-tree variant into the file system.

System Structure

The proposed system has four subsystem layers:

- File System (user interface, top layer)

- Search trees, indexing

- Reliable data block storage

- DHT (data block location and routing)

The top layer is a conventional file system interface: it should show to users files, attributes, directories, and should provide tools to add, modify, copy, delete files and their metadata. DHTs offer a scalable and completely decentralized storage mechanism [1, 2, 3, 4]. However, DHTs only solve the problem of Decentralized Object Location and Routing (DOLR), and do not provide any means for indexing and searching contents of the network. Many solutions have been proposed to build an indexing layer on top of the DHT, and offer file system functionality [5, 6]. In the Cooperative File System, a DHash layer provides redundancy and caching of file system data and metadata blocks, on top of which a File System layer provides the file system interface to users. In our system, we replace the simple replication of DHash with a more efficient block-coding algorithm, like IDA [7] or Reed- Solomon block coding. In our system, we add a search tree/indexing layer between the reliable data block storage and the File System. Modern conventional file systems use B-trees to improve performance [8], and outperform non-B-tree file systems by a factor of 10-15x in certain usage scenarios. We hope to use B-trees in DHT-based file systems to reduce the number of routing steps necessary to locate data blocks, which can have a significant impact in the case of a globally distributed network.

Search Trees

The search tree we propose is a B+-tree, we call C-tree, C for content addressing. Taking advantage of content addressing, we can increase the branching factor, while node sizes in bytes can stay the same. The concept is the following: Let the address of tree-nodes be calculated from the first and last index value, one can reach using the node in question. I proposed a hash function to spread nodes data uniformly over the DHT network. Now child nodes can be addressed without a pointer, we need only the "lowerthan- node" and "higher-than-node" index values the B+ algorithm already stores, and use the hash function. A little downside is that we must store the low and high limit of the node, but it is out weighted by the benefit of not having to store node addresses implicitly. An extra advantage is that the root blocks address will be always well known, because it must be the hash of the index of all zero bits and all one bits. To install multiple file systems over one DHT can be guaranteed to use of keyed cryptographic hash functions, encrypted storage, user specific keys.

Conclusion

By incorporating B-trees into DHT-based file systems, we hope to take DHT-based file systems beyond the simple initial architectures used in CFS or Pastis.

References

[1] I. Stoica, R. Morris, D. Karger, M.F. Kaashoek, and H. Balakrishnan. Chord: A scalable peer-to-peer lookup service for internet applications, In *Proc. ACM SIGCOMM* (San Diego, aug. 2001).

[2] S.P. Ratnasamy. A scalable content-addressable network (2002)

[3] P. Druschel and A. Rowstron. Past: Persistent and anonymus storage in a peer-to-peer networking environment, In *Proceedings of the 8th IEEE workshop on hot topics in operating systemy* (HOTOS 2001), (Elmau/Oberbayern, Germany, may 2001), PP. 65-70.

[4] B.Y. Zhao, J. Kubiatowicz, and A.D. Joseph. TAPESTRY: An infrastructure for fault-tolerant wide area location routing, (2001)

[5] F. Dabek. A cooperative file system, (2001)

[6] J.-M. Busca, F. Picconi, and P. Sens. PASTIS: A highly-scalable multi-user peer-to-peer file system, (2004)

[7] M.O. Rabin. Efficient dispersal of information for security, Load balancing, and fault tolerance, *JACM*, Vol. 36, no. 2, april 1989.

[8] REISERFS4, `http://www.namesys.com/V4/V4.html`

Model-Free Control Based on Reinforcement Learning

Zoltán Dávid and István Vajk

When controlling physical systems, the aim of the control mechanism is to influence the behavior of the process plant to achieve the desired design goals. An objective can be, for instance, to keep the output of the system near a reference signal. The performance metrics of a control can be measured in terms of the settling time and the error signal. Traditional control algorithm designs are based not on the physical plant, but on its mathematical model. However, the incorrectness or complexity of the model can make it hard to design an appropriate controller.

In several problem domains the model of the process is either not known or hard to handle. Applying a model-free controller in these domains can ease the design of the control structure. Several model-free control strategies have previously been introduced [1, 2, 3]. Here we discuss a method based on reinforcement learning. Reinforcement learning is a computational approach to learning how to map states of the environment to actions, in order to maximize a numerical reward signal. In the reinforcement learning scenario, the agent is in a continuous interaction with its discrete time environment, performing actions on it, and receiving state and reward signals of the environment. Any algorithm is considered to be a reinforcement learning algorithm if it is able to provide a mapping from states to actions, while maximizing the discounted reward (weighted sum of future rewards) in the long run.

Thus, reinforcement learning techniques can be applied in any environment, where the goal of the learning process can be represented with a numerical reward signal, and the environment can be modeled with a Markov Decision Process.

The main requirement of applying reinforcement learning in process control is to find an appropriate mapping from the closed loop system to a Markov Decision Process. In this paper, we examine several mapping techniques and reinforcement learning controller implementation. Moreover, we evaluate their appropriateness for process control by comparing them to a traditional PID controller.

Keywords: *reinforcement learning, process control*

References

[1] C.R. Edgar and B. E. Postlethwaite. MIMO fuzzy internal model control, *Automatica*, 2000, Vol. 34, pp. 867 - 877.

[2] K.K. Biasizzo, I. Skrjanc, and D. Matko. Fuzzy predictive control of highly nonlinearity pH process, *Computers and Chemical Engineering*, 1997, Vol. 21, pp. s613 - s618.

[3] A.P. Loh, K.O. Looi, and K.F. Fong. Neural network modeling and control strategies for a pH process, *Journal of Process Control*, 1995, Vol. 6, pp. 355 - 362.

Programming Language Elements for Correctness Proofs

Gergely Dévai

The study of formal methods to reason about program properties is getting a more and more important research area, as a considerable part of a software product's lifecycle is testing and bugfixing. The theoretical basis — such as formal programming models and reasoning rules [1, 2, 3, 5, 6, 7] — has been developed so far, but these are rarely used in industry [10]. The main reason for this fact is that formally proving a program property usually takes much more time than writing the program itself.

The goal of this research is to use programming language elements to make the construction of these proofs easier. The basic idea is to develop a new programming language where the source code contains the formal specification and the correctness proof of the implementation. The proofs are built up using *stepwise refinement* [4, 8, 9], as this technique provides *correctness by construction* [12], and also helps the programmers to make the right decisions during software development. The compiler of the language has to check the soundness of the proof steps and to generate the program code in a target language using the information of the proof.

Similarly to programs, proofs also contain schematic fragments. These can be managed efficiently using *"proof templates"* that have the same role in proof construction as procedures have in traditional program development. This leads to a special kind of *generative programming* [13]: by the instantiation of the templates a proof is constructed (and checked) in compile time and from the proof a target language program is generated which automatically fulfils all the requirements stated in the specification.

In this paper we show how language elements can be used to make specification statements easy to understand and to write, we introduce the basic refinement rules of the model and show how templates can be used to construct the proof. The algorithms used by the compiler to check the soundness of the proof and to generate the target language program are discussed too.

References

[1] C.A.R. Hoare. An axiomatic basis for computer programming, *Commun. ACM*, 12, 10, 1969, 0001-0782, 576–580, http://doi.acm.org/10.1145/363235.363259, ACM Press, New York, NY, USA.

[2] E.W. Dijkstra. A Discipline of Programming, *Prentice-Hall*, 1976, DIJ e 76:1 1.Ex,

[3] F. Kröger. Temporal Logic of Programs, *Springer*, 1987, Berlin, Heidelberg.

[4] J.M. Morris. A theoretical basis for stepwise refinement and the programming calculus, *Sci. Comput. Program.*, 9, 3, 1987, 0167-6423, 287–306, http://dx.doi.org/10.1016/0167-6423(87)90011-6, Elsevier North-Holland, Inc., Amsterdam, The Netherlands.

[5] K.M. Chandy and J. Misra. Parallel Program Design, A Foundation, *Addison-Wesley*, 1988.

[6] Z. Horváth. Párhuzamos programok relációs programozási modellje, ELTE, TTK, Informatika Doktori Iskola, 1996.

[7] Z. Horváth, T. Kozsik, and M. Tejfel. Proving Invariants of Functional Programs, *Proceedings of the Eighth Symposium on Programming Languages and Software Tools*, 115-127, 2003.

[8] C. Morgan. Programming from specifications, Second, *Prentice Hall International (UK) Ltd.*, 1994.

[9] J.-R. Abrial. The B-book: assigning programs to meanings, *Cambridge University Press*, New York, NY, USA, 1996, 0-521-49619-5.

[10] J.P. Bowen and M.G. Hinchey. Ten commandments revisited: a ten-year perspective on the industrial application of formal methods, *FMICS '05: Proceedings of the 10th international workshop on Formal methods for industrial critical systems*, 2005, 1-59593-148-1, 8–16, lLisbon, Portugal, http://doi.acm.org/10.1145/1081180.1081183, ACM Press, New York, NY, USA.

[11] D. Pavlovic and D. Smith. Software development by refinement, In *UNU/IIST 10th Anniversary Colloqium*, Formal Methods at the Crossroads: From Panaea to Foundational Support. Springer-Verlag, 2003., 2003,
`citeseer.ist.psu.edu/pavlovic03software.html`.

[12] J. McDonald and J. Anton. SPECWARE - Producing Software Correct by Construction, *Kestrel Institute Technical Report*, KES.U.01.3., March 2001., 2001.

[13] M. Czarnecki and U. Eisenecker. Generative Programming: Methods, Tools, and Applications, *Addison-Wesley*, CZA k 00:1 1.Ex, 2000.

Pliant Ranking

József Dombi and Norbert Győrbíró

Multi-criteria decision management applications are important tools for decision makers. One of the difficulties in decision management is the comparison of the alternative choices. The alternatives are often described with human words or given with fuzzy boundaries and as such cannot be ranked easily. Fuzzy theory [10] provides a mathematical foundation for modelling imprecise values and elusive human words [11]. There has already been a lot of effort to aid the decision process with fuzzy ranking algorithms.

One of the earliest fuzzy ranking function comes from Shimura and is based on a comparison method which is adopted for psychological test [9]. Buckley and Chanas [1, 2, 3] provided a fast ranking method using interval analysis. The ranking algorithm introduced by Cheng [4] is based on calculating dis- tances. Delgado et al. [5] gave an ordering procedure using fuzzy relations and fuzzy measures. The preference relation described by Kundu [8] utilizes a fuzzy leftness relation on intervals.

Despite the many different approaches, there is still no consensus which ranking method is the best suitable for applications. The problem arises from the requirements that the ranking algorithm should run fast and have all the properties that a ranking procedure over crisply defined objects has.

In this paper we propose a novel preference ranking method based on the pliant concept [7]. Pliant ranking provides a pairwise choice between two alternatives and can be calculated easily. Our algorithm models the various alternatives with fuzzy numbers and defines a preference relation over them. We prove that pliant ranking fulfills all the necessary properties associated with preference methods, i.e. shift invariance, transitivity and shift monotonicity.

References

[1] J.J. Buckley. Ranking alternatives using fuzzy numbers, *Fuzzy Sets and Systems*, 15 (1985) 21-31.

[2] J.J. Buckley. Fuzzy hierarchical analysis, *Fuzzy Sets and Systems*, 17 (1985) 233-247.

[3] J.J. Buckley and S. Chanas. A fast method of ranking alternatives using fuzzy numbers, *Fuzzy Sets and Systems*, 30 (1989) 337-339.

[4] C-H. Cheng. A new approach for ranking fuzzy numbers by distance method, *Fuzzy Sets and Systems*, 95 (1998) 307-317.

[5] M. Delgado, J.L. Verdegay, and M.A. Vila. A procedure for ranking fuzzy numbers using fuzzy relations, *Fuzzy Sets and Systems*, 26 (1988) 49-62.

[6] J. Dombi. General class of fuzzy operators, the DeMorgan class of fuzzy operators and fuzziness measures induced by fuzzy operators, *Fuzzy Sets and Systems*, 8 (1982) 149-168.

[7] J. Dombi. Fuzzy arithmetic operations on various inequalities: The pliant concept, manuscript, paper submitted to *Fuzzy Sets and Systems*.

[8] S. Kundu. Preference relation on fuzzy utilities based on fuzzy leftness relation on intervals, *Fuzzy Sets and Systems*, 97 (1998) 183-191.

[9] M. Shimura. Fuzzy sets concept in rank-ordering objects, *Journal of Mathematical Analysis and Applications*, 43 (1973) 717-733.

[10] L.A. Zadeh. Fuzzy Sets *Inf. Control*, 8 (1965) 338-353.

[11] L.A. Zadeh. A computational approach to linguistic quantifiers in natural language, *Comp. Math. Appl.*, 9 (1983) 149-184.

Calibration of CCD cameras for computer aided surgery

Krisztina Dombi

A navigation system is planned and implemented in order to use it in computer aided surgery. The idea is that three cameras are collecting images of the same object and from these projections the 3D coordinates of the points can be computed. In order to perform such a positioning we have to solve the calibration of the cameras. The calibration needs a special object, called calibration cross. The images of the calibrated cameras can be used later for determining point positions.

A calibration program has been developed which is able to the determined the precision of the calibrated Navigation System. Several test experiment have been performed in order to check the positioning of the system.

The experimental result shows that the positions of 3D points can be determined with an error cca. 0.3 cm. We work further for improving this result.

References

[1] General Elecctric Company. Image Guided Surgery.
http://www.gehealthcare.com/rad/savi/nav/igs.html, 2004.

[2] G. Fichtinger. Surgical Navigation, *Registration, and Tracking*. 2004.

[3] C.E. Thorpe, T. Kanade, K.D. Gremban. Geometric camera calibration using szstem of linear equations. *Robotics and Automation*, 1998 IEEE International Conference, 1988

[4] W.T. Vetterling, B.P. Falnnery, W.H. Press, and S.A. Teukolsky. Numerical Recepies in C. *Cambridge University Press*, 1992.

[5] Z. Zhang. A flexible new technique for camera calibration, *IEEE Transactions on PatternAnalzsis and Machine Intelligence*, 2000.

An Optimization Approach for Integrated Design and Parameter Estimation in Process Engineering

Jose A. Egea

Typical problems in (bio)chemical processes such as integrated design and parameter estimation can be stated as non-linear programming problems subject to non-linear differential constraints. These are frequently non-convex and/or ill-conditioned. Besides, the computational time associated to each function evaluation (i.e. simulation) can be considerably high. To surmount these disadvantages, the use of global optimization methods is proposed. In particular, those methods that use any sort of surrogate model instead of the original model are intended to provide good solutions in relatively short computation times. We will focus in a computationally expensive model to test the capabilities of some global optimization methods based in surrogate models which are currently being investigated in order to effectively apply them to costly process engineering problems.

Towards a unified model of computation with correctness proofs

Gáspár Erdélyi

In the past decades several programming models have been developed - each with a different approach and different focus. These models usually give the programmer the possibility of designing the program and proving its correctness at the same time. Each approach and each paradigm has its strong and weak sides (for example, a language manipulation tool is easily described by a grammar, a network protocol by Petri-nets for more usual problems structured programs are more suitable. There is no single paradigm which fits well all cases.

The programmer has to make the decision: either to use a single model or to design the whole program as a composition of different parts each of a different approach. The prior case he has to tackle with describing the whole program in the very same model which can turn very difficult at some parts and mess up perscipuity. At least correctness could be dealt by the usual proofs inside the model. The latter case he may choose whatever approach is suitable for each subproblem, however in the practice the possibility of a totally correct correctness proof is lost as integration is done by linking the object codes of the different parts together and the hiiker has almost nothing to do with correctness.

How couid we provide correctness proving methods for programs built up using many models? Either we could improve the compilers to issue certificates of the behaviour of the compiled modules and also the linker to investigate if the modules are integrated in a correct way, or we could try to bisimulate programs across models. This second alternative holds out the chance to reveal correlations between the different models that might help us improving our proof techniques - badly needed for improvement of the linker to integrate the proofs.

It's not an easy job to bisimulate programs of various models. The correlation between the models are often flimsy or at least hardly palpable. In the article we will review some key concepts of programming and design a model for simulating programs of other models. These key aspects include state of data or variables the program works on, and also the state of the program which tells us what could be done in the next execution step. Many models do not have both of them, either do not allow the use of variables of different types (grammars, finite state automatons for example) while others do not focus that much on state transitions (structured programs). We introduce a model where we can have both kinds of state separately which is very helpful in simulation. Then the operational semantics is specified so wc could reason about the correctness of the bisimulation of programs of various models such as structured programs, finite state automatons, Petri-nets, grammars. Finally we give some ideas on topics for further research.

References

[1] J. Baeten and W.P. Weijland. Process Algebra, *Cambridge University Press*, Cambridge, 1990.

[2] K.M. Chandy and J. Misra. Parallel Program Design: A Foundation, *Addison-Wesley*, Reading, USA, 1988.

[3] E.W. Dijsktra. A Discipline of programming, *Praetice-Hall*, 1976.

[4] W.H.J. Feijen and A.M.J. Gasteren. On a method on multiprogramming, *Spinger-Verlag*, New York, 1999 ISBN; 0-387-98870-X.

[5] R. Herken (editor). A half century survey on The Universal Turing Machine *Oxford University Press*, Oxford, 1988 ISBN: 0-19-853741-7.

[6] C.A.R. Hoare. Communicating sequential processes, *Prentice-Hall Int.*, Englewood Cliffs, 1985 ISBN: 0-13-153271-5.

[7] R.Y. Kain. Automata theory, *McGraw-Hill*, New York, 1972.

[8] H.R. Nielson and F. Nielson. Semantics with Applications: A Formai Introduction, *Wiley*, Chichester, 1992.

[9] S. Owiczki and D. Gries. An Axiomatic Proof technique for Parallel Programs, *Acta Informatica*, 6(4):319-340, 1976.

[10] J.L. Peterson. Petri Net Theory and the Modelling of Systems, *Pratice-Hall Int.*, Englewood Cliffs, 1981.

Implementing and comparing impact algorithms

Szabolcs Faragó

As software systems are growing, in parallel the code analysis techniques are spreading as well. The aim of these techniques is to help the work of software developers or testers. As it is hard to understand complex systems, the size of the software usually affects the length of testing.

The aim of impact analysis is to determine how the methods of a program affect other methods. Impact set for a method, or a method set includes the methods affected by the given method/method set. Impact analysis has two modes:

- static impact analysis

- dynamic impact analysis

Both methods have advantages and disadvantages.

Static impact analysis algorithms return with much precise result because these analyze the whole source code. Because of this, the static algorithms are slower than other techniques.

Dynamic impact analysis is a new research area. It determines the impact set to one execution of the program. This method is faster than static methods, but it is not as precise as the old analysis techniques.

The impact algorithms have been used in more areas, mostly in help for testing, but during developing as well.

During my PhD work I have worked in a project team, where we have researched static and dynamic techniques, and we have tested the precision and the speed of analysis algorithms. We have written implementations for static and dynamic algorithms in Java language. Additionally, we have executed several tests on these implementations, and have compared the results. The results have proved that static algorithms are much precise than dynamic algorithms but slower too.

In my presentation I would like to introduce the algorithms, which we have implemented and tested, and I would also like to present the results of tests.

Extensions and Applicability of the Membership Driven Reasoning Scheme

Zsolt Gera

In this paper we use sigmoid-like membership functions which take values in the open unit interval, and investigate the Membership Driven Inference (MDI) reasoning scheme. With sigmoid-like membership functions one can avoid the so-called indetermination part of the conclusion, which occur in reasoning with the Compositional Rule of Inference (CRI). Moreover, the MDI and the min and product based CRI are closed under such membership functions. The good axiomatic properties of the MDI reasoning scheme are shown, including not only the generalized modus ponens, but also the generalized modus tollens, the generalized chain rule, and more. As a special sigmoid-like function, we present the so-called squashing function by which piecewise-linear fuzzy intervals can be arbitrarily approximated. We show that by utilizing approximated fuzzy intervals in rules and premises, the MDI reasoning scheme can be efficiently calculated only by the parameters of the rule and premise memberships. In the second part of the paper we extend the original MDI reasoning scheme to multiple rules and multiple dimensions. We also investigate its applicability in fuzzy control i.e. with precise inputs and outputs.

References

[1] J.F. Baldwin. A new approach to approximate reasoning using a fuzzy logic. *Fuzzy Sets and Systems*, 2:309-325, 1979.

[2] T.C. Chang, K. Hasegawa, and C.W. Ibbs. The effects of membership function on fuzzy reasoning. *Fuzzy Sets and Systems*, 44:169-186, 1991.

[3] N.N. Morsi and A.A. Fahmy. On generalized modus ponens with multiple rules and a residuated implication. *Fuzzy Sets and Systems*, 129:267-274, 2002.

[4] B. Bouchon-Meunier, R. Mesiar, C. Marsala, and M. Rifqi. Compositional rule of inference as an analogical scheme. *Fuzzy Sets and Systems*, 138:53-65, 2003.

[5] J. Dombi and Zs. Gera. The approximation of piecewise linear membership functions and ALukasiewicz operators. *Fuzzy Sets and Systems*, 154:275-286, 2005.

An Agent-based Framework for Supporting e-Health

Pedone Gianfranco

The increasing level of mobility of patients across the European countries obliges governments to offer health services more and more flexible but at the same time uniform. If globally seen, patients could saturate the national health care services effort and increase related costs. The idea of the research project is to prospect the possibility of building an intelligent ICT (Information & Communication Technology) platform by which granting home care personal services close to the patient needs. The platform should: integrate information of different types and from different sources, be integrated with ICT whilst ensuring private and customized data access; use ontologies to define the profile of subjects and objects, have a mechanism to combine and refine the ontologies to personalize the system, incorporate knowledge from the different clinical sources, configure a knowledge-based decision support system that can supply e-Services to all subjects involved in the Home Care Model, extract evidence from real patients and integrate it with published evidence derived from real care treatments. The objective of the research project is to analyse how the agent-based technologies can supply the necessary intelligent effort to provide human-oriented services.

Objectives

Next to general objectives that directly come from the business context, such as generate a new ICT Sanitary Model, propose a knowledge-based platform for care services and integrate knowledge about HCPs (Home Care Patients) assistance, there are consistent and scientifically relevant goals to be achieved, such as: integrating different data sources, integrating information coming different EU members countries, defining ontologies for representing the necessary knowledge contained in all actors of the context, defining ontologies for representing pathologies, defining formal intervention plan, personalizing the access to the platform, personalizing the assistance by intelligent approaches.

Agents technology

The objective of project is to investigate how community care can be developed in the Internet age through the use of multi-agent technology. The motivation for this comes from the consideration of the agent society's social abilities in:

- Promoting effective care systems (provide better services and resources to clients, enhance social interaction between them, and with their carers, deliver more effective care)

- Providing the high-abstraction level care management strategies by linking all relevant agencies into a single framework of accountability;

- Giving an in-depth understanding of the health information framework that underpins the delivery of high quality, effective community care, including the formularisation of the links between the disparate agencies involved;

- Establishing a single agent-based care monitoring facility that can be used by all care professionals to assist in effective monitoring and diagnosis;

- Developing cooperative structures within the community structures to change service provision and care policies through the use of automated agents involvement in planning, scheduling, organizing;

- Devolving care management and responsibility and client centred environment that adapts rapidly to changing needs.

Modeling of grid monitoring systems

Gábor Gombás

Grid systems play an increasingly significant role in many scientific areas and are also being used in commercial systems. Grids provide enormous computing power for a relatively low price. The cost is the increased complexity. Different components of a grid system must be able to communicate with each other and must have up-to-date view of available resources. The need of interoperability between different components often results in "black box" designs where the internal working of components is completely hidden for others. While this makes building a grid easier it provides obvious difficulties for users and system administrators who want to diagnose problems on remote resources. The need for timely information in grids is therefore evident.

Information can have many properties. This research focuses on monitoring, where the basic charasteristics of information include limited validity period, low latency and usually small size of individual pieces of data.

Over the years several grid monitoring solutions were created. Some of them were designed to solve very specific problems only while others were meant to be generally usable. While many characteristics of existing monitoring systems differ, there are several features that all of them share. The Grid Monitoring Architecture (GMA, [1]) specification created by the Global Grid Forum tried to collect the commonalities of the existing systems and create a recommendation for the architecture of future systems. The main goal of the GMA was to help in creating interoparable monitoring systems, but unfortunately it did not contain enough specifics to be really useful for that.

The research presented here considers the same problem but from a different point of view. Instead of the informal nature of the GMA, a formal model of grid monitoring systems is described. The model is created by abstracting and formalising the common features of existing systems. The model gives a formal specification of the components involved in grid monitoring (traditionally called *producer*, *consumer* and *registry*) by utilising methodologies like abstract state machines and temporary logic. The required interactions between the monitoring components are also studied and a formal definition of a minimalistic producer-consumer protocol is given.

Real monitoring systems usually target specific application areas and therefore have specific extra requirements that are not neccessarily suitable for other monitoring systems. A few examples are provided to demonstrate how this formal model can be extended and refined to support such extra requirements while still maintaining the properties of the original model. The examples include the introduction of actuators and the modeling of aggregation services that can be used for building complex monitoring hierarchies.

In the future the model can be used both for the verification of existing systems as well as the base of new monitoring system designs. Using a formal approach may help system integrators to better understand the differences and similarities of existing systems.

References

[1] B. Tierney, R.A. Aydt, D. Gunter, W. Smith, V. Taylor, R. Wolski, and M. Swany. A grid monitoring architecture. *Informational Document* GFD-I.7, Global Grid Forum, January 2002.

Verification of Reconfiguration Mechanisms in Service-Oriented[1]

László Gönczy

As the automated integration of heterogeneous software environments becomes widespread, there is a growing demand for resilient software architectures. The Service-Oriented Architecture is an emerging paradigm in this field, however, in its present form, it does not cover the fault-tolerant aspects and no verifiable system reconfiguration mechanisms are modeled.

As the configuration of the underlying services may often change, (e.g. consider a network of mobile services), the system must be able to react to effects of the changes of the environment by means of *dynamic reconfiguration*. On the other hand, certain properties (such as the availability of the minimal number of service instances of a given service type, the presence of a particular service or some quantitative requirements) must be guaranteed during the entire lifetime of the system. Therefore, the aim of my ongoing research is twofold:

First, the *components* of a typical Service Oriented Architecture (services, ports, messages, etc.), their non-functional *properties* (such as the guaranteed response time or the acknowledgement options), the *reconfiguration* mechanisms (e.g. searching for and invoking a new service if a call fails or the response time decreases below a certain limit), and the *fault model* (e.g. a service crashes or becomes overloaded) are described in a high-level model while the underlying technology, for instance, SOAP messages or Grid technologies will remain hidden [1]. The basis of the followed approach is discussed in [2].

Second, I implemented a system for the *verification* of reconfiguration algorithms. Based on the high-level, technology-independent system description, i) the state space of the model will be investigated by *reachability analysis* to generate the set of the possible succeeding system configurations and ii) the fulfillment of the requirements against these configurations (and the transient states) are verified. During this analysis, a graph transformation tool with a model checking support can be used, like Groove [3]. The most important future research task is the investigation of the applicability model-based synthesis of Service-Oriented Architectures and service configurations.

References

[1] L. Gönczy and D. Varró. Modeling of Reliable Messaging in Service Oriented Architectures, In *Proc. Intern. Workshop on Web Service Modeling and Testing (WS-MATE 2006)*, To appear.

[2] L. Baresi, R. Heckel, S. Thöne, and D. Varró. Style-Based Modeling and Refinement of Service-Oriented Architectures - A graph transformation-based approach, To appear in *Software and Systems Modeling* - Springer-Verlag, 2006.

[3] Graphs for Object-Oriented Verification (GROOVE project),
http://groove.sourceforge.net/groove-index.html

Towards A Unified Stylesheet Format for Paper-Based and Electronic Documents

Hajnalka Hegedűs and László Csaba Lőrincz

Nowadays more and more originally paper-based documents appears on the Web (usually in HTML or XML format). New documents are often designed to be good quality on both media. In both cases the goal of having two instance of the document, the more alike, the better, but optimised for the specialities of the medium on which it is published. This is a tedious job, since the two medium (paper and web) has different possibilities and limitations, and authors or editors have to maintain two different stylesheets with a lot of redundant information..

To aid the work of authors writing documents for multiple media, a unified stylesheet language can be made. In this language authors can express their requirements both toward the paper-based and the electronic version of document. This can decrease redundancy, and maintenance is easier as well.

References

[1] D.E. Knuth. The TeXbook, *Computers and Typesetting*, 1986.

[2] L. Lamport. LaTeX: A Document Preparation System: User's Guide and Reference Manual, 1994.

[3] Sal Mangano, Simon St. Laurent, XSLT Cookbook, *O'Reilly & Associates, Inc.*, Sebastopol, CA, USA, 2002.

[4] T. Bray and J. Paoli, and C.M. Sperberg-McQueen. Extensible markup language, *World Wide Web J.*, 2, 4, 1997, 1085-2301, 29–66, O'Reilly & Associates, Inc., Sebastopol, CA, USA.

Automatic Generation of Compiled Model Transformations[1]

Ákos Horváth

Nowadays, Model Driven Architecture (MDA [1]) is an emerging paradigm in software development. Based on high-level modelling standards (e.g. UML), MDA separates business and application logic from underlying platform technology by using platform independent models (PIM) to capture the core functionality of the target system, and platform specific models (PSM) to specify the target system on the implementation platforms (Java, C#, C++). PSMs and platform-specific source codes are automatically generated from PIM and PSMs, respectively, by using model transformation (MT) methods, thus, the role of MT is unquestionable for the success of the overall process of MDA.

Models in MDA have a graph structure, so model transformations can be specified by graph transformation (GT), which is a declarative and rule based specification paradigm. The execution of a GT rule performs local manipulation on graph models by finding a matching of the pattern prescribed by its left-hand side graph in the model, and changing it according to the right-hand side graph. In order to be able to handle complex model transformations, control structures embedding GT rule applications are also needed.

As the complexity of model transformations is growing, a new demand has been arisen to separate the design from transformation execution by using high-level models at design time and by automatically deriving source code for the target platform from these high-level models. While high-level models ease the development, testing, debugging and validation of model transformations within a single transformation framework without relying on a highly optimized target transformation technology, compiled standalone versions of a model transformation in an underlying platform (e.g. Java, C#) are more efficient from runtime performance aspects.

Code generators that are used in the process of standalone version generation, are typically implemented in a standard programming language for specific model transformations, thus, it is difficult to reuse existing source code generators to different platforms with conceptual similarities (e.g. OO based programming languages) or to modify the output source code generation in order to integrate it to other MT tools.

The current paper presents a new approach using high order (generic and meta [2]) graph transformation rules for the source code generation from high-level model transformation specifications defined by a combination of graph transformations and abstract state machine (ASM) constructs (as used within the VIATRA2 framework [3], which is a general Eclipse based modelling framework tool having been developed at the TUB-DMIS). The essence of the approach is to store model transformation rules as ordinary models in the model space. This way the source code generator of the transformations can be handled within the modelling framework. As a result, the code generator can be reused by replacing only the output generation rules in order to port the transformations to new underlying platforms, and the correctness of the code generators can be validated by a huge range of graph algorithms (e.g mutant graphs).

References

[1] Model Driven Architecture — A Technical Perspective, *Object Management Group* (`http://www.omg.org`), September 2001.

[2] D. Varró and A. Pataricza. Generic and meta-transformations for model transformation engineering, in: *T. Baar, A. Strohmeier, A. Moreira and S. Mellor, editors, Proc. UML 2004: 7th International Conference on the Unified Modeling Language*, LNCS 3273 (2004), pp. 290–304.

[3] VIATRA2 Framework,
An Eclipse GMT Subproject (`http://www.eclipse.org/gmt/`).

Comparison of static- and dynamic call graphs

Endre Horváth and Judit Jász

The construction of the call graph is one of the most important tasks during the static analysis of a program. In object-oriented languages, like C++, the presence of pointers and polymorphism can make the results of the analysis imprecise, and the efficiency of the tool or algorithms that use these results can significantly drop.

To make the call graph more precise we have to refine the results from our static analysis. For example we can use some well known algorithms, like CHA or RTA, to resolve the virtual function calls in the program, but in most cases even these algorithms will not give us the most precise call graph. Also we could use a precise pointer analysis, to handle the use of pointers, but in case of large programs, a context- and flow sensitive points-to algorithm requires a large amount of storage space and running time. Generally we can use these accurate pointer analyses only on small systems.

In many cases, tough, it is enough to calculate the call graph from the execution of the program for a set of test cases. This way we do not have to compute the precise call graph, but we can get a set of dynamic call graphs to approximate their static counterpart. We have to notice that the cost of building a dynamic call graph greatly increases with the size of the analyzed program. So we have to find the tools that can build the static and dynamic call graphs of a large system under acceptable circumstances.

We use a program analysis tool for C/C++, called Columbus [1]. It is capable of analyzing real-life, large software products and building the structure of the program in the form of an abstract syntax tree. With the use of Columbus we have different algorithms to compute the static call graph of such programs. We have developed a tool as part of Columbus what can instrument the source code of the analyzed programs, and can make the programs capable of generating their own dynamic call graph. Having run the code, what was instrumented this way, we can get the set of dynamic call graphs and we can use these graphs to approximate the static call graph for the whole system.

We have analyzed some large, real-life software, and have made their static and dynamic call graphs. We have compared these graphs, and have observed whether the static call graphs can be approximately calculated using their dynamic counterparts.

References

[1] R. Ferenc and Á. Beszédes, M. Tarkiainen, and T. Gyimóthy. Columbus – Reverse Engineering Tool and Schema for C++, *Proceedings of the 18th International Conference on Software Maintenance (ICSM 2002)*, IEEE Computer Society.

Design and reengineer metadata with the help of business process modelling

Gábor Horváth

We discussed the major issues of Metadata Repositories from a business process modelling point of view earlier [1]. The "Managed Metadata Environment" approach as an important mind shift supports the modelling framework we suggested.

Here we would like to investigate design and redesign issues related to the content of Metadata Repositories, namely metadata itself. Using the goal-oriented process modelling framework, teleonics, we discuss the why and what behind the main processes of the metadata lifecycle: design, collection, storage, maintenance and application. Industrial best practices are also visited and integrated into the language and view of the teleonic approach, in order to prove the usefulness of the application of a general system thinking methodology to our study.

Our main area of interest, international statistical data and metadata exchange programs provide an excellent opportunity to examine the upcoming international standard of the field: SDMX [2], which is supported by the major European and international organisations. The current purpose of the research is to complete the teleonic model of the exchange of time series data and their corresponding metadata based on SDMX. This includes the study of the key teleonic views: environment, goals, processes plus the data involved in the data exchange. An example of challenging and changing currently supported metadata will be discussed by the model together with the details of possible future solutions.

References

[1] G. Horvath. Teleonics in BP modelling and IT system design, accepted for publication, 2005

[2] SDMX - Statistical Data and Metadata Exchange, http://www.sdmx.org

Refactoring Erlang Programs[1]

Zoltán Horváth, László Lövei, Tamás Kozsik, Anikó Víg, and Tamás Nagy

We present here the prototype of a refactoring toolset for Erlang programs where one can incrementally carry out programmer-guided meaning-preserving program transformations. We discuss an approach to the problems of storing and extracting the syntactic and also the static semantic information in order to be flexible enough to perform the desired transformations. In our approach the program to be redesigned is stored in a relational database. The Erlang ODBC interface is used to access the database. The refactoring tool is integrated into a development environment using Emacs as an interface. The refactoring steps are implemented in Erlang, using the standard Erlang parser to construct the database from the source code. This backend is connected to Emacs through Distel. The tool has two different modes. In the first mode, the programmer can choose from the safe refactor steps, which results an other safe position. Editing the source code is prohibited in this mode, which eliminates the need for reparsing and rebuilding the database. In the second mode, editing is possible in two ways: the programmer can choose from menupoints like "insert a function", which is partly controlled, or can edit the source code freely. Further refactor steps need syntactically correct source code.

References

[1] H. Li, C. Reinke, and S. Thompson. Tool Support for Refactoring Functional Programs, *Haskell Workshop: Proceedings of the ACM SIGPLAN workshop on Haskell*, Uppsala, Sweden, Pages: 27–38, 2003.

[2] Á. Fóthi, Z. Horváth, and J. Nyéky-Gaizler. A Relational Model of Transformation in Programming, *Proceedings of the 3rd International Conference on Applied Informatics*, Eger-Noszvaj, Hungary, Aug. 26-28, 1997. 335-349.

[3] M. Fowler, K. Beck, J. Brant, W. Opdyke, and D. Roberts. Refactoring: Improving the Design of Existing Code, *Addison-Wesley*, 1999.

[4] *Martin Fowler's refactoring site*, `http://www.refactoring.com`

[5] R. Szabó-Nacsa, P. Diviánszky, and Z. Horváth. Prototype Environment for Refactoring Clean programs, *CSCS 2004, The Fourth Conference of PhD Students in Computer Science Szeged*, Hungary, July 1-4, 2004.

[6] Distel: Distributed Emacs Lisp, `http://fresh.homeunix.net/~luke/distel`

[7] J. Armstrong, R. Virding, M. Williams, C. Wikstrom. Concurrent Programming in Erlang, *Prentice Hall*, 1996

A Novel Performance Model of J2EE Web Applications Considering Application Server Settings

Gábor Imre and Hassan Charaf

Managing business processes, the improper performance of a web application can cause serious financial loss to a company. The performance-related requirements of an Internet application are often recorded in a Service Level Agreement (SLA). SLAs can specify an upper limit for the average response time, a lower limit for availability, while the application guarantees a certain throughput level.

These performance metrics depend on several factors, such as hardware, software, network, and client workload. This paper focuses on the settings of the application server software that serves the HTTP requests of the browsers. More precisely, the performance of a test web application is measured under different client load with different values of two parameters of the application server. These tuning parameters are the maximum size of the thread pool, and the maximum size of the HTTP connection queue. In the application server, accepted HTTP connections are placed into a connection queue. The size of the connection queue is limited by an adjustable parameter of the given application server. When this limit is reached, it refuses to serve the request. The threads in the thread pool take connections from the queue and serve the requests. The server can decide to create more threads (i. e. increase the size of the thread pool), but cannot exceed a certain configurable maximum. When the maximum thread pool size is reached, however, the requests are not dropped, as long as they find free space in the connection queue.

Setting up a performance model that is capable of creating a quantitative relationship between the performance factors and the performance is a key issue to meet the performance requirements of SLAs. Using queueing networks is a popular method of performance modelling. In [1], a queueing model for multi-tier internet applications is presented, that faithfully captures concurrency limits at the tiers. The maximum size of the connection queue, as presented earlier, can be considered as a concurrency limit of the web tier in this model, but it cannot handle the maximum size of the thread pool. A powerful combination of the queueing network and the Petri net formalism is presented in [2]. Using queueing Petri nets, the authors successfully model the performance of a web application, considering the maximum size of thread pools. Their model, however, does not take the maximum size of the connection queue into account.

Our paper shows that the limits configured both for the connection queue and the thread pool have a considerable effect on the performance and presents a queueing network based performance model that considers both of them.

Keywords: *performance evaluation, Web technologies, queueing networks*

References

[1] B. Urgaonkar, G. Pacifici, P. Shenoy, M. Spreitzer and A. Tantawi. An Analytical Model for Multi-Tier Internet Services and Its Applications, *ACM SIGMETRICS Performance Evaluation Review*, 33(1), 2005, pp. 291-302.

[2] S. Kounev, and A. Buchmann. Performance Modelling of Distributed E-Business Applications Using Queueing Petri Nets, in *Proceedings of IEEE International Symposium on Performance Analysis of Systems and Software*, Austin, Texas, 2003, pp. 143-155.

Grid Meta-Broker Architecture: Requirements of an Interoperable Grid Resource Brokering Service

Attila Kertész and Péter Kacsuk

The Grid was originally proposed as a global computational infrastructure to solve grand-challenge, computational intensive problems that cannot be handled within reasonable time even with state of the art supercomputers and computer clusters [1]. Grids can be realized relatively easily by building a uniform middleware layer, on top of the hardware and software resources, the programming concept of such distributed systems is not obvious. To enhance the manageability of grid resources and users Virtual Organizations were founded. This kind of grouping started an isolation process in grid development, too. Interoperability among these "islands" will play an important role in grid research.

As resource management is a key component of grid middlewares, many solutions have been developed. After examining the existing resource brokers we created a taxonomy, which helps identifying the relevant properties of these brokers. Utilizing the existing, widely used and reliable resource brokers and managing interoperability among them could be new point of view in resource management. This paper introduces an abstract architecture of a Meta-Broker that enables the users to access resources of different grids through their own brokers. Designing such architecture the following guidelines are essential: As standards play an important role of today's grid development, the interfaces must provide standard access. The architecture must be "plug-in based" - the components should be easily extended by all means. The main components of the system: the *Converter* is responsible for translating the user requests to the language of the appropriate broker that the Meta-Broker wants to invoke; the *Information System* stores the properties of the reachable brokers and historical data of the previous submissions; the *Matchmaker* selects the proper broker for a user request; the *Submitter* communicates with the connected brokers, invokes them with a job request and collects the results. The job description contains the user request, and the Information System provides the broker information needed for the Meta-Broker to decide where to submit the job. The interconnected brokers' tasks are to perform the actual job submissions; to find the best resource within their scopes, i.e. the VOs they have access to. The Meta-Broker only needs to communicate with them. In this sense meta-brokering stands for brokering over resource brokers instead of resources.

Some related works also deal with interoperability: The Grid Interoperability Project [2] has some results on resource brokering between Unicore [4] and Globus [5] Grids. The goal of their work was to create a semantic matching of the resource descriptions, but their ontological mappings specialize only in these two middlewares. The Gridbus Grid Service Broker [3] is designed for computational and data-grid applications and supports all Globus middlewares and Unicore in experimental phase. Both solutions aim at accessing resources from different grids, but their architecture stays on the level of direct resource brokering.

Grid portals give a user friendly access to grid resources and other grid services. Using a Web-based portal, the user can submit a job easily, regardless of location. The P-GRADE Portal [6] is a workflow-oriented, multi-grid portal that provides all the functions needed for job submission. P-GRADE portal is already connected to different grids and brokers. Integrating the Meta-Broker to this portal provides the next step supporting interoperability in grids.

The introduced meta-brokering approach opens a new way for interoperability support. The design and the abstract architecture of the Grid Meta-Broker follow the latest results and standards in grid computing. This architecture enables a higher level brokering called meta-brokering by utilizing resource brokers for different middlewares. This service can act as a bridge among the separated "islands" of the current grids, therefore it enables more beneficial resource utilization and collaboration.

References

[1] I. Foster and C. Kesselman. Computational Grids, The Grid: Blueprint for a New Computing Infrastructure, *Morgan Kaufmann*, 1998. pp. 15-52.

[2] John Brooke, Donal Fellows, Kevin Garwood, and Carole Goble. Semantic Matching of Grid Resource Descriptions, *Lecture Notes in Computer Science*, Volume 3165, Jan 2004, pp. 240 – 249.

[3] S. Venugopal, R. Buyya, and Lyle Winton. A Grid Service Broker for Scheduling e-Science Applications on Global Data Grids, *Journal of Concurrency and Computation: Practice and Experience*, Wiley Press, USA (accepted in Jan. 2005).

[4] D.W. Erwin and D.F. Snelling. UNICORE: A Grid Computing Environment, In *Lecture Notes in Computer Science*, volume 2150, Springer, 2001, pp. 825 – 834.

[5] I. Foster and C. Kesselman. The Globus project: A status report, in *Proc. of the Heterogeneous Computing Workshop*, IEEE Computer Society Press, 1998, pp. 4-18.

[6] Cs. Németh, G. Dózsa, R. Lovas, and P. Kacsuk. The P-GRADE Grid Portal, *Lecture Notes in Computer Science*, Volume 3044, Jan 2004, pp. 10-19.

Numerical simulations of stochastic electrical circuits using C#

Edita Koláŕová

Modelling of physical systems by ordinary differential equations ignores stochastic effects. By incorporating random elements into the differential equations, a system of stochastic differential equations (SDEs) arises.

A general N-dimensional SDE can be written in vector form as

$$\mathrm{d}\mathbf{X}(t) = \mathbf{A}(t, \mathbf{X}(t))\,\mathrm{d}t + \sum_{j=1}^{M} \mathbf{B}^j(t, \mathbf{X}(t))\,\mathrm{d}W^j(t),$$

where $\mathbf{A} : \langle 0, T\rangle \times \mathbb{R}^N \to \mathbb{R}^N$, $\mathbf{B} : \langle 0, T\rangle \times \mathbb{R}^N \to \mathbb{R}^{N \times M}$ are functions and $W^1(t), \ldots, W^M(t)$ are independent Wiener processes representing the noise. (A process $W(t)$ is called the Wiener process if it has independent increments, $W(0) = 0$ and $W(t) - W(s)$ distributed $N(0, t-s)$.) The solution is a stochastic vector process $\mathbf{X}(t) = (X^1(t), \ldots, X^N(t))$.

By an SDE we understand in fact an integral equation

$$\mathbf{X}(t) = \mathbf{X}_0 + \int_{t_0}^{t} \mathbf{A}(s, \mathbf{X}(s))\,\mathrm{d}s + \sum_{j=1}^{M} \int_{t_0}^{t} \mathbf{B}^j(s, \mathbf{X}(s))\,\mathrm{d}W^j(s),$$

where the integral with respect to $\mathrm{d}s$ is the Lebesgue integral and the integrals with respect to $\mathrm{d}W^j(s)$ are stochastic integrals, called the Itô integrals (see [1]). Although the Itô integral has some very convenient properties, the usual chain rule of classical calculus doesn't hold. Instead, the appropriate stochastic chain rule, known as Itô formula, contains an additional term, which, roughly speaking, is due to the fact that the stochastic differential $(\mathrm{d}W(t))^2$ is equal to $\mathrm{d}t$ in the mean square sense, i.e. $E[(\mathrm{d}W(t))^2] = \mathrm{d}t$, so the second order term in $\mathrm{d}W(t)$ should really appear as a first order term in $\mathrm{d}t$.

We present an application of the Itô stochastic calculus to the problem of modelling inductor-resistor electrical circuits. The electrical current $i(t)$ at time t in a simple RL electrical circuit satisfies the differential equation

$$L\,\frac{\mathrm{d}\,i(t)}{\mathrm{d}t} + R\,i(t) = u(t), \quad i(0) = i_0,$$

where the resistance R and the inductance L are constants and $u(t)$ denotes the potential source at time t (see [2]). Now we allow some randomness in the electrical source as well as in the resistance. The SDE describing this situation is

$$\mathrm{d}I(t) = \frac{1}{L}(u(t) - RI(t))\,\mathrm{d}t - \frac{\alpha}{L}\,I(t)\,\mathrm{d}W_1(t) + \frac{\beta}{L}\,\mathrm{d}W_2(t), \quad I(0) = I_0,$$

where α and β are non negative constants. Their magnitudes determine the deviation of the stochastic case from the deterministic one. We consider both the initial condition and the current at time t as random variables and denote them by capital letters. We present the analytical solution of this SDE and also show the expectation and the second moment equation of $I(t)$.

In order to simulate $I(t)$ numerical techniques have to be used (see [3]). To generate numerical solutions and their graphical representations we use the programming language C# (see [4]), which is a part of the new MS .NET platform. To make matrix manipulation and visualization simpler we use the component library LinAlg described in [5]. LinAlg

is a set of classes that enables vectorial programming and incorporates a wide range of numerical, statistical and graphical methods. We also compute the confidence intervals for the trajectories of the solution. The results were verified in an experiment by measurements on inductor-resistor electrical circuits.

References

[1] B. Øksendal. Stochastic Differential Equations, An Introduction with Applications, *Springer-Verlag*, 2000.

[2] D. Halliday, R. Resnick, and J. Walker. Fundamentals of Physics, *John Wiley & Sons*, 1997.

[3] P. Kloeden, E. Platen, and H. Schurz. Numerical Solution Of SDE Through Computer Experiments, *Springer-Verlag*, 1997.

[4] Cs. Török et al. Professional Windows GUI Programming: Using C#, *Chicago: Wrox Press Inc*, 2002, ISBN 1-861007-66-3

[5] Cs. Török. Visualization and Data Analysis in the MS .NET Framework, *Communication of JINR*, Dubna, 2004, E10-2004-136

Formal analysis of existing checkpointing systems and introduction of a novel approach

József Kovács

One of the important goals of the Grid is to provide a dynamic collection of resources from which applications requiring very large computational power can select. In that sense Grid is a natural extension of the concepts of supercomputers and clusters towards a distributed and dynamic parallel program execution platform and infrastructure.

Due to the huge number of collaborating execution components the Grid is an inherently dynamic and un-reliable execution environment and hence this environment provides an error-prone system. Any component can fail at any time which may lead to abnormal termination or erroneous result. A parallel application containing communicating processes running on a cluster containing nodes is endangered on a bigger degree.

There are several methods and systems providing rollback-recovery support for parallel applications to survive any failure related to infrastructure. These systems are designed in different ways to support different goals. Most of them are lying on a checkpoint mechanism. Checkpoint means to save and restore the overall state of a distributed application where each surrounding component holding any state must be accessed and retrieved.

Checkpoint can be implemented at different levels (kernel, user, application), different ways (coordinated, uncoordinated, etc.) and in different components on a cluster (application, message-passing environment, scheduler, os, etc.). Each system providing „checkpoint" support uses different combination of these options.

The aim of this research is to introduce a formal framework in which any checkpoint system can be modelled. The model is focusing on the architectural components of the software and hardware environment of a cluster. The basic components are application, process, cluster and node. Relations and functions are defined among them according to the principles of Abstract State Machines that emphasize the main important features of the key components related to the checkpoint systems.

In this work the basic checkpoint methods are formalised. It is shown that existing checkpoint system can be defined in this environment and the different systems can be compared to each other while architectural classification of these systems are defined.

Based on the comparison a novel checkpoint approach is defined in this model. The approach aims at defining rules in a way that the resulting checkpoint system does not require any cluster specific checkpoint environment to serve the application in creating and restoring checkpoint information. Checkpoint and resumption of a distributed application can be realised among different clusters without any checkpoint-related restrictions regarding the software environment. This method is able to provide an environment independent checkpoint system combined with unmodified user code. No existing system can give this degree of freedom for cluster developers and programmers at the same time.

Simulation and Formal Analysis of Workflow Models Using Model Transformations[1]

Máté Kovács

Today major organizations tend to use more and more IT infrastructure that include the use of workflow execution engines as well. The *workflow* coordinates the activities inside the company, this way it becomes crucial concerning the correct functionality.

There are several high level executable languages for workflow implementation for example: BPEL (Business Process Execution Language). However, the testing of such workflows is problematic. The engine may be in contact with several independent databases that may be in the possession of different organizations. The rollback of the effects of each test run and the establishment of a test environment with all the data is equally expensive. My proposal aims to solve this problem with the formal model checking of workflows.

The BPEL is a semiformal notation. In order to examine its properties with mathematical accuracy, a workflow has to be transformed into an exact formalism. There are several approaches that use Petri nets or nondeterministic automata for this purpose.

I chose *dataflow networks* as the model of workflows. Dataflow networks consist of data processing nodes that are finite state automata interconnected with data conveyer channels. The formalism of dataflow networks gives a *verifiable semantics* to workflows.

In my work [1] I discuss a method, which allows us to carry out the formal analysis of a workflow implemented in BPEL without human intervention. This method relies on graph-transformations that are executed by the Viatra (Visual Automated Model Transformations). The first step (i) is to transform the BPEL model into a dataflow network model, which can be done in a deterministic way, due to the properties of the workflow description. The second transformation (ii) maps the dataflow into a Promela (Process Meta Language) model. The third (iii) step is to automatically generate the real Pomela program. The requirements set against the Promela code have to be formulated as Linear Temporal Logical expressions (LTL) that can be evaluated. This way the dynamic properties of a BPEL model can be formally checked.

The dataflow network model of workflows makes fault simulation also possible. With a small extension of the BPEL language *fault injector activities* are introduced. LTL expressions may be evaluated questioning whether the fault of one activity affects the other. This way the *error confinement region* of a faulty activity can be determined, and it can be verified whether a possible planned redundancy in the workflow model reached its goal.

There are general properties that a workflow definitely has to meet. The concrete LTL expressions, which are specific to the model, can be generated from the general requirements.

If the evaluation of a specific LTL expression results in failure, the sequence of events leading to the counterexample is also presented. The *back annotation* of such event sequence to the BPEL domain is a future goal.

References

[1] M. Kovács and L. Gönczy. Simulation and Formal Analysis of Workflow Models, In *Proceedings of 5th International Workshop on Graph Transformation and Visual Modeling Techniques* (GT-VMT), ENTCS, pp. 215-225, Vienna, Austria, 2006.

Optical delay buffer optimization in packet switched optical network

Miklós Kozlovszky and Tibor Berceli

Nowadays the delay sizing and optimization within all-optical networks is one of the most crucial points in the field of optical network communication. The information in al-optical networks is transformed to light and transmitted with the help of optical waves. We are currently investigating optical buffers that could meet the requirements of high speed/low latency communication within these networks. The information in such systems is traveling with the speed of light, but in the active nodes it is very likely experiencing delay, due traffic congestion, path switching or just because of hardware processing time. During delay in the network node, the information can not be stored and restored easily as we can do in „normal" electrical networks. During operation, light signal can not be stopped, and the electrical-optical (OEO) conversion takes huge amount of time comparing with the communication speed in such networks. As a result, the information should remain in form of optical waves all the time. To artificially introduce delay for the information in the network node we could use optical delay buffers. We can define the optical buffer as a device with both its input and output data streams in optical format without optical-electrical-optical conversion.

In the first section of the paper we give a short overview how to introduce time delays in the network nodes, what kind of optical delay elements are available and what are their basic characteristics. After that we focus on one widely used solution, namely the usage of the simple and cost effective optical delay line. The simplest implementation of an optical delay line is introducing a physical distance, such as a length of optical fiber, or free space. Although simple this implementation suffers from two drawbacks; significant tenability is difficult to achieve, and long time-delays require large physical distances. In the paper we will describe fully the basic symptoms of the usage of optical delay lines.

In the next section we compare the „normal" electrical delay buffers to optical delay buffers. Show the differences in latency and throughput and usable buffer sizes. With the help our investigation we introduce and build up an optical delay buffer model which can help hardware manufacturer to make accurate sizing and optimization of the optical buffer according to the system needs. According to our theoretical approach and simulations we optimize multistage optical delay line structures, to handle delay problems with multiple in/output channels. In the next section we give formal description and functional model of the proposed optical delay buffer.

We have carried out various simulations with network simulation software, to analyze the proposed delay buffer model and to make the functional analysis of such system. To search for the highest throughput, and minimize latency during network communication, we have worked with single and multistage optical delay buffer simulations as well. The major topics investigated include: latency evolution during single and multistage buffer usage, optimal buffer sizes for different data communication speed and variable packet sizes. In the last section of the paper we discuss the performance results of simulations and draw conclusions on our optical delay buffer model.

Keywords: *all-optical network, optical delay line, modeling, optical buffer*

A linear programming background for the R_{HFF}^{∞} upper bound proof

Máté Labádi

In the last 30 years intensive research was carried on the behavior of different kinds of bin-packing (BP) algorithms, where the goal is to put different tall items into the minimum number of unit height bins. There are several versions of BP algorithms, like on-line, off-line, d-dimensional and k-repacking problem.

In the on-line BP problem you have to put the items one by one into their final position without knowing the size of the remaining items. In the d-dimensional BP problem the items have d dimensions, and you have to pack them into d-dimensional one unit „width" bins.

Our research focused on the worst case scenario of the BP algorithms. The asymptotic worst case ratio of algorithm A (R_A^{∞}) shows us, for any input series of items, how many times more bins are needed in the worst case to pack the items with algorithm A, comparing to the optimum packing requires.

In 1982 Chung, Garey and Johnson introduced their off-line two-dimensional bin-packing algorithm, called the Hybrid First Fit (HFF)[1] algorithm. In their work the authors proved the lower bound of R_{HFF}^{∞} to be at least $\frac{182}{90}$ by giving a special input series of items. They also proved the upper bound to be at most ($\frac{191.25}{90}$). We expect the exact value of R_{HFF}^{∞} to be somewhat less than $\frac{187}{90}$.

In the upper bound proof the authors used a „horizontal lines through the bins" method, where they had 5 item classes and they crossed the bins with lines at different positions and counted up all the items more times. Unfortunately in [1] there is no word about how the authors found the proper number and position of the crossing lines.

We have developed two Linear Programming (LP) models, where solving these LPs give the number and position of the crossig lines and we get exact the same numbers that Chung, Garey and Johnson found in their work. The question is if we can use this LP model to generate the number and position of the lines, than are we able to improve the R_{HFF}^{∞} by using more and more item classes?

References

[1] F. R. K. Chung, M. R. Garey, and D. S. Johnson. On packing two-dimensional bins. *SIAM Journal on Algebraic and Discrete Methods*, 3:66–76, 1982.

[2] M. R. Garey, R. L. Graham, D. S. Johnson, and A. C. Yao. Resource constrained scheduling as generalized bin packing. *J. Combinatorial Theory Ser. A*, 21:257–298, 1976.

[3] D. S. Johnson. Near-optimal bin packing algorithms. *PhD thesis, Massachusetts Institute of Technology*, Cambridge, Mass., 1973.

[4] D. S. Johnson, A. Demers, J. D. Ullman, M. R. Garey, and R. L. Graham. Worst-case performance bounds for simple one-dimensional packing algorithms. *SIAM Journal on Computing*, 3:299–325, 1974.

Aspect-UML-Driven Model-Based Software Development

László Lengyel, Tihamér Levendovszky, and Hassan Charaf

Visual Modeling and Transformation System (VMTS) is an n-layer metamodeling environment which supports editing models according to their metamodels, and allows specifying OCL constraints. Models and transformation steps are formalized as directed, labeled graphs. VMTS uses a simplified class diagram for its root metamodel ("visual vocabulary"). Also, VMTS is a UML-based model transformation system, which transforms models using graph rewriting techniques. Moreover, the tool facilitates the verification of the constraints specified in the transformation steps during the model transformation process.

Many concerns pertaining to software development have a crosscutting impact on a system. Using current technologies (Unified Modeling Language and Object-Oriented Programming), such kinds of concerns are difficult to identify, understand, and modularize at design and implementation time, as they cut across the boundaries of many components of a system. Crosscutting concerns typically include design constraints and features, as well as architectural qualities and system-level properties or behaviors, such as transactions, logging and error recovery.

Aspect-oriented (AO) techniques are popular today for addressing crosscutting concerns in software development. Aspect-oriented software development (AOSD) methods enable the modularization of crosscutting concerns within software systems. AOSD techniques and tools, applied at all stages of the software lifecycle, are changing the way in which software is developed in various application domains, ranging from enterprise to embedded systems.

Aspect-oriented modeling (AOM) is a critical part of AOSD that focuses on techniques for identifying, analyzing, managing and representing crosscutting concerns in software design. The design models consist of a set of aspects and a primary model that can be weaved together by AOM weaving. Most current implementations of aspect-oriented programming (AOP) start directly from programming level. First, aspects are identified either by source code reading or document reading. Second, the system will be implemented in the code level based on the current defined aspects.

To meet the limitation of AOP we combine AOP and AOM to utilize the strength of both approaches. AOM is used in the requirement and design phase to ensure a proper aspect-oriented design. It can test any conflicting situations and enhance optimization between aspects.

Some AOM researchers propose that the aspects and the primary model are woven together in a way that the aspect is integrated into the primary model before the code generation. However, a major problem of this solution is that the separation of the aspects and the primary model is lost once the weaving is done. In the VMTS approach, model transformations are used to generate CodeDOM tree from models. The CodeDOM tree is a language-independent model representation of the source code, from which the code is generated automatically. Therefore, in VMTS, the CodeDOM tree corresponds to the code phase. In our process, we propose to keep the separation until the code phase. AOM weaving is only used to test the validity of the aspect models. In this way, we can keep the primary model and aspect models separately in the coding phase. It is then easy to backtrack if there are any changes in the CodeDOM tree since the generated primary code can be mapped directly back to the primary and aspect models. Current work introduces the methods required to realize the approach: (i) VMTS Aspect-UML and (ii) a weaving method that composes aspect models and primary model as well as CodeDOM models. An extended Aspect-UML is needed to express aspect models. Pointcuts, joinpoints and advice cannot be expressed exactly using current UML tools. As aspects are context-specific and need initialize, Aspect- UML needs to be able to customize the aspect model based on different contexts. Another requirement for Aspect-UML is that the connections between aspects and classes should be detailed enough to make the tracking easy. This requires that both the advice and pointcuts are represented in the aspect model. Aspects should be handled on two

levels: both at the modeling level (AOM) and the programming language level (AOP). At the AOM level, aspects are identified and woven together by AOM weaving to verify and optimize them. However, models are not woven together for the purpose of code generation based on a combined model. The actual weaving is performed on CodeDOM models.

Fitting the statistical module of the adaptive grid scheduler to the data of the NIIF

Ágnes Lipovits, Előd Kovács, and Zoltán Juhász

Distributed computer systems are becoming prominent in various application areas. An important requirement is to create such large, geographically distributed systems (grids) whose elements are connected by a wide-area network, heterogeneous in every aspect, yet provide a traditional desktop environment for users. These systems must support various tasks, such as execution of computation intensive and data intensive applications, supporting distributed collaboration and providing problem solving environments, whose resource demand and execution characteristics are very different.

There are many unsolved problems in large distributed heterogeneous environments. One of the most important ones is grid scheduling. It is not enough to know the hardware parameters of the resources for the best scheduling; we also need to estimate the overall run-time (waiting + wallclock time) of a job of a given queue. Job execution is a stochastic process; therefore the grid scheduler needs to be an adaptive system. The data processing module of the adaptive grid scheduler estimates the distribution of the typical run-time of the system as a random variable, after this it identifies the parameters of the given distribution. As the procedure requires a large database, for being efficient the purpose is to solve this part of the task with process migration and to send the output parameters towards the statistical module with minimal network overload. Another input parameter of the statistical module is the length of the queue. The overall run time as a random variable is the sum of independent random variables with the same distribution and parameters. ($\eta_k = \xi_1 + \xi_2 + \cdots + \xi_k$, where k is the length of the queue) Calculating the distribution of η_k in some particular cases - e.g. when the distribution of ξ_i -s being exponential - is not so difficult. In case of more complicated distributions the statistical module can only identity some characteristic values of the distribution. The expected value is for example $E(\eta_k) = k \cdot E(\xi_1)$. The estimation of the quintiles will be done by simulations.

In the presentation we will provide an overview of the architecture and the operating mechanism of the planned grid scheduler. We will present the distribution and parameters that fit the best the run-time as a random variable calculated from the job logs of the NIIF supercomputer. Then we characterize the calculated lengths of the job queues. Using the results we compare the calculated values of our model with the real run-times.

Comparing Specification with Proved Properties of Clean Dynamics[1]

László Lövei, Máté Tejfel, Mónika Mészáros, Zoltán Horváth, and Tamás Kozsik

Clean dynamics can be used for implementing mobile code in a functional programming language. Dynamics are type safe, but other semantical properties are not checked before application of the dynamically linked code in the consumer. A language for expressing semantical requirements and an algorithm for comparing requirements with proven properties of the dynamics are presented in the current paper. Sparkle, the dedicated theorem prover of Clean is extended for dealing with open specifications. The properties of the consumer application can be proved based on the requirements are satisfied by the dynamics. New kinds of propositions and tactics are introduced in the paper for dealing with such proofs. The applicability of the new concepts and tools are demonstrated by a running example. The model of comparing properties with requirements is designed in language independent way, so the results may be applicaple for Erlang and other functional languages supporting dynamic code loading.

References

[1] K.M. Chandy and J. Misra. Parallel program design: a foundation. *Addison-Wesley*, 1989.

[2] Z. Horváth, P. Achten, T. Kozsik, and R. Plasmeijer. Verification of the Temporal Properties of Dynamic Clean Processes. *Proceedings of Implementation of Functional Languages, IFL'99*, Lochem, The Netherlands, Sept. 7–10, 1999. pp. 203–218.

[3] M. de Mol, M. van Eekelen, and R. Plasmeijer. Theorem Proving for Functional Programmers, *Sparkle: A Functional Theorem Prover*, Springer Verlag, LNCS 2312, p. 55 ff., 2001.

[4] Z. Horváth, T. Kozsik, and M. Tejfel. Extending the Sparkle Core language with object abstraction. *Acta Cybernetica* 17 (2005), pp. 419-445.

[5] Z. Horváth, T. Kozsik, and M. Tejfel. Proving Invariants of Functional Programs. *Proceedings of Eighth Symposium on Programming Languages and Software Tools*, Kuopio, Finland, June 17-18, 2003., pp. 115-126

[6] Z. Horváth, T. Kozsik, and M. Tejfel. Verifying invariants of abstract functional objects—a case study. *6th International Conference on Applied Informatics*, Eger, Hungary January 27-31 2004.

Simulations on a fractal growth model

Julia Marchis

Fractals are everywhere in nature. To understand why nature gives rise to fractal structures implies the formulation of models of fractal growth based on physical phenomena.

In the random deposition model from a randomly chosen site over the surface, a particle falls vertically until it reaches the top of the column under it. In the ballistic deposition model the particle sticks to the first particle, which it meets, so it is not necessary to reach the top of the column under it (see Barabasi, Stanley, Fractal concepts in surface growth, Cambridge University Press). There are situation, when in the system different type of particles has different behavior. We use a model, where a particle follows the rules of the first model with probability p, and the rules of the second model with probability $1 - p$. We see the obtained cluster for $p = \frac{1}{2}$ in Figure 1, and for $p = \frac{3}{4}$ in Figure 2. This model can have applications in Chemistry, where different type of particles has different behavior in the reactions.

We study this model mainly numerically. For this we study the interface width, given by the formula

$$w(L,t) = \sqrt{\frac{1}{L} \sum_{i=1}^{L} [h(i,t) - \bar{h}(t)]^2},$$

where L is the system size, t is the time, $h(i,t)$ is the height of column i at time t and $\bar{h}$ is the mean height of the surface defined by

$$\bar{h}(t) = \frac{1}{L} \sum_{i=1}^{L} h(i,t).$$

w increases until a time t_{sat}, then reaches a saturation value w_{sat}. For $t \leq t_{sat}$ $w(L,t) \sim t^{\beta}$ and for the saturation regime $w_{sat}(L) \sim L^{\alpha}$. We determine the value of the parameter α and β numerically for different value of p.
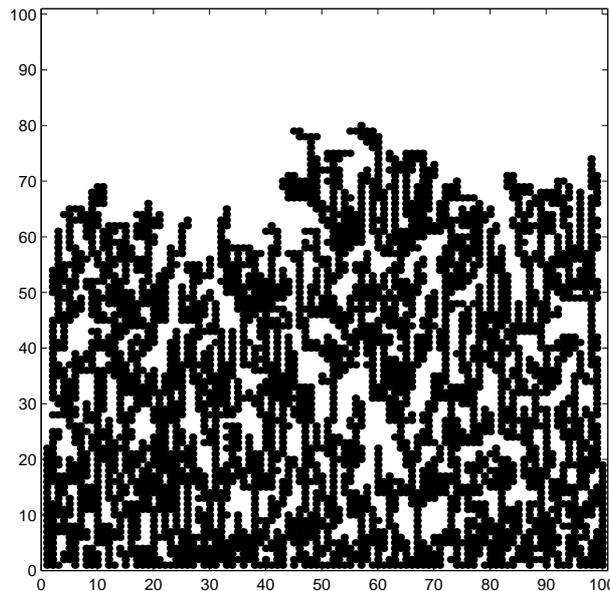
Figure 1: A cluster obtained by depositing 4000 particles on a substrate of horizontal size $L = 100$, for $p = \frac{1}{2}$.
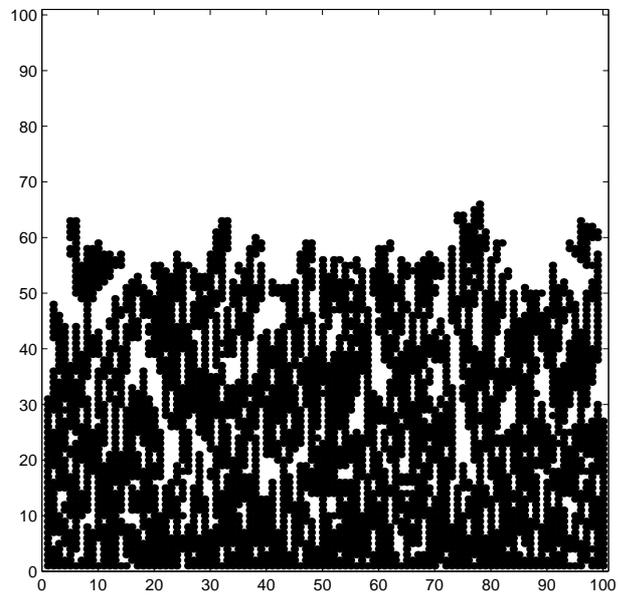
Figure 2: A cluster obtained by depositing 4000 particles on a substrate of horizontal size $L = 100$ for $p = \frac{3}{4}$

Proving Quality of Service Constraints of Multimedia Systems[1]

Mónika Mészáros

Some constraints on the quality of service of multimedia systems can be expressed based on a series of values changing in time. So temporal logics can be used for expressing and comparing quality constraints and properties of streams. A first step towards specification language and a proof system is presented in the current paper for proving the properties of multimedia applications. The model is based on the temporal logic, which was developed for proving correctness of distributed and parallel programs and used also for dealing with proven properties of mobile functional code. The applicability of the new concepts and methods are demonstrated by a running example.

References

[1] L. Böszörményi, C. Becker, H. Kosch, and C. Stary. Quality of Service in Distributed Object Systems and Distributed Multimedia Object/Component Systems, *Technical Reports of the Institute of Information Technology*, University Klagenfurt, TR/ITEC/01/2.05. In Objet-Oriented Technology ECOOP 2001 Workshop Reader Budapest, Hungary, June 2001, pp 7-30 Springer Verlag LNCS 232

[2] K.M. Chandy and J. Misra. *Parallel program design: a foundation. Addison-Wesley*, 1989.

[3] Z. Horváth, T. Kozsik, and M. Tejfel. Extending the Sparkle Core language with object abstraction. *Acta Cybernetica 17* (2005), pp. 419-445.

Optimization algorithms for constraint handling

Gergely Mezei, Tihamér Levendovszky, and Hassan Charaf

Model-based development is one of the most focused research fields. Domain-specific languages can describe the target domain in a flexible and highly customizable way, but they require a proper domain specification: the abstract syntax. Model-based development also requires efficient model transformation techniques to transform the models to source code, or to another model according to the aspects to model. Metamodeling is a means to avoid coding the language definitions manually and to create DSMLs in a visual environment, but the information represented by metamodels has a tendency to be incomplete, and sometimes inconsistent. Besides other issues, there is a need to describe additional constraints about the objects in the model. Transformation steps are also often imprecise without the ability to create constraints in the transformation steps. Therefore, constraint specification and validation lies at the heart of modeling and model transformation.

One of the most wide-spread approaches to constraint handling is the Object Constraint Language (OCL). OCL is a formal language that remains easy to read and write. OCL was originally created to extend the capabilities of UML, and define constraints for the model items. OCL can be used also in generic metamodeling environments to validate the models, or to define constraints in the model transformations. There are several interpreters and compilers that handle OCL constraints in modeling. These tools can extend the metamodel definitions, but they do not support optimization, or the constraint handling in model transformations, therefore, they are not always efficient enough.

Visual Modeling and Transformation System (VMTS) is an n-layer metamodeling and model transformation tool that grants full transparency between the layers (each layer is handled with the same methods). VMTS uses OCL constraints in both model validation and in the graph rewriting-based model transformation. This paper presents algorithms used in VMTS to optimize OCL constraint handling. Since the selection of the model items and their attributes referenced in the constraints has the most serious computational complexity, thus, the optimization algorithms focus on minimizing the number of navigation steps, and attribute queries required to check the constraint. The paper presents an efficient method to minimize the navigation steps of the constraints, and a caching mechanism to optimize the number of database queries during the validation process. Proofs are also provided to show that the optimized and the unoptimized code are functionally equivalent, and that the optimized code is never less efficient than the original. The presented algorithms do not use system-specific features, thus they can be used in any other modeling or model transformation environment.

Automatic vessel segmentation from CDSA image sequences

Balázs Michnay and Kálmán Palágyi

An automatic segmentation of coronary arteries from cardiac digital subtraction angiograms (CDSA) is hereby presented. The basic idea of the research is based on the term used in interventional cardiology: Myocardial Blush (MB). Assessment of MB in general practice of invasive cardiology is performed by a physician with eyeball estimation. The main goal of this research is the quantitative assessment of the myocardial blush grade in the human heart using coronary angiograms.

For assessment, a software toolkit is under development, which consists of the following distinct modules:

- CDSA image processor

- Vessel Mask Calculator (VMC)

- Moving Region Of Interest Tracker

- Measurement Unit

To protect the patient from high-dose X-ray, we have to consider significant noise level throughout the whole sequence. Depending on when and for how long the contrast medium is injected, the intensities of vessels and the myocardium vary by time. The algorithm has to deal with these issues. An ideal DSA sequence only contains contrast medium without any background structure, but in most cases artifacts are present due to patient movement, therefore subtraction errors are visible, which can easily be mistaken for vessels. Creation of DSA sequence is performed by CDSA image processor module.

The first module produces sequences of 8 bits/pixel, gray-scale images with the size of 512x512, containing 8-12 cardiac cycles at the frame rate of 15 fps.

The proposed, fully automated segmentation algorithm exploits the fact that the segmentation is performed on sequences, not on stand-alone images, so it has the opportunity to look back to previous frames or to look ahead to forthcoming ones. Due to the nature of CDSA sequences, we also utilize the fact that on identical frames of the cardiac cycles, vessels - most likely - appear at the same position. In spite of the fact that the images are very noisy due to the low-dose X-ray, the algorithm shall detect thin vessel structures as well. Not having the myocardium segmented enables us to measure the grade of perfusion.

Keywords: *vessel segmentation, cardiac digital subtraction angiogram, myocardial blush, myocardial perfusion*

DRM systems in wireless environment

Rita Móga, Tamás Polyák, and István Oláh

Nowadays there is a growing interest in achieving security in digital rights management (DRM). That means there is a need for securing, watermarking and authenticating video streams. Transmitting a video stream ciphered ensures that only the addressed recipient can view the content. Authentication is needed, to be sure of the sender's identity and also of the quality. Watermarking is for detecting image manipulation and for tracking the way of the media. Creating such a system suiting the needs of the wireless environment is a great challenge, as the wireless transmission channel can easily be attacked or spoofed, and it is noisy, so much higher bit error ratio (BER) is to be expected, than a wired transmission channel. The existing algorithms in ciphering, watermarking and authentication cannot handle the elevated BER, plus they are too complex to be applied in mobile phones or PDAs all at the same time. They either need to be redesigned, or new algorithms shall be created.

In our paper we discuss ciphering, authenticating and watermarking solutions more detailed with some tests proving our results. We also show how these three solutions can work together to give a complete solution on the problems related to secure video transmission in wireless environment.

For ciphering, the major problem is that the whole video stream cannot be ciphered, as it would have too high demand on resources. The algorithm known as Video Encryption Algorithm is suitable for our purposes, as it only ciphers the I frames. According to our measurements, utilizing the statistical randomness of the compressed video information the really expensive ciphering computations (ciphering with AES or RC4) can be reduced to one selected half of the I frames. For the other half non-expensive XOR operation is used for encryption.

The challenge of the authentication is that due the higher BER in wireless environment, the methods of the wired environment cannot be used, because of their sensitivity to single bit errors. In real-time video transmission throwing away too many frames due to single bit errors would be a waist. In our paper we propose a novel approximate authentication algorithm, which is able to accept not just exact, but closely similar match as well.

To protect the stored media stream we should improve security using watermarking. Aside from imperceptibility, a very important property of the watermark is its robustness. As we said, in wireless environment the transmitted media can suffer many different distortions and attacks. We proposed a method to improve the robustness of the watermark-embedding algorithm. To achieve this we hide information bits in larger blocks instead of pixels. The algorithm also uses synchronization information to resist to attacks in the time domain, like frame dropping.

According to our measurements the proposed methods work well together in wireless environments, they offer enough protection to the media, and are suitable to run on devices low on computing resources.

The Power of Deterministic Alternating Tree-Walking Automata[1]

Loránd Muzamel

The concept of a *tree-walking automaton (twa)* was introduced in [1] for modeling the syntax-directed translation from strings to strings. Recently twa are rather used in XML theory as sequential finite state recognizers for tree languages. A twa A, obeying its state-behaviour, walks on the edges of the input tree s and accepts s if the (only) accepting state q_{yes} is accessed. Every tree language recognized by a twa is regular, although it was an open problem for more than 30 years whether twa can be determinized or whether twa can recognize all regular tree languages. The answer for these two questions were provided in [2] and [3] saying that (1) twa cannot be determinized and (2) twa do not recognize all regular tree languages. Hence $dTWA \subset TWA \subset REG$, where $dTWA$ and TWA denote the tree language classes recognized by deterministic twa and twa, respectively, and REG is the class of regular tree languages.

The concept of alternation for various kinds of tree automata was introduced in [7] as a natural extension of nondeterminism. Alternation for twa is considered in [6] and in [4] (without formal definition in the later paper), although they define semantically different computation models. The tree language classes recognized by the two models agree in the nondeterministic case, but do not agree in the deterministic one.

We give a formal definition of the *alternating tree-walking automaton (atwa)* of [4]. A computation of an atwa A on an input tree s starts in the initial state with the reading head at the root node of s. Depending on the applicable rules it generates new parallel computations (such that each has its own copy of s with the current position of the reading head). A accepts s if all the computations spawned from the initial configuration terminate in the accepting state q_{yes}. We denote the tree language class recognized by a (deterministic atwa) atwa by $(dATWA)$ $ATWA$. One can prove that atwa recognize exactly the class of regular tree languages, hence $ATWA = REG$, however, it is still an open problem whether the inclusion $dATWA \subseteq REG$ is strict or not.

Roughly speaking, an atwa A is circular if there is an input tree s such that one of the computations of A on s gets into an infinite loop. Otherwise A is noncircular. We denote the class of tree languages recognized by noncircular deterministic twa (atwa) by $dTWA_{nc}$ ($dATWA_{nc}$).

We investigate the recognizing power of deterministic noncircular atwa and prove that they are more powerful than deterministic twa, but less powerful than atwa. Thus we can write the strict inclusions $dTWA \subset dATWA_{nc} \subset ATWA$. Since $dATWA_{nc} \subseteq dATWA$, we also obtain that $dTWA \subset dATWA$.

Finally we raise an open question. In [5] it was proved that circularity does not give extra power for deterministic twa, i.e. $dTWA = dTWA_{nc}$. This result suggests that possibly $dATWA_{nc} = dATWA$. If it does, then the strict inclusions $dTWA \subset dATWA \subset ATWA$ holds.

References

[1] A.V. Aho and J.D. Ullman. Translations on a context–free grammar. *Inform. Control*, 19:439–475, 1971.

[2] M. Bojanczyk and T. Colcombet. Tree-walking automata cannot be determinized. In *Proceedings of the 31st International Conference on Automata, Languages, and Programming*, pages 246–256. Springer-Verlag, 2004.

[3] M. Bojanczyk and T. Colcombet. Tree-walking automata do not recognize all regular languages. In *STOC '05: Proceedings of the thirty-seventh annual ACM symposium on Theory of computing*, pages 234–243, New York, NY, USA, 2005. ACM Press.

[4] J. Engelfriet and H.J. Hoogeboom. Automata with nested pebbles capture first-order logic with transitive closure. Technical Report 05-02, Leiden University, The Netherlands, April 2005.

[5] A. Muscholl, M. Samualides, and L. Segoufin. Complementing deterministic tree-walking automata. *To appear in IPL*, 2005.

[6] T. Milo, D. Suciu, and V. Vianu. Typechecking for XML transformers. *J. of Comp. Syst. Sci.*, 66:66–97, 2003.

[7] G. Slutzki. Alternating tree automata. *Lect. Notes Comput. Sci.,CAAP*, 159:391–404, 1983.

2D Pattern Repetition Based Lossless Image Compression

Dénes Paczolay

Recently several lossless compression methods were proposed for low-gradient (continuous-tone) images. The majority of everyday photos, and also medical images and satellite images tend to fall into this category. These algorithms were developed for grey-scale images but, analogously, they can be applied to the colour components of coloured images as well. The efficient lossless image compression algorithms developed for low-gradient images are based on the coherence of pixel intensity. That is, based on the intensity of the neighboring pixels they are able to give a good estimate of the intensity of the actual pixel. [4] The more peaked the Gaussian distribution error, the more efficient the estimation will be. It is usually easier to compress the resulting error-image, which is the difference between the original image and the estimated pixel intensities for it. The estimation uses a few parameters at most, so we can get a good compression ratio. An earlier trend was to apply lossy image compression for the purpose of estimation, but the precision of the estimate obtained this way greatly depended on the size of the lossy compressed image, and in many cases this method resulted in worse, rather than better compression ratios [3].

In practice not only low-gradient images need to be compressed, but also high contrast or colour palette images. In such cases, it frequently occurs that the pixels cannot be efficiently estimated from their neighbours, especially when the pixels represent colour indices. But even in these cases the image may contain repetitions which can be exploited for efficient compression. The algorithm which was developed here for such images is based on traditional and general compression techniques. For example, the GIF format uses LZW [1] and the PNG format uses LZ77 [2], these algorithms actually forming part of a well-known zip packer. To improve the compression ratio on low-gradient images the PNG compressed image format uses some simple filters (estimations) – which may be different for each row. A common feature of these widely used compression methods is that the compression and search is done one line at a time. Interestingly, a common "deficiency" of all these methods is that if we rotate the image by 90 degrees, we can get a noticeably different compressed image size.

The novel lossless image compression algorithm described in this paper was developed for images that contain repetitions of identical regions. Such images are produced when using remote desktop access (rdesktop), virtual network computing (VNC), document faxing, or document-to-image conversion. The algorithm we propose here compresses not just the homogenous areas, but also repeated background patterns, so it works very efficiently for patterned background documents and the screenshots of a web-page. If the image contains many texts (take, for example, a screen shot of this abstract), the method finds the blocks of the characters and words and then compresses them.

When searching for the repetitions of 2D patterns, we have to solve the problem of handling overlapping regions. For this we can store information about which pixels have already been processed (compressed), so at these points the equivalence of pixels does not have to be examined. Because the regions tend to overlap the compression algorithm processes the pixels in a non-linear order. A further optimization step we might take is to determine the optimal number of bits required for storing the region sizes and the search distance; we can separate large and small regions if necessary. The pixels that are not compressed (for example, the first pixel) can be stored using Huffman Coding, Arithmetic Coding or some other way. Huffman Coding is simple, but requires storing the Huffman table for decoding or the usage of a dynamic Huffman Code. In experiments we found that in some cases it is better to use a "pixel cache" than some other method. The "pixel cache" contains the recently used 2^m pixel values (or colours). If the next uncompressed pixel is in the "cache" we store its index using $m+1$ bits, otherwise we store the pixel or colour using $n+1$ bits and add this pixel to the "cache" (where the original image stored the pixels in n bits and $m+1 < n$). If necessary, we can separate the

short and long series of stored pixels and hence optimize the number of bits needed to show each storage block (short/long stored pixels, small/large repeated region).

Our algorithm usually proves more efficient than PNG or GIF on the sort of images described above. Furthermore, similar to PNG, our algorithm is suitable for palette and RGB images as well.

References

[1] T.A. Welch. A Technique for High Performance Data Compression, *IEEE Computer*, 17, 6, 8-19, 1984.

[2] J. Ziv and A. Lempel. A Universal Algorithm for Sequential Data Compression, *IEEE Transactions on Information Theory*, 23, 3, 337-343, 1977, citeseer.ist.psu.edu/ziv77universal.html.

[3] X. Wu. An Algorithmic Study on Lossless Image Compression, *Data Compression Conference*, 150-159, 1996, citeseer.ist.psu.edu/wu96algorithmic.html.

[4] N.V. Boulgouris, D.Tzovaras and M.G. Strintzis. Lossless Image Compression Based on Optimal Prediction, Adaptive Lifting and Conditional Arithmetic Coding, *IEEE Trans Transactions on Image Processing*, 10, 1, 1-14, 2001.

Robust Recognition of Vowels In Speech Impediment Therapy Systems

Dénes Paczolay, András Bánhalmi, and András Kocsor

Learning to speak with damaged or missing auditory feedback is very dificult. It would be a real help for the hearing impaired if a computer were able to provide a real-time visual feedback of the quality of the uttered vowels. For speech impediment therapy and teaching reading the SpeechMaster software package was developed in 2004 [7] and was financially supported by the Hungarian Ministry of Education. The members of the Research Group on Artificial Intelligence of the Hungarian Academy of Sciences and University of Szeged were the coordinators of the software development, and SpeechMaster is freely available from our Internet site: http://www.inf.u-szeged.hu/beszedmester.

The software package has been downloaded numerous times and many people use it successfully at many schools and at home. The speech impediment therapy part of SpeechMaster was tested at the School for the Hearing Impaired in Kaposvár. The results of the experiments show that the project achieved its ambitions [5]. Following therapy each young subject pronounced vowels more intelligibly, and the number of mistakes became noticeably fewer. Hence the use of the SpeechMaster can indeed speed up the learning of the utterance of vowels to a significant degree. The speech therapy usually starts in early childhood with introductory exercises like loudness and pitch control drills. The success of our package among children is due to its simple user interface and the playful sound formation exercises. Often children open up more quickly and easily and start uttering sounds using a computer than with a therapist who they do not know. The eficiency of the software can be attributed to two factors. First, the therapy is based on a series of varied and customizable drills. Second, the software produces effective real-time vowel recognition with machine learning [2, 9, 12] and gives a clear visual feedback. Although the project came to an end in 2004, the development of software is still ongoing. We fixed several bugs and added some new features. Currently we are working on improving the accuracy of the real-time vowel recogniser based on experiments and test results.

The goal of the research is twofold. The first goal is to make the software package more user independent. This is needed because the recogniser has to provide on objective rating of the quality of the uttered vowels for all sorts of speakers. To achieve this goal we tried applying real-time speaker normalization [3] and classifier combinations [4]. We found these techniques to be valuable for improving the recognition accuracy [6, 8]. In the current system we divide our vowel database into male, female and child speaker types. We trained three distinct recognisers on these data sets, and trained another to separate these three groups and detect which one is active. In this article we examine what happens if we create categories automatically using a machine learning algorithm. We performed our tests with 3-8 categories using the K-Means method and the Unweighted Pair Group Method with Arithmetic Mean [10] with several configurations.

The second goal is the research to improve the accuracy and stability of the recogniser near the boundary of vowels in the transient period. This is required when the students practice the pronunciation of vowels in words. In the current realisation we trained three-three recognisers separately on isolated vowels and non-isolated vowels e.g. which were uttered in words (separately pronounced vowels and ones uttered in a word). Then we get the scores of the quality of the uttered vowels as follows. First we select the active group (man, woman or child), next we evaluate the two active classfiers (isolated and not-isolated ones), and lastly we get a result using a classifier combination. To improve the stability we tried out new schemes, like training a machine learner to decide whether the actual sound is a vowel or a nasal or if it is a border. Earlier the SpeechMaster package applied three-layer Artificial Neural Networks [1], but in this paper we describe the effects of using a Core Vector Machine [11] classifier as well.

References

[1] C.M. Bishop. Neural Networks for Pattern Recognition. *Oxford Univ. Press*, 1995.

[2] R.O. Duda, P.E. Hart, and D.G. Stork. Pattern Classification. *John Wiley and Son*, New York, 2001.

[3] E. Eide and H. Gish. A parametric approach to vocal tract length normalization. In *ICASSP*, pages 1039-1042, Munich, 1997.

[4] T.K. Ho. Data complexity analysis for classifier combination. *Lecture Notes in Computer Science*, 2096:53-68, 2001.

[5] A. Kocsor. Acoustic technologies in the speechmaster software package. VI(2):3-8, 2005.

[6] D. Paczolay, L. Felföldi, and A. Kocsor. Classifier combination schemes in speech impediment therapy systems. *Acta Cybernetica*, 17:385-399, 2005.

[7] D. Paczolay, A. Kocsor, Gy. Sejtes, and G. Hégely. A 'beszédmester' csomag bemutatása, informatikai és nyelvi aspektusok. IV(1):57-79, 2004.

[8] D. Paczolay, A. Kocsor, and L. Tóth. Real-time vocal tract length normalization in a phonolgical awareness teaching system. *In Text Speech and Dialogue*, volume 2807, pages 4-37, Czech Republic, 2003. Springer.

[9] L.R. Rabiner and B.H. Juang. Fundamentals of Speech Recognition. *Englewood Cliffs, NJ, Prentice Hall*, 1993.

[10] P.H.A. Sneath and R.R. Snokal. Numerical Taxonomy: The Principles and Practice of Numerical Classification. *W.H. Freeman and Company*, 1973.

[11] I.W. Tsang, J.T. Kwok, and P. Cheung. Core vector machines: Fast svm training on very large data sets. 6:363-392.

[12] V.N. Vapnik. Statistical Learning Theory. *John Wiley and Son*, 1998.

Performance Modeling of Hierarchical List Structures

Sándor Palugyai and János Miskolczi

Sequential Lists are widely used in the filed of telecommunication. These lists are exist in firewall systems, in Access Control Lists (ACLs) in routers, or in filter lists in personal computers (for example in LINUX or BSD). With the aid of lists we can filter the incomming/outgoing traffic of our PC or router. In base list-structures, when a packet arrives into the router or PC, the processor matches it against the rules of the list. For long lists this searching time plays a significant role in the packet delay.

In different platforms and in different systems the searching method in these lists can be different. The Access Control Lists in IP routers use simple sequential searching method, or hash structures. The BSD in personal computers use some trick to reduce the execution time of searching. It creates groups from the list-rules and it decides in which group is the searched match. Other systems use balanced tree structures for reducing the searching time. In our work we created a mathematical model, which can describe these hierarchical list structures. From the proposed model we also want to derive performance indices, such as packet delay and packet loss, what packets suffer in the list.

The proposed model uses a Markovian structure, called the Discrete-time Quasi Birth-Death (DQBD) process. That can be described and solved in a matrix-geometrical way. With the aid of this type of process we can also calculate parameters that other structures are unable to handle, e.g. state-distribution or the effect of a finite buffer. Since

using DQBDs the model can derive performance parameters, like packet delay and packet loss ratio, from a parameterized hierarchical list. In our model parameters are the rule system of the hierarchical list together with the matching probabilities and the description of the incoming traffic. The input/output traffic can be described by a Markovian Arrival Processes (MAP) or by a Phase-Type (PH) distribution, which are widely used in the field of Markovian performance modelling.

With the aid of our model, lots of performance parameters can be calculated. Because of this fact the model can be used efficiently in testing and development of filter systems.

Structural Complexity Metrics on SDL Programs[1]

Norbert Pataki, Ádám Sipos, and Zoltán Porkoláb

Structural complexity metrics play an important role in modern software engineering. Testing, bugfixing cover an increasing percentage of the software lifecycle. The most significant part of the cost spent on software design is connected to the maintenance of the product. The cost of software maintenance highly correlates with the structural complexity of the code. With the aid of a good complexity-measurement tool the critical parts of the software can be predicted. Metrics can help writing good quality code by predicting development and testing efforts as early as the design phase.

The Specification and Description Language (SDL) [1] is a high-level specification language widely used in the telecommunation industry (e.g. in protocol specification). The language supports numerous special constructs, like non-determinstic decisions, axiomatic abstract datatype definition, and others. The language is able to work with both text-based representation and graphical representation. The aforementioned language constructs are rarely found in widely used programming languages – like C, C++ or Java –, thus they cause difficulties when applying conventional metrics on SDL programs.

A multiparadigm structural complexity metric is described in [2]. Based on earlier experiments [3, 4] we extend this metric to cover the SDL-specific language constructs. We validate the metric using a database of five-year-long test results originated from real-world applications developed at a major multinational telecommunications company.

References

[1] ITU-T Recommendation Z.100(11/99). Languages for telecommunications applications – Specification and Description Language (SDL), *International Telecommunication Union*, Geneva, 2000,

[2] Z. Porkoláb and Á. Sillye. Towards a multiparadigm complexity measure, *QAOOSE Workshop*, ECOOP 2005, Glasgow, pp. 134-142.

[3] D. Zage and W. Zage. An analysis of the fault correction process in a large-scale SDL production model, *Proc. of the 25th International Conference on Software Engineering*, 2003, pp. 570-577.

[4] J.J. Li, W.E. Wong, W. Zage, and D. Zage. Validation of Design Metrics on a Telecommunication Application, *SERC-TR-171-P*, 1997.

Concept-based C++ template overload compilation with XML transformations

Szabolcs Payrits and István Zólyomi

Parametric polymorphism provides a high level of abstraction by enabling type arguments in function and type definitions. This programming paradigm is implemented by templates in C++. Type safety of parametric polymorphism is greatly improved by concept checking which allows formal specification of type parameter requirements. On one hand, concepts help application developers to implement correct parameter types conforming to static type-safety. On the other hand, concepts also make it possible to overload polymorphic functions based on selected characteristics of the parameter types.

Currently there is no established C++ standard for defining concepts, although several proposals compete to be adopted into the standard. The common approach in these proposals is to define concepts similarly to interfaces in other languages, e.g. Java. In this paper we suggest another possible alternative, similar to current type traits definitions. We assume a set of elementary predicates about C++ types and we define concepts as a composite logical expression of these elementary predicates.

Advanced C++ metaprogramming techniques also require a template overloading mechanism based on template type arguments. However, currently C++ template overloading is fundamentally based on the Substitution-Failure-Is-Not-An-Error (SFINAE) rule. This rule makes template overloading dependent on tracing of internal compiler behaviour, what makes template metaprogramming extremely vulnerable to compiler bugs and compiler-dependent relaxations of the standard.

In this paper we show that in selected cases it is possible to replace the current template overloading approach based on SFINAE with template overloads based on explicitly defined concepts. By eliminating SFINAE, we can separate the compilation of template-containing C++ programs into two phases. During the first compilation phase, resolution of template overloads based on our concept predicates and template instantiation is done. Second phase is the compilation of a c++ program containing no templates into binary code.

With this separation of phases, it is possible to use an independent meta-compiler tool for the execution of the first phase of the compilation. The second compilation phase can be done with any standard C++ compiler.

As a proof of concept, we have started to implement the-phase compilation mechanism. Currently, our C++ concepts are defined as C++ comments for template parameters. Compilation consists of three steps: in first step, we use the Columbus C++ parser to transform our concept-containing C++ program into an XML-based representation called CPPML. Template instantiation with concept-based template overloading is implemented as an XSLT transformation on the CPPML language, resulting in another CPPML document with all templates instantiated. In the third step we convert the transformed document back into standard (template-less) C++ code.

References

[1] B. Stroustrup. The C++ Programming Language (3rd Edition), *Addison-Wesley Professional*, June 1997.

[2] J. Siek, D. Gregor, R. Garcia, J. Willcock, J. Järvi, and A. Lumsdaine. Concept for C++0x, *Technical Report N1758=05-0018, ISO/IEC SC22/JTC1/WG21*, January 2005.

[3] G.D. Reis and B. Stroustrup. Specifying C++ concepts, *N1886, JTC1/SC22/WG21 C++ Standard Committee*, October, 2005.

[4] F. Rudolf and Á. Beszédes. Data Exchange with the Columbus Schema for C++, In *Proceedings of the 6th European Conference on Software Maintenance and Reengineering (CSMR 2002)*, pages 59-66, IEEE Computer Society, March 2002.

Comparison of convolution based interpolation techniques in digital image processing

Miklós Póth

In this paper, we compare different types of interpolation techniques. Both sinc based (linear and cubic) and spline based interpolators are presented in detail. The quality of the interpolation techniques are tested on doubling the size of the images in both directions. We present the accuracy of all techniques, and the frequency responses of the interpolators.

Declarative mapping between concrete and abstract syntax of domain-specific visual languages[1]

István Ráth

Nowadays, the relevance of domain-specific visual languages is rapidly increasing due to the fact that they enable engineers to better concentrate on the problem itself.

In this paper, I present the VIATRADSM framework, a tool developed by Dávid Vágó and myself at the Department of Measurement and Information Systems of the Budapest University of Technology and Economics, which utilizes VIATRA2[2] to provide uniform support for creating editors, transformations, simulators and code generators for domain specific visual languages within the Eclipse framework.

The VIATRADSM framework is based on a plug-in architecture in order to enable the user to view the same modelspace from different domain-specific perspectives, which is an important advantage over current DSM implementations, since those tools focus on generating a separate editor program for each domain. VIATRADSM's tree view-based, syntax-driven editors can be constructed by simply specifying the domain metamodel (abstract syntax).

In most of the current DSM tools, such as MetaEdit+[3], concrete syntax representation is directly mapped to the abstract syntax, meaning that logical entities are always visualised as nodes, and logical relationships as edges. This is acceptable for simple languages, however our experience has shown that using more complex metamodels, especially those conceived for automated model transformations, not only results in visual models being too complicated to overview, but it can also drain system resources heavily.

Thus, modern approaches, such as Eclipse's Graphical Modeling Framework[4], employ a separate *visualisation metamodel*, which describes the structural appearance of diagrams. This technique allows the toolsmith to hide unnecessary detail, however it is still limited in the sense that classes can only be mapped to nodes and references to edges.

The declarative mapping technology presented in this paper, developed for the VIATRADSM framework, extends this idea by using VIATRA2's graph pattern matching engine to give the language engineer complete freedom to define how models are visualized. Thus, complex mappings such as aggregations can be easily defined using VIATRA2's native pattern description language (based on the Visual and Precise Metamodeling[1] language). The goal of the research is to provide full declarative support for specifying these mappings, meaning that language engineers should be able to construct visually appealing and effective tools without any manual coding.

References

[1] D. Varró and A. Pataricza. VPM: A visual, precise and multilevel metamodeling framework for describing mathematical domains and UML (The Mathematics of Metamodeling is Metamodeling Mathematics), in: *Journal of Software and Systems Modeling*, October, 2003.

[2] VIATRA2 Framework, *An Eclipse GMT Subproject*, http://www.eclipse.org/gmt.

[3] MetaCase MetaEdit+, http://www.metacase.com

[4] General Modeling Framework, http://www.eclipse.org/gmf

Occupancy Grid Based Robot Navigation with Sonar and Camera

Richárd Szabó

Autonomous mobile robots is a young interdisciplinary scientific field of growing importance with strong connections to electronics engineering, informatics and cognitive sciences ([5]). Development in this domain will substantially influance our lives in the near future.

Mobile robot simulators offers rapid prototyping environment for modelling, programming, and analizing different robotic tasks. Although difficulties and drawbacks are arising when using simulators, their obvious advantages make them unavoidable ([4]). Webots ([1]) is a well-known representant of these programs, a three-dimensional mobile robot simulator with the possibility to program and control various type of wheeled and legged robots.

At CSCS 2002 the author presented a metric navigation module using occupancy grid in the Webots mobile robot simulation environment ([2]). During this task the robot covers the surface of square-shaped environment while it creates the map. As a continuation of the research a topological graph is placed on top of occupancy grid. The implementation of a topologic graph of the explorable places using the metric map enables the robot to navigate in a more efficient manner as it was presented at CSCS 2004 ([3]).

During this talk the author presents an extension of the former methods. An on-top camera complement the perception of the sonar sensors determining obstacle distance by floor plane extraction ([6]). Efficiency of the three methods is compared and conclusions are drawn about their usefulness.

Keywords: *mobile robots, simulation, navigation, occupancy grid, topological graph, image processing*

References

[1] O. Michel. Professional Mobile Robot Simulation, *International Journal of Advanced Robotic Systems*, Cyberbotics Ltd., 39–42, 2004

[2] R. Szabó. Navigation of simulated mobile robots in the Webots environment, *Periodica Polytechnica — Electrical Engineering*

[3] R. Szabó. Combining metric and topological navigation of simulated robots, *Acta Cybernetican*

[4] R. Szabó. Robotika és szimuláció - Miért nehéz robotnak lenni?, *Élet és tudomány*,

[5] R. Siegwart and I.R. Nourbakhsh, Introduction to Autonomous Mobile Robots, *MIT Press*, 2004

[6] M.C. Martin. The Simulated Evolution of Robot Perception, *Carnegie Mellon University*, 2001, Technical Report CMU-RI-TR-01-32.

Word Boundary Detection Based on Phoneme Sequence Constraints[1]

György Szaszák and Zsolt Németh

Most of ASR applications are based on statistical approach, often using Hidden Markov models to model phonemes. For the purpose of speech recognition, words in the dictionary of speech recognizers are mapped to phoneme model sequences. In case of continuous speech recognition even word networks are connected together during the decoding process, which needs much more computation performance and which also might lead to more confusions due to the longer speech patterns increasing the number of canditate word sequences. In ASR, the Viterbi algorithm is used for recognition to find the most probable match. Using word level segmentation or word boundary information in ASR systems can decrease the searching space during the decoding process thus increase recognition accuracy.

In human speech perception word boundary detection depends on many factors, including prosodic, semantic and sintactic cues [4]. The interaction among these factors is difficult to model, hence prosodic or sintactic information are usually handled and processed separately. Speech production is a continuous movement of the articulating organs, producing a continuous acoustic signal. In human speech processing, linguistic content and phonological rules help the brain to separate syntactic units, such as sentences, phrases (sections between two intakes of breath), sintagms or even words.

Since the mid-eighties there have been several trials to exploit prosodic information in human speech [5, 6]. In the Laboratory of Speech Acoustic of the Budapest University of Technology and Economics, we have already investigated word boundary detection possibilities based on prosodic features, reliing mainly on fundamental frequency and energy level data derived from the acoustic signal [1]. We have also trained a prosodic HMM based word boundary segmentator in order to use it as a front-end module for Viterbi decoding in ASR [2].

Using phoneme sequence information is an alternative for word boundary detection. Phoneme sequence constarints can be derived by matching a complete set of 3 phoneme sequences that can occur accross word boundaries [3, 8]. Phoneme assimilations must be taken into account over word boundaries.

In this paper we would like to investigate the use of phoneme sequence constraints in word boundary detection for Hungarian language. We have collected a phonetically rich, large Hungarian text database from the Internet. This material was transcribed phonetically taking into account all possible assimilations over word boundaries, and also the pronunciation variants of words [7]. We matched against this corpus complete sets of 2 or 3 phoneme sequences in order to get a frequency statistics of each phoneme sequence element word internally and over word boundaries. In this way we obtained a frequency database for all possible phoneme sequences in Hungarian. Based on this information we derived a set of rules (eg. phoneme sequence constraints) which can be used for word boundary detection for Hungarian language. We also investigated the adaptability of the method for phoneme based speech recognizer's output that contains some phoneme confusions, since phoneme recognition accuracy in ASR - just like in case of human listeners - is around 70-80%. A new challenge is the paralel use of prosody based and phoneme sequence based word boundary detection methods, which might be of interest in our future work. In this paper we would like to present our methodology and evaluate our results obtained by phoneme sequence based word boundary detection for Hungarian language.

References

[1] K. Vicsi and Gy. Szaszak. Automatic Segmentation of Continuous Speech on Word and

Phrase Level based on Suprasegmental Features. In: *Proceedings of Forum Acusticum* 2005 Budapest, Hungary, 2669-2673.

[2] Gy. Szaszák and K. Vicsi. Folyamatos beszéd szószintű automatikus szegmentálása szupraszegmentális jegyek alapján, *MSZNY 2005*, Szeged, Hungary, 360-370.

[3] J. Harrington, G. Watson, and M. Copper. Word Boundary Identification from Phoneme Sequence Constraints in Automatic Continuous Speech Recognition, 1998, COLING: 225-230.

[4] S.L. Mattys and P.W. Jusczyk. Phonotactic and Prosodic Effects on Word Segmentation in Infants, *Cognitive Psichology*, 1999, 38: 465-494.

[5] Di Cristo. Aspects phonétiques et phonologiques des éléments prosodiques, *Modeles linguistiques Tome III*, 2:24-83.

[6] M. Rossi. A model for predicting the prosody of spontaneous speech (PPSS model), *Speech Communication*, 1993, 13:87-107.

[7] K. Vicsi and Gy. Szaszák., Examination of Pronunciation Variation from Hand- Labelled Corpora, *TSD 2004 Brno*, Czech Republic: 473-480.

[8] K. Müller. Probabilistic Context-Free Grammars for Phonology. *Workshop on Morphological and Phonological Learning at Association for Computational Linguistics 40th Anniversary Meeting (ACL-02)*, University of Pennsylvania, Philadelphia, USA.

Caching in Multidimensional Databases

István Szépkúti

One utilisation of multidimensional databases is the field of On-line Analytical Processing (OLAP). The applications in this area are designed to make the analysis of shared multidimensional information fast.

On one hand, speed can be achieved by specially devised data structures and algorithms. On the other hand, the process of analysis is a cyclic one. In other words, the user of the OLAP application runs his or her queries in a row. The output of the latest enquiry may be there (at least partly) in one of the previous results. Therefore caching also plays an important role in the operation of these systems.

However, caching in itself may not be enough to ensure acceptable performance. Size does matter: The more memory is available, the larger we gain by loading and keeping information in there.

Oftentimes, the cache size is fixed. This limits the performance of the multidimensional database, as well, unless we try and condense the data in order to move a greater proportion of them into the memory. Caching combined with proper compression methods promise further performance improvements.

In this paper, we investigate how caching influences the rapidity of OLAP systems. Different physical representations (multidimensional and table-based) are evaluated. For the thorough comparison, a model is proposed. We draw conclusions based on this model, and these conclusions are verified with empirical data. In particular, using benchmark databases, we show examples when one physical representation is more beneficial than the alternative one and vice versa.

Testing and optimization of a continuous speech recognizer with a middle sized vocabulary[1]

Csaba Teleki, Szabolcs L. Tóth, and Klára Vicsi

This paper describes testing and optimization methods of a speaker independent, continuous and automatic speech recognizer, developed at the Laboratory of Speech Acoustics. The recognizer is based on acoustic and language models built using statistical methods. It is capable of training and recognizing middle sized vocabulary-based texts, containing 1000-20 000 different words. New methods were developed in acoustical preprocessing [1], in statistical model developing and language modeling. The recognizer uses HMM acoustic models for phonemes [2] and bigram language model, using non-linear smoothing [3]. During the tests we varied the acoustical models and the language model of the recognizer.

The acoustic models were trained, using the Hungarian Reference Database (MRBA) [4].The optimization of these models were done earlier. In the end, the recognizer uses Quasi-Continuous HMM (QCHMM) models with 4-5 states based on recordings made with 16 kHz sample rate, 17 derivates in Bark frequency domain, 17 time derivates, 17 second time derivates, and energy as input vector. For details on optimization of acoustic level see [5].

The language model were developed using a text corpus collected from 2700 medical reports from the of Medical Clinic nr 2 of SOTE in Budapest and 6365 medical reports from the Medical School of Szeged.

For testing purposes we used a couple of different type of recordings. One of them was made at Medical Clinic nr 2 of SOTE in Budapest. The test material contains spoken medical reports of 5 different doctors each physician dictating 4 different medical reports. Therefore, 20 different recordings, from 5 different speakers were used for testing. The recordings were made in an examination room with bad acoustic condition. The other type of testing material was recorded in our laboratory by 2 speakers, containing the same text, but in better acoustic condition.

The recognizer was evaluated from multiple points of view: word error rate (WER), real time factor (RT), memory capacity used during recognition (MEM).

This paper presents results that show that the goodness of the recognizer is influenced by:

- the type training material,

- the acoustic conditions of the testing material,

- the pronunciation quality of the speaker,

- the size of the searching space,

- the weighting of the role of the bigram language model.

We trained three different types of acoustic models: one based only on male speech, another one based on only on female speech and one based on mixed (male and female) speech from the MRBA database. So far, for the recordings made at the Medical Clinic, the word error rate is 30% for lexemes, using the language model built based on the reports made at the Medical Clinic from Budapest. Using the recordings made at our laboratory, in better acoustical conditions, the results got significantly better (WER = 21%). Mean value for the real time factor were 1 and for the memory capacity used were 300MB. The pronunciation quality of the speaker influences the WER. If the pronunciation quality is higher, de recognizer gives better results. If the size of the searching space is bigger, the results will be better, but then increases the processing time and the used memory.

The above mentioned preliminary results showed that optimization of the acoustic and language model of the recognizer is required. Beyond tuning the mentioned parameters to the appropriate values, optimization at different levels of the training material is necessary. At the level of acoustic models we will show results regarding the effectiveness of using supervised speaker adaptation. At the level of the language model we tried to optimize the vocabulary, using perplexity and active vocabulary adaptation. Using these optimization methods, the word error rate decreased under 10 %.

References

[1] M. Szarvas and S. Furui. Evaluation of the Stoachic Morphosyntactic Language Model on a One Million Word Hungarian Dictation Task, In: *EUROSPEECH 2003 - Genova*, p. 2297-2300

[2] S. Chen, D. Beeferman, and R. Rosenfeld. Evaluation Metrics For Language Models, In: *DARPA98*, National Institue of Standards and Technology (NIST)

[3] P. Clarkson, and T. Robinson. Towards improved language model evaluation measures

[4] K. Vicsi, A. Kocsor, Cs. Teleki, and L. Tóth. Hungarian Speech Database for computer-using environments in offices, *Second Conference on Hungarian Computational Linguistics (MSZNY 2004)*, Szeged, Hungary, 2004, p. 315-319.

[5] Sz. Velkei and K. Vicsi. Speech recognizer model-building experiments at the level of acoustic and phonetics, on behalf of developing a speech recognizer for medical reporting, *Second Conference on Hungarian Computational Linguistics (MSZNY 2004)*, Szeged, Hungary, 2004, p. 307-315.

[6] C. Becchetti, L.P. Ricotti. Speech Recognition,*Theory and C++ implementation*, Fondazione Ugo Bordoni, Rome, 1999. ISBN 0-471-97730-6

[7] HUMOR, Morphological analyzer for developers,
    http://www.morphologic.hu/h_humor.htm

Model-based Development of Graphical User Interfaces[1]

Dániel Tóth and Eszter Jósvai

The development of computer technology vastly broadened the range of potential users. Previously the primary users of computer applications were developers themselves, nowadays most applications are developed for end users who use them for regular work and have no special understanding of the inner workings of the system. These users need graphical user interfaces (GUIs) designed with these requirements in mind. Another important trend in application design is the requirement to be able to run on a wide range of different platforms.

Present day GUI designer tools are very advanced in that they make the design process a very simple task. They usually feature code generation facilities to remove the requirement of lengthy coding. However they are typically tied to a single programming language and platform. There are attempts to create platform-independent GUI description techniques, but they typically approach the problem by adding another layer above existing frameworks and widget libraries. Usually some interpreter is used to translate the platform-independent description to the native widget set in run-time.

True platform independence can be achieved trough the use of OMG's Model Driven Architecture (MDA) guidelines. The MDA defines a development procedure where the application is designed as a platform-independent model which is automatically transformed into platform specific models which serve as a base for automatic code generation for a wide variety of platforms. The ability to modify the application model easily, then propagate the modification automatically to the generated appliation is particularly important in GUI design, since it requires frequent conciliation with the end users.

We created a GUI modelling system using the VIATRA2 transformation and modelling framework that allows the generation of native code for different platforms. It can also take the platform independent GUI model from different kinds of model sources. Our prototype implementation used UML models for input because it is an industry standard for software design and documentation. UML by itself is not useable for GUI modelling and this has been a major obstacle in utilizing the MDA concept for such development. We created a UML profile to make platform independent GUI modelling in UML possible.

We have found that user interface design is not as easy in UML as in the special GUI design tools. Therefore we conducted further research in the field of domain specific modelling (DSM). DSM allows the creation of an easy to use visual GUI model editor, similar to the plaform specific design tools, while maintaining the platform independence and the ability to output UML models for unified documentation of the application.

References

[1] D. Varró and A. Pataricza. VPM: A visual, precise and multilevel metamodeling framework for describing mathematical domains and UML, *Springer-Verlag*, 2003.

[2] P.P. da Silva and N.W. Paton. User Interface Modelling with UML, 2000.

[3] M. Iseger. Domain-specific modeling for generative software development, 2005,
`http://www.itarchitect.co.uk/articles/display.asp?id=161`

Hybrid algorithm for sentence alignment of Hungarian-English parallel corpora

Krisztina Tóth, Richárd Farkas, and András Kocsor

Parallel corpora, i.e. texts available in two or more languages with their segments aligned, are playing an increasingly important role in natural language processing. The processing of parallel corpora facilitates more easier for example, the building of automated machine translation systems and multilingual lexica. This paper describes an efficient hybrid technique for aligning sentences with their translation in a parallel corpus.

The Algorithm

The accuracy of a sentence-aligning algorithm depends to a great extent on the preceding sentence segmentation. Thus we proposed a sentence segmenting algorithm that works both on Hungarian and English texts and has an error rate of 0.5% or less.

In the process of alignment the sentence segmenting algorithm divides the English and Hungarian raw texts first into paragraphs, then into sentences. The output of the alignment is a TEI-compliant file containing the identifiers of the aligned Hungarian and English sentences.

In the sentence aligning algorithm we soon realised that sentences could be not extended beyond paragraph boundaries. The majority of the sentence-aligning algorithms are based on a kind of length-measure [6].

We chose the number of characters as the length-measure, and also took into consideration the fact that the number of characters within sentences have a correlation, thus in general a long sentence corresponds to a long sentence and a short sentence to a short one [3]. Obviously, it happens only in the simplest cases that a single source sentence corresponds to a single target sentence. It might happen, due to the translator's freedom of interpretation, that one source sentence corresponds to several possible target sentences and that several source sentences correspond to a single target sentence. For this reason we present a novel algorithm that is capable of recognising 1-N, N-1 and N-M correspondences as well.

We combined statistical, length-based alignment with an anchor-searching method that forms the basis of partial text-alignment. We gain such lexical information via anchor-recognition, the latter of which is present both in the target and source sentences. The anchor-searching processes published for Hungarian use the normalised form of numbers and words starting with capital letters as anchors. We added acronyms and abbreviations as anchors and employ named entities (NE) instead of words that start with a capital letter. We extracted the named entities via a quasi language-independent NE recogniser built by FARKAS et al [2]. The reason why we omitted the use of words starting with capital letter as an anchor was to avoid filtering anchors moreover, we gain valuable information by retaining inflected, lower-case named entities characteristic of agglutinative languages.

During the implementation of our hybrid algorithm we made use of both the statistical and anchor methods to examine similarities between the source and target sentences. Our costs are calculated on the basis of the length-proportion of corresponding sentences and costs assigned to the applied synchronising categories [6], just like the Gale and Church method used as a reference algorithm. We set the tolerance limit of length-proportion to 0.85 since Hungarian texts are generally 15% longer than their English equivalents. This data came from the sentence-length statistics of an English-Hungarian translation memory. We also assigned costs to similarities calculated on the basis of anchor-information, then the costs were totalled. Our cost thus derives from the costs of length-proportions measures and the costs of anchor-distances. In the sentence-alignment process the matches (mappings) with a minimal cost are chosen.

Experimental results

The implemented sentence segmenting algorithm was tested on a randomly selected part of the Szeged Treebank [1] with a result of 99.7-99.85%. This result surpasses earlier results published for the Hungarian language [4, 5].

The testing of the hybrid algorithm described above was carried out on the English-Hungarian parallel corpus built by the Research Group on Artificial Intelligence of the Hungarian Academy of Science at the University of Szeged. This corpus is comprised of 8,000 sentence pairs. The corpus contains mostly bilingual newspaper articles and general descriptions. The corpus tries to represent the Hungarian and English everyday languages. Due to the use of anchors the hybrid algorithm identified the synchronising units correctly, thus we were able to achieve more accurate results than with the length-based alignment of the Gale and Church method used as the reference algorithm. In the cases of alignment without anchors, this was determined solely by the cost of length-based matching (mapping).

References

[1] D. Csendes, J. Csirik, and T. Gyimóthy. The Szeged Corpus: A POS tagged and Syntactically Annotated Hungarian Natural Language Corpus, In *Proc. of TSD 2004*, Brno, LNAI vol. 3206, pp. 41-49, 2004.

[2] R. Farkas, Gy. Szarvas, and A. Kocsor. Named Entity Recognition for Hungarian Using Various Machine Learning Algorithm, In *Acta Cybernetica* submitted paper, 2006.

[3] W.A. Gale and K.W. Church. A Program for Aligning Sentences in Bilingual Corpora, In *29th Annual Meeting of the Association for National Linguistics*, 1991.

[4] Hunglish cd-rom

[5] A. Mihaczi, L. Németh, and M. Rácz. Magyar szövegek természetes nyelvi előfeldolgozása, In: *Magyar Számítógépes Nyelvészeti Konferencia*, pp. 40., 2003.

[6] G. Pohl. Szinkronizációs módszerek, hibrid bekezdés- és mondatszinkronizációs megoldás, In: *Magyar Számítógépes Nyelvészeti konferencia*, pp. 254-259, 2003.

Database Design Patterns

Ovidiu Vasutiu and Florina Vasutiu

The purpose of this communication is to analyze the classical database design solutions to some classical problems like multi-language support or "history data" maintenance in software applications. We shall focus on the web-application and on the ERP Systems.

As internationalization is no longer a paper concepts, the applications address users that speak different languages - so multi-language support is one of the issues to be considered when designing an application. We have analyzed the way in which several applications implement the multi-language support.

Among of aspects to be considered when implementing the language database design are: the possibility to switch the language during the runtime, the measure in which the application data needs to be translated, the most important, the dynamism of the data and the data type.

If the translation is to be stored in the same table with the "original" data, or in "dictionary" table is to be discussed. If the data used in application is relatively static, than a single dictionary table (more dictionary tables grouped by different criteria) properly indexed is doing the job. For dynamic, different in type application data, different solutions involving auto-referred tables or more dictionary tables are used.

Maintaining the old version of the data, saving intermediary data and the archive of date to be rarely used are some aspects to be considered when we speak about "history data" maintenance. Depending about the amount of legacy data that the business needs required, different design are used and they are to be presented and subject to discussions, as a definitive pattern is not yet defined. The need of additional columns to the table or maybe the need of a parallel table might be noticed, but also auto-referred tables might be used.

The way in which the classical database operations: insert, update, delete are performed for each of this solutions, the optimal indexations of the tables and the advantages of each method are to be discussed. An exemplification of the applications of these remarks to the classical ERP solutions (SAP mostly) are to be presented and the recommendations for implementing web-applications are further to be discussed.

The modeling method used is, generally, the entity-relationship model, but the object-role modeling is going to be used were further details are needed.

Transformation and simulation of domain specific languages[1]

Dávid Vágó

In the last decade, as computer applications have become more and more complex, modeling became an essential and unavoidable part of software development. Models represent software at a higher level of abstraction, allowing developers to focus mainly on the business intelligence, and delay implementation-specific questions to a later stage. In the early 1990s UML (Universal Modeling Language) was conceived to support the design of object-oriented systems. UML gives a language (graphical notation and semantics) to describe the structure and overall behavior of software components, but due to its imprecise semantics [1], formal verification of UML models is hardly possible. Its other weakness is that it tries to be as universal as possible, which makes application field specific modeling elaborate, generally done by extensions called *UML profiles*. Certain modeling aspects (such as reliability) already have a well-established formalisms (in case of reliability, that would be Petri nets). Thus a reliability experts, who had always worked with Petri nets, may find it difficult to transform their ideas to the language of UML.

Domain specific modeling (DSM) is a more advanced design methodology, which introduces the term *domain*. A domain is the set of concepts of a specific application field. For example, the domain of mobile communication includes the concepts *SMS*, *dialing*, *contact list*, etc. DSM allows the developers to create *domain-specific models*, models made up of these application-specific concepts. Thus an user interface designer would use familiar concepts like *window*, *scrollbar* or *menu* when designing the appearance of the system, and on the other hand, the reliability expert working on the same system could use Petri nets.

There are various DSM tools available (MetaEdit+, Microsoft DSL Tools and many others), and all of them provides support for graphical editing of domain-specific models and (generally template-based) source code generation. However there are two other important areas of domain-specific modeling, simulation and model transformations, which are not widely supported. Model-level simulation makes it easier for the developer to test the behavior of his/her model. Model transformation on the other hand is useful for formal model verification. Some tools make it possible to follow the execution of the generated code on the model level, but interactive simulation or model verification support is rare. In my paper, I focus on VIATRADSM, a tool designed by István Ráth and Dávid Vágó at the Department of Measurement and Information Systems of the Budapest University of Technology and Economics. This tool is based on VIATRA2[2], a universal model transformation framework made by other members of the same research group.

The underlying VIATRA2 framework has strong transformation capabilities (based on graph transformation rules and abstract state machines), and in my paper I examine how these capabilities may be used in a domain-specific modeling tool. My goal is to integrate interactive simulation and model transformation features into the existing VIATRADSM framework. Many existing DSM tools lack simulation support because they provide no way of describing model semantics. In the first part of my paper, I discuss how model semantics can be expressed using the declarative GTASM language of VIATRA2. Using GTASM, model semantics can be given using simple, pattern-based transformation steps. Using Petri nets as an example, I demonstrate how interactive simulation works in our DSM tool. Finally, I show that complex model transformations can be expressed and executed as simply as simulations, and I give an example how such a transformation can be used for model verification.

References

[1] D. Varró and A. Pataricza. VPM: A visual, precise and multilevel metamodeling framework for describing mathematical domains and UML (The Mathematics of Metamodeling is Metamodeling Mathematics), in: *Journal of Software and Systems Modeling*, October, 2003

[2] `VIATRA2` Framework. *An Eclipse GMT Subproject*
`http://www.eclipse.org/gmt/.`

Model transformations on the Preprocessor Metamodel - Graph Transformation approach

László Vidács

There is an important trend in software engineering that focuses on development methods where models are used besides concentrating on source code only. In previous work we designed a metamodel for C/C++ preprocessing (Columbus Schema for C/C++ Preprocessing [1]). The metamodel describes the source code from the preprocessor's point of view and enables the studying the preprocessor constructs in detail.

Refactoring techniques are considered as promising means for software development. There exists extensive literature on refactoring object-oriented programs. Refactoring on C++ code is supported by a variety of tools. Preprocessor means obstacle for these tools. Since preprocessing loses information about macros and different configurations, it is not enough to work on the preprocessed code. Tool developers face the problem of having two different languages: C/C++ and the preprocessor language. There are several research papers about solving this problem, but the aim of these contributions is to implement only the C/C++ refactorings. A recent work of Garrido [2] combines directive usages into the C language. The contribution contains several preprocessor related refactorings, which are not related to the C language itself but to the preprocessing directives. Refactorings made on directives may be not so complicated as language dependent refactorings combined with directives. But they may have important role in refactoring real C/C++ programs. However, refactorings are usually considered from the implementation point of view only, not from a conceptual view.

A conceptual view on preprocessor refactorings on the basis of metamodels and graph transformations allows to express the properties of refactorings more precise. A conceptual view also opens the possibility for viewing refactorings on the semantical level. In this contribution we provide transformations on preprocessor constructs (like adding a new macro parameter) as graph transformations. We use a single pushout approach providing the left-hand side and right-hand side of the transformation. The transformed graph is a directed, labelled and attributed graph. It is a program graph according to the UML class diagram of the metamodel. To be more general we extend the usual notation with the concept of multi-nodes. It is common to use NACs (Negative Application Condition) to prevent application of a graph transformation rule, but we chose to use OCL expressions instead, because in program transformation context the application conditions are as complex as the structural changes in the graph. OCL is appropriate not only to formalize conditions but to check them in a program graph.

To validate this approach we choose the USE system instead of existing graph transformation engines. USE can handle UML metamodels and models, this concept fits to our existing program representation. The graph transformation rule is given as a USE description and converted into basic operations on the model instance. The USE system evaluates OCL expressions to check the application conditions and performs the operations. The time consuming part of a transformation is to find the appropriate place in the graph where it is applicable. In case of refactoring the user has to determine the place of the transformation, in our implementation the place is a parameter of the transformation rule.

Our approach uses the result of graph transformation theory in software (re)engineering. Transformations on preprocessor constructs are not well studied yet. In this paper they are formalized using a higher level representation which also enables condition checking. The metamodel-model approach of the USE system provides a flexible framework to manipulate graphs and to visually check the transformation rules. Besides model to model transformations the reverse engineering tools of the Columbus system give the possibility of source code to source code level transformations.

References

[1] L. Vidács, Á. Beszédes, and F. Rudolf. Columbus Schema for C/C++ Preprocessing, *CSMR*, IEEE Computer Society, Tampere, Finland

[2] A. Garrido. Program Refactoring in the Presence of Preprocessor Directives *Ph.D. thesis, UIUC* 2005

High-level Restructuring of TTCN-3 Test Suites

Antal Wu-Hen-Chang, Dung Le Viet, and Gyula Csopaki

Telecommunication software provides the foundation of the communication infrastructure. These systems must be reliable, efficient and compatible with systems from different vendors. Consequently, their development must be accompanied by quality assurance activities, thus testing plays a vital role during the development process of each telecommunication system. The purpose of testing is to find all errors and shortcomings of the system. This is a very resource-demanding and time-consuming task, because it requires the manual effort of many well-trained developers, therefore its support is an important challenge.

TTCN-3 (Testing and Test Control Notation 3) [1] is the new industry-standard test specification language that was developed and standardized by the European Telecommunication Standards Institute (ETSI). It is a very powerful language that has been tried out on different application areas of testing. It can be applied for all kinds of black-box testing for reactive and distributed systems and it is suitable to test virtually any system including telecommunication and mobile systems, Internet and CORBA based protocols.

The general testing process with TTCN-3 includes the following main steps: The developed test suite is compiled and extended with an adaptor that provides the connection between the tested system and the executable test suite. Then the executable test suite is executed against the system under test. Finally, the results are evaluated.

TTCN-3 has a special language element – the template – that provides sophisticated means for describing test data. In order to test complicated systems, the TTCN-3 templates can be created either in a manual or in an automatic way, but in neither case is the result optimal, since developers cannot cope with the enormous number of huge data structures, and automatic methods focus primarily on the generation problem. According to our empirical experiences test data definition occupies at least 60-70 percent of a complete test specification and they are highly redundant. Consequently, these modules are unnecessarily large that leads to several problems. In case of very large TTCN-3 modules the compilation time can be surprisingly long, which sets back the development process. It is not uncommon, that the compilation of the test specification takes more than an hour on an average computer and complicated test suites consist of several different modules. Besides, executable test suites derived from large modules have performance drawbacks, that makes it harder to develop performance or scalability test suites, where performance is a critical issue. Furthermore, we must take into account, that the development process of a test suite is cyclic, therefore these problems appear repeatedly.

By eliminating the redundant and unused data structures, the quality of the generated implementation code can be significantly improved. In our paper, we introduce a re-engineering method that can be applied without human intervention to test data templates defined in TTCN-3. The approach analyzes and restructures [2, 3] an already existing TTCN-3 template specification, so that it becomes more compact, redundancy-free and the compilation time is reduced. Naturally, the alterations retain semantical correctness, only syntactical changes are introduced.

Beside the static test data restructuring, it is worth analysing the TTCN-3 description of the dynamic behavior. The test specifications often contain repeated test steps, but because of their complex nature, they can't be detected easily. We examine the possibility of applying dynamic behavior restructurings to TTCN-3, that would extract and reuse frequently appearing test steps.

References

[1] ETSI ES 201 873-1 V2.2.1 (2002-08). Methods for Testing and Specification (MTS); The Testing and Test Control Notation Version 3; Part1: TTCN-3 Core Language.

[2] E.J. Chikofsky and J.H. Cross. Reverse Engineering and Design Recovery: A Taxonomy, *IEEE Software*, Vol 7(1):13-17, 1990.

[3] T. Mens and T. Tourwe. Survey of Software Refactoring, *IEEE Trans. on Software Engineering*, Vol. 30(2):126-139, 2004.

List of Participants

Aczél, Kristóf: Budapest University of Technology and Economics, H-1111 Budapest, Goldman György tér 1., Hungary, E-mail: `aczelkri@aut.bme.hu`

Asztalos, Domonkos: Software Engineering Group, Ericsson Hungary Ltd., H-1300 Budapest P.O. Box 107, Hungary, E-mail: `Domonkos.Asztalos@ericson.com`

Balázs, Péter: Institute of Informatics, University of Szeged, H-6701 Szeged P.O. Box 652, Hungary E-mail: `pbalazs@inf.u-szeged.hu`

Balogh, Ádám: Eötvös Loránd University, H-1117 Budapest, Hungary E-mail: `bas@elte.hu`

Balogh, János: University of Szeged, JGYTF, Deptartment of Computer Science, H-6725 Szeged, Hungary, E-mail: `balogh@jgytf.u-szeged.hu`

Bánhalmi, András: Research Group on Artificial Intelligence of the Hungarian Academy of Sciences and University of Szeged, H-6720 Szeged, Aradi vértanúk tere 1., Hungary E-mail: `bandras@inf.u-szeged.hu`

Bánhelyi, Balázs: Institute of Informatics, University of Szeged, H-6701 Szeged P.O. Box 652, Hungary, E-mail: `banhelyi@inf.u-szeged.hu`

Bátori, Gábor: Software Engineering Group, Ericsson Hungary Ltd., H-1300 Budapest P.O. Box 107, Hungary, E-mail: `Gabor.Batori@ericson.com`

Békési, József: University of Szeged, JGYTF, Deptartment of Computer Science, H-6725 Szeged, Hungary, E-mail: `bekesi@jgytf.u-szeged.hu`

Berceli, Tibor: Budapest University of Technology and Economics, H-1111 Budapest, Goldman György tér 1., Hungary

Biczó, Mihály: Eötvös Loránd University, H-1117 Budapest, Hungary E-mail: `mihaly.biczo@axelero.hu`

Bilicki, Vilmos: Institute of Informatics, University of Szeged, H-6701 Szeged P.O. Box 652, Hungary, E-mail: `bilickiv@inf.u-szeged.hu`

Bogárdi-Mészöly, Ágnes: Budapest University of Technology and Economics, H-1111 Budapest, Goldman György tér 1., Hungary, E-mail: `agi@aut.bme.hu`

Busa-Fekete, Róbert: Research Group on Artificial Intelligence of the Hungarian Academy of Sciences and University of Szeged, H-6720 Szeged, Aradi vértanúk tere 1., Hungary E-mail: `busarobi@inf.u-szeged.hu`

Charaf, Hassan: Budapest University of Technology and Economics, H-1111 Budapest, Goldman György tér 1., Hungary, E-mail: `hassan@aut.bme.hu`

Csendes, Tibor: Institute of Informatics, University of Szeged, H-6701 Szeged P.O. Box 652, Hungary E-mail: `csendes@inf.u-szeged.hu`

Cserkúti, Péter: Budapest University of Technology and Economics, H-1111 Budapest, Goldman György tér 1., Hungary, E-mail: `csiri@aut.bme.hu`

Csirik, János: Institute of Informatics, University of Szeged, H-6701 Szeged P.O. Box 652, Hungary
E-mail: csirik@inf.u-szeged.hu

Csopaki, Gyula: Budapest University of Technology and Economics, H-1111 Budapest, Goldman György tér 1., Hungary, E-mail: csopaki@tmit.bme.hu

Csorba, Kristóf: Budapest University of Technology and Economics, H-1111 Budapest, Goldman György tér 1., Hungary, E-mail: kristof@aut.bme.hu

Csorba, Máté J.: Budapest University of Technology and Economics, H-1111 Budapest, Goldman György tér 1., Hungary
E-mail: csorba@tmit.bme.hu, mate.csorba@ericsson.com

Csörnyei, Zoltán: Eötvös Loránd University, H-1117 Budapest, Hungary
E-mail: csz@elte.hu

Dávid, Ákos: University of Veszprém, H-8200 Veszprém, Egyetem u. 10., Hungary
E-mail: davida@almos.vein.hu

Dávid, Róbert: Budapest University of Technology and Economics, H-1111 Budapest, Goldman György tér 1., Hungary, E-mail: dr420@hszk.bme.hu

Dávid, Zoltán: Budapest University of Technology and Economics, H-1111 Budapest, Goldman György tér 1., Hungary, E-mail: dzoli@aut.bme.hu

Dévai, Gergely: Eötvös Loránd University, H-1117 Budapest, Hungary
E-mail: deva@elte.hu

Dombi, József: Institute of Informatics, University of Szeged, H-6701 Szeged P.O. Box 652, Hungary
E-mail: dombi@inf.u-szeged.hu

Dombi, József Dániel: Institute of Informatics, University of Szeged, H-6701 Szeged P.O. Box 652, Hungary, E-mail: dombijd@inf.u-szeged.hu

Dombi, Krisztina: Institute of Informatics, University of Szeged, H-6701 Szeged P.O. Box 652, Hungary, E-mail: mkriszta@inf.u-szeged.hu

Domes, Ferenc: Faculty of Mathematics, University of Vienna, Vienna, Austria
E-mail: ferenc.domes@univie.ac.at

Egea, Jose A.: Process Engineering Group, IIM-CSIC, Vigo (Spain)
E-mail: jegea@iim.csic.es

Erdélyi, Gáspár: Eötvös Loránd University, H-1117 Budapest, Hungary
E-mail: gaspar.erdelyi@elte.hu

Faragó, Szabolcs: Institute of Informatics, University of Szeged, H-6701 Szeged P.O. Box 652, Hungary, E-mail: faragosz@inf.u-szeged.hu

Farkas, Richárd: Research Group on Artificial Intelligence of the Hungarian Academy of Sciences and University of Szeged, H-6720 Szeged, Aradi vértanúk tere 1., Hungary
E-mail: rfarkas@inf.u-szeged.hu

Galambos, Gábor: University of Szeged, JGYTF, Deptartment of Computer Science, H-6725 Szeged, Hungary, E-mail: galambos@jgytf.u-szeged.hu

Gera, Zsolt: Institute of Informatics, University of Szeged, H-6701 Szeged P.O. Box 652, Hungary
E-mail: gera@inf.u-szeged.hu

Gianfranco, Pedone: E-mail: gianfranco.pedone@medcfs.hu

Gombás, Gábor: MTA SZTAKI, H-1518 Budapest, P. O. Box 63, Hungary
E-mail: gombas@sztaki.hu

Gyimóthy, Tibor: Institute of Informatics, University of Szeged, H-6701 Szeged P.O. Box 652, Hungary, E-mail: gyimi@inf.u-szeged.hu

Gönczy, László: Budapest University of Technology and Economics, H-1117 Budapest, Magyar Tudósok krt. 2., Hungary, E-mail: gonczy@mit.bme.hu

Győrbíró, Norbert: Institute of Informatics, University of Szeged, H-6701 Szeged P.O. Box 652, Hungary, E-mail: gyorbiro@inf.u-szeged.hu

Hegedűs, Hajnalka: Eötvös Loránd University, H-1117 Budapest, Hungary
E-mail: heha@aszt.inf.elte.hu

Horváth, Ákos: Budapest University of Technology and Economics, H-1111 Budapest, Goldman György tér 1., Hungary, E-mail: hor.akos@gmail.com

Horváth, Endre: Institute of Informatics, University of Szeged, H-6701 Szeged P.O. Box 652, Hungary
E-mail: hendre@inf.u-szeged.hu

Horváth, Gábor: Eötvös Loránd University, H-1117 Budapest, Hungary
E-mail: ghorvath@elte.hu

Horváth, Zoltán: Eötvös Loránd University, H-1117 Budapest, Hungary
E-mail: hz@inf.elte.hu

Imre, Gábor: Budapest University of Technology and Economics, H-1111 Budapest, Goldman György tér 1., Hungary, E-mail: gabor@aut.bme.hu

Jász, Judit: Institute of Informatics, University of Szeged, H-6701 Szeged P.O. Box 652, Hungary
E-mail: jasy@inf.u-szeged.hu

Jósvai, Eszter: Budapest University of Technology and Economics, H-1111 Budapest, Goldman György tér 1., Hungary, E-mail: eszter.josvai@gmail.com

Juhász, Zoltán: University of Veszprém, H-8200 Veszprém, Egyetem u. 10., Hungary
E-mail: juhasz@irt.vein.hu

Kacsuk, Péter: MTA SZTAKI, H-1518 Budapest, P. O. Box 63, Hungary
E-mail: kacsuk@sztaki.hu

Kertész, Attila: Institute of Informatics, University of Szeged, H-6701 Szeged P.O. Box 652, Hungary
E-mail: keratt@inf.u-szeged.hu

Kocsor, András: Research Group on Artificial Intelligence of the Hungarian Academy of Sciences and University of Szeged, H-6720 Szeged, Aradi vértanúk tere 1., Hungary
E-mail: kocsor@inf.u-szeged.hu

Kolářová, Edita: Ústav matematiky Fakulta elektrotechniky a komunikačních technologií Vysoké Učení Technické, Brno Česká Republika, E-mail: kolara@feec.vutbr.cz

Kovács, Előd: University of Veszprém, H-8200 Veszprém, Egyetem u. 10., Hungary
E-mail: kovacse@almos.vein.hu

Kovács, József: MTA SZTAKI, H-1518 Budapest, P. O. Box 63, Hungary
  E-mail: `smith@sztaki.hu`

Kovács, Kornél: Research Group on Artificial Intelligence of the Hungarian Academy of Sciences
  and University of Szeged, H-6720 Szeged, Aradi vértanúk tere 1., Hungary
  E-mail: `kkornel@inf.u-szeged.hu`

Kovács, Máté: Budapest University of Technology and Economics, H-1117 Budapest, Magyar Tudó-
  sok krt. 2., Hungary, E-mail: `kovmate@gmail.com`

Kozlovszky, Miklós: MTA SZTAKI, H-1518 Budapest, P. O. Box 63, Hungary
  E-mail: `m.kozlovszky@ieee.org`

Kozma, László: Eötvös Loránd University, H-1117 Budapest, Hungary
  E-mail: `kozma@ludens.elte.hu`

Kozsik, Tamás: Eötvös Loránd University, H-1117 Budapest, Hungary
  E-mail: `kto@inf.elte.hu`

Labádi, Máté: Institute of Informatics, University of Szeged, H-6701 Szeged P.O. Box 652, Hungary
  E-mail: `labadi@inf.u-szeged.hu`

Lengyel, László: Budapest University of Technology and Economics, H-1111 Budapest, Goldman
  György tér 1., Hungary, E-mail: `lengyel@aut.bme.hu`

Levendovszky, Tihamér: Budapest University of Technology and Economics, H-1111 Budapest,
  Goldman György tér 1., Hungary, E-mail: `tihamer@aut.bme.hu`

Le Viet, Dung: Budapest University of Technology and Economics, H-1111 Budapest, Goldman
  György tér 1., Hungary, E-mail: `leviet@tmit.bme.hu`

Lipovits, Ágnes: University of Veszprém, H-8200 Veszprém, Egyetem u. 10., Hungary
  E-mail: `lipovitsa@szt.vein.hu`

Lőrincz, László Csaba:

Lövei, László: Eötvös Loránd University, H-1117 Budapest, Hungary
  E-mail: `lovei@inf.elte.hu`

Marchis, Julia: Mathematics and Computer Science Faculty, Babes-Bolyai University, 400084 Cluj-
  Napoca, Kogalniceanu 1, E-mail: `imarchis@math.ubbcluj.ro`

Mezei, Gergely: Budapest University of Technology and Economics, H-1111 Budapest, Goldman
  György tér 1., Hungary, E-mail: `gmezei@aut.bme.hu`

Mészáros, Mónika: Eötvös Loránd University, H-1117 Budapest, Hungary
  E-mail: `bonnie@inf.elte.hu`

Michnay, Balázs: Institute of Informatics, University of Szeged, H-6701 Szeged P.O. Box 652, Hun-
  gary, E-mail: `michnay@inf.u-szeged.hu`

Miskolczi, János: Ericsson Magyarorszag Kft., E-mail: `janos.miskolczi@ericsson.com`

Móga, Rita: Budapest University of Technology and Economics, H-1111 Budapest, Goldman György
  tér 1., Hungary, E-mail: `moga@tmit.bme.hu`

Muzamel, Loránd: Institute of Informatics, University of Szeged, H-6701 Szeged P.O. Box 652, Hun-
  gary, E-mail: `muzamel@inf.u-szeged.hu`

Nagy, Tamás: Eötvös Loránd University, H-1117 Budapest, Hungary

Németh, Zsolt: Budapest University of Technology and Economics, H-1111 Budapest, Goldman György tér 1., Hungary, E-mail: `zsola@safranek.hu`

Oláh, István: Budapest University of Technology and Economics, H-1111 Budapest, Goldman György tér 1., Hungary, E-mail: `olah@tmit.bme.hu`

Paczolay, Dénes: Research Group on Artificial Intelligence of the Hungarian Academy of Sciences and University of Szeged, H-6720 Szeged, Aradi vértanúk tere 1., Hungary
E-mail: `pdenes@inf.u-szeged.hu`

Palágyi, Kálmán: Institute of Informatics, University of Szeged, H-6701 Szeged P.O. Box 652, Hungary, E-mail: `palagyi@inf.u-szeged.hu`

Palugyai, Sándor: Budapest University of Technology and Economics, H-1111 Budapest, Goldman György tér 1., Hungary
E-mail: `palugyai@tmit.bme.hu`, `sandor.palugyai@ericsson.com`

Pataki, Norbert: Eötvös Loránd University, H-1117 Budapest, Hungary
E-mail: `patakino@elte.hu`

Payrits, Szabolcs: Eötvös Loránd University, H-1117 Budapest, Hungary
E-mail: `quador@gmail.com`

Pócza, Krisztián: Eötvös Loránd University, H-1117 Budapest, Hungary
E-mail: `kpocza@kpocza.net`

Polyák, Tamás: Budapest University of Technology and Economics, H-1111 Budapest, Goldman György tér 1., Hungary, E-mail: `polyak@tmit.bme.hu`

Porkoláb, Zoltán: Eötvös Loránd University, H-1117 Budapest, Hungary
E-mail: `gsd@elte.hu`

Póth, Miklós: E-mail: `pmiki@vts.su.ac.yu`

Pozsgai, Tamás: University of Veszprém, H-8200 Veszprém, Egyetem u. 10., Hungary
E-mail: `pozsgai@almos.vein.hu`

Ráth, István: Budapest University of Technology and Economics, H-1111 Budapest, Goldman György tér 1., Hungary, E-mail: `istvan.rath@gmail.com`

Reinelt, Gerhard: Institute of Computer Science, University of Heidelberg
E-mail: `gerhard.reinelt@informatik.uni-heidelberg.de`

Sipos, Ádám: Eötvös Loránd University, H-1117 Budapest, Hungary, E-mail: `shp@elte.hu`

Szabó, Richárd: Eötvös Loránd University, H-1117 Budapest, Hungary
E-mail: `richard_szabo@invitel.hu`

Szaszák, György: Budapest University of Technology and Economics, H-1111 Budapest, Goldman György tér 1., Hungary, E-mail: `szaszak@tmit.bme.hu`

Szépkúti, István: ING Service Centre Budapest Ltd.H-1068 Budapest, Dózsa György út 84/b, Hungary, E-mail: `szepkuti@inf.u-szeged.hu`

Tejfel, Máté: Eötvös Loránd University, H-1117 Budapest, Hungary
E-mail: `matej@aszt.inf.elte.hu`

Teleki, Csaba: Budapest University of Technology and Economics, H-1111 Budapest, Goldman György tér 1., Hungary, E-mail: `teleki@tmit.bme.hu`

Theisz, Zoltán: Software Engineering Group, Ericsson Hungary Ltd., H-1300 Budapest P.O. Box 107, Hungary, E-mail: `Zoltan.Theisz@ericson.com`

Tóth, Dániel: Budapest University of Technology and Economics, H-1111 Budapest, Goldman György tér 1., Hungary, E-mail: `daniel.toth@gmail.com`

Tóth, Szabolcs L.: Budapest University of Technology and Economics, H-1111 Budapest, Goldman György tér 1., Hungary

Tóth, Krisztina: Research Group on Artificial Intelligence of the Hungarian Academy of Sciences and University of Szeged, H-6720 Szeged, Aradi vértanúk tere 1., Hungary
E-mail: `tothkr@inf.u-szeged.hu`

Vajk, István: Budapest University of Technology and Economics, H-1111 Budapest, Goldman György tér 1., Hungary, E-mail: `vajk@aut.bme.hu`

Vasutiu, Florina: Academy of Economic Studies, Bucuresti Faculty of Cybernetics, Statistics and Economic Informatics, E-mail: `fvasutiu@hotmail.com`

Vasutiu, Ovidiu: Institut National des Sciences Applique de Lyon Laboratoire d'InfoRmatique en Images et Systems d'information, E-mail: `ovidiu.vasutiu@cnedi69.cnafmail.fr`

Vágó, Dávid: Budapest University of Technology and Economics, H-1111 Budapest, Goldman György tér 1., Hungary, E-mail: `david.vago@t-online.hu`

Vidács, László: Institute of Informatics, University of Szeged, H-6701 Szeged P.O. Box 652, Hungary
E-mail: `lac@inf.u-szeged.hu`

Vicsi, Klára: Budapest University of Technology and Economics, H-1111 Budapest, Goldman György tér 1., Hungary, E-mail: `vicsi@tmit.bme.hu`

Víg, Anikó: Eötvös Loránd University, H-1117 Budapest, Hungary

Vincze, Gábor: Budapest University of Technology and Economics, H-1111 Budapest, Goldman György tér 1., Hungary, E-mail: `vincze@tmit.bme.hu`

Wu-Hen-Chang, Antal: Budapest University of Technology and Economics, H-1111 Budapest, Goldman György tér 1., Hungary, E-mail: `wuhen@tmit.bme.hu`

Zólyomi, István: Eötvös Loránd University, H-1117 Budapest, Hungary

Notes

Notes

Notes