

About Us

[Environment](#)
[Advisors](#)
[Doctoral students](#)
[Administrative office](#)

Program

[Structure](#)
[Preparatory phase](#)
[Dissertation phase](#)
[Admissions](#)
[Online Application](#)

Information

[Arrival](#)
[In Saarbrücken](#)
[On campus](#)
[Financial matters](#)

Events

[Social](#)
[Talks](#)
[Soft Skills](#)
[Guest program](#)



[Home](#) > [Doctoral students](#) > [PhD thesis defenses](#) > [PhD thesis defenses 2012](#)

2012

December

Rawia AWADALLAH

Methods for Constructing an Opinion Network Politically Controversial Topics

(Advisor: Prof. Gerhard Weikum)

Fri, 21 Dec 2012, 12:00h, building E1 4 (MPI-Inf), room 024

The US presidential race, the re-election of President Hugo Chavez, and the economic crisis in Greece and other European countries are some of the controversial topics being played on the news every day. To understand the landscape of opinions on political controversies, it would be helpful to know which politician or other stakeholder takes which position – support or opposition - on specific aspects of these topics. The work described in this thesis aims to automatically derive a map of the opinions-people network from news and other Web documents. The focus is on acquiring opinions held by various stakeholders on politically controversial topics. This opinions-people network serves as a knowledge-base of opinions in the form of <opinion holder><opinion><topic> triples. Our system to build this knowledge-base makes use of on-line news sources in order to extract opinions from text snippets. These sources come with a set of unique challenges. For example, processing text snippets involves not just identifying the topic and the opinion, but also attributing that opinion to a specific opinion holder. This requires making use of deep parsing and analyzing the parse tree. Moreover, in order to ensure uniformity, both the topic as well the opinion holder should be mapped to canonical strings, and the topics should also be organized into a hierarchy. Our system relies on two main components: i) acquiring opinions which uses a combination of techniques to extract opinions from on-line news sources, and ii) organizing topics which crawls and extracts debates from on-line sources, and organizes these debates in a hierarchy of political controversial topics. We present systematic evaluations of the different components of our system, and show their high accuracies. We also present some of the different kinds of applications that require political analysis. We present some application requires political analysis such as identifying flip-floppers, political bias, and dissenters. Such applications can make use of the knowledge-base of opinions.

Ernst Moritz HAHN

Model Checking Stochastic Hybrid Systems

(Advisor: Prof. Holger Hermanns)

Fri, 21 Dec 2012, 9:00h, building E1 7, room 001

The interplay of random phenomena with discrete-continuous dynamics deserves increased attention in many systems of growing importance. Their verification needs to consider both stochastic behaviour and hybrid dynamics. In the verification of classical hybrid systems, one is often interested in deciding whether unsafe system states can be reached. In the stochastic setting, we ask instead whether the probability of reaching particular states is bound by a given threshold. In this thesis, we consider stochastic hybrid systems and develop a general abstraction framework for deciding such problems. This gives rise to the first mechanisable technique that can, in practice, formally verify safety properties of systems which feature all the relevant aspects of nondeterminism, general continuous-time dynamics, and probabilistic behaviour. Being based tools for classical hybrid systems, future improvements in the effectivity of such tools directly carry over to improvements in the effectivity of our technique.

We extend the method in several directions: First, we discuss how we can handle continuous probability distributions. We then consider systems which we are in partial control of. Next, we consider systems in which probabilities are parametric, to analyse entire system families at once. Afterwards, we consider systems equipped with rewards, modelling costs or bonuses. Finally, we consider all orthogonal combinations of the extensions to the core model.

Ndapandula NAKASHOLE

Automatic Extraction of Facts, Relations, and Entities for Web-Scale Knowledge Base Population

(Advisor: Prof. Gerhard Weikum)

Thur, 20 Dec 2012, 15:15h, building E1 4 (MPI-Inf), room 024

Equipping machines with knowledge, through the construction of machine-readable knowledge bases, presents a key asset for semantic search, machine translation, question answering, and other formidable challenges in artificial intelligence. However, human knowledge predominantly resides in books and other natural language text forms. This means that knowledge bases must be extracted and synthesized from natural language text.

When the source of text is the Web, extraction methods must cope with ambiguity, noise, scale, and updates. The goal of this dissertation is to develop knowledge base population methods that address the afore mentioned characteristics of Web text. The dissertation makes three contributions. The first contribution is a method for mining high-quality facts at scale, through distributed constraint reasoning and a pattern representation model that is robust against noisy patterns. The second contribution is a method for mining a large comprehensive collection of relation types beyond those commonly found in existing knowledge bases. The third contribution is



a method for extracting facts from dynamic Web sources such as news articles and social media where one of the key challenges is the constant emergence of new entities. All methods have been evaluated through experiments involving Web-scale text collections.

Christian HAHN

A Platform-Independent Domain –Specific Modeling Language for Multiagent Systems

(Advisor: Prof. Jörg Siekmann)

Thur, 20 Dec 2012, 14:00h, building D3 4 (DFKI, VisCenter), room -1.63 NB

Associated with the increasing acceptance of agent-based computing as a novel software engineering paradigm, recently a lot of research addresses the development of suitable techniques to support the agent-oriented software development. The state-of-the-art in agent-based software development is to (i) design the agent systems basing on an agent-based methodology and (ii) take the resulting design artifact as a base to manually implement the agent system using existing agent-oriented programming languages or general purpose languages like Java. Apart from failures made when manually transform an abstract specification into a concrete implementation, the gap between design and implementation may also result in the divergence of design and implementation. The framework discussed in this dissertation presents a platform-independent domain-specific modeling language for MASs called DSML4MAS that allows modeling agent systems in a platform-independent and graphical manner. Apart from the abstract design, DSML4MAS also allows to automatically (i) check the generated design artifacts against a formal semantic specification to guarantee the well-formedness of the design and (ii) translate the abstract specification into a concrete implementation. Taking both together, DSML4MAS ensures that for any well-formed design, an associated implementation will be generated closing the gap between design and code.

Victor ALVAREZ

Selected Topics in Algorithmic Geometry

(Advisor: Prof. Raimund Seidel)

Thur, 20 Dec 2012, 11:00h, building E1 7, room 0.01

Let P be a set of n points on the plane with no three points on a line. A crossing-free structure on P is a straight-edge plane graph whose vertex set is P . In this thesis we consider problems of two different topics in the area of algorithmic geometry: Geometry using Steiner points, and counting algorithms. These topics have certain crossing-free structures on P as our primary objects of study. Our results can roughly be described as follows:

i) Given a k -coloring of P , with $k \geq 3$ colors, we will show how to construct a set of Steiner points $S = S(P)$ such that a k -colored quadrangulation can always be constructed on $(P \cup S)$. The bound we show of $|S|$ significantly improves on previously known results.

ii) We also show how to construct a set $S = S(P)$ of Steiner points such that a triangulation of $(P \cup S)$ having all its vertices of even (odd) degree can always be constructed. We show that $|S| \leq n/3 + c$, where c is a constant. We also look at other variants of this problem.

iii) With respect to counting algorithms, we show new algorithms for counting triangulations, pseudo-triangulations, crossing-free matchings and crossing-free spanning cycles on P . Our algorithms are simple and allow good analysis of their running times. These algorithms significantly improve over previously known results. We also show an algorithm that counts triangulations approximately, and a hardness result of a particular instance of the problem of counting triangulations exactly.

iv) We show experiments comparing our algorithms for counting triangulations with another well-known algorithm that is supposed to be very fast in practice.

Christoph ENDRES

PRESTK: Situation –Aware Presentation of Messages and Infotainment Content for Drivers

(Advisor: Prof. Wolfgang Wahlster)

Wed, 19 Dec 2012, 14:30h, building E1 1, room 407

The amount of in-car information systems has dramatically increased over the last few years. These potentially mutually independent information systems presenting information to the driver increase the risk of driver distraction.

In a first step, orchestrating these information systems using techniques from scheduling and presentation planning avoid conflicts when competing for scarce resources such as screen space. In a second step, the cognitive capacity of the driver as another scarce resource has to be considered. For the first step, an algorithm fulfilling the requirements of this situation is presented and evaluated. For the second step, I define the concept of System Situation Awareness (SSA) as an extension of Endsley's Situation Awareness (SA) model. I claim that not only the driver needs to know what is happening in his environment, but also the system, e.g., the car. In order to achieve SSA, two paths of research have to be followed: (1) Assessment of cognitive load of the driver in an unobtrusive way. I propose to estimate this value using a model based on environmental data. (2) Developing model of cognitive complexity induced by messages presented by the system. Three experiments support the claims I make in my conceptual contribution to this field.

A prototypical implementation of the situation-aware presentation management toolkit PRESTK is presented and shown in two demonstrators.

Manuel GORIUS

Adaptive Delay-constrained Internet Media Transport

(Advisor: Prof. Thorsten Herfet)

Wed, 19 Dec 2012, 13:30h, building C6 3, room 9.05

Reliable transport layer Internet protocols do not satisfy the requirements of packetized, real-time multimedia streams. The available thesis motivates and defines predictable reliability as a novel, capacity-approaching transport paradigm, supporting an application-specific level of reliability under a strict delay constraint. This paradigm is being implemented into a new protocol design -- the Predictably Reliable Real-time Transport protocol (PRRT).

In order to predictably achieve the desired level of reliability, proactive and reactive error control must be optimized under the application's delay constraint. Hence, predictably reliable error control relies on stochastic modeling of the protocol response to the modeled packet loss behavior of the network path. The result of the joined modeling is periodically evaluated by a reliability control policy that validates the protocol configuration under the application constraints and under consideration of the available network bandwidth. The adaptation of the protocol parameters is formulated into a combinatorial optimization problem that is solved by a fast search algorithm incorporating explicit knowledge about the search space. Experimental evaluation of PRRT in real Internet scenarios demonstrates that predictably reliable transport meets the strict QoS constraints of high-quality, audio-visual streaming applications.

David GÜNTHER

Topological Analysis of Discrete Scalar Data

(Advisor: Dr. Tino Weinkauff)

Tues, 18 Dec 2012, 10:15h, building E1 4 (MPI-Inf), room 0.19

This thesis presents a novel computational framework that allows for a robust extraction and quantification of the Morse-Smale complex of a scalar field given on a 2- or 3-dimensional manifold. The proposed framework is based on Forman's discrete Morse theory, which guarantees the topological consistency of the computed complex. Using a graph theoretical formulation of this theory, we present an algorithmic library that computes the Morse-Smale complex combinatorially with an optimal complexity of $O(n^2)$ and efficiently creates a multi-level representation of it. We explore the discrete nature of this complex, and relate it to the smooth counterpart. It is often necessary to estimate the feature strength of the individual components of the Morse-Smale complex – the critical points and separatrices. To do so, we propose a novel output-sensitive strategy to compute the persistence of the critical points. We also extend this well-founded concept to separatrices by introducing a novel measure of feature strength called separatrix persistence. We evaluate the applicability of our methods in a wide variety of application areas ranging from computer graphics to planetary science to computer and electron tomography.

Marc SCHLICKLING

Timing Model Derivation – Static Analysis of Hardware Description Languages

(Advisor: Prof. Reinhard Wilhelm)

Mon, 17 Dec 2012, 8:30h, building E1 7, room 0.01

Safety-critical hard real-time systems are subject to strict timing constraints. In order to derive guarantees on the timing behavior, the worst-case execution time (WCET) of each task comprising the system has to be known. The aiT tool has been developed for computing safe upper bounds on the WCET of a task. Its computation is mainly based on abstract interpretation of timing models of the processor and its periphery. These models are currently hand-crafted by human experts, which is a time-consuming and error-prone process. Modern processors are automatically synthesized from formal hardware specifications. Besides the processor's functional behavior, also timing aspects are included in these descriptions. A methodology to derive sound timing models using hardware specifications is described within this thesis. To ease the process of timing model derivation, the methodology is embedded into a sound framework.

A key part of this framework are static analyses on hardware specifications. This thesis presents an analysis framework that is build on the theory of abstract interpretation allowing use of classical program analyses on hardware description languages. Its suitability to automate parts of the derivation methodology is shown by different analyses. Practical experiments demonstrate the applicability of the approach to derive timing models. Also the soundness of the analyses and the analyses' results is proven.

Sandra EBERT

Semi-Supervised Learning for Image Classification

(Advisor: Prof. Bernt Schiele)

Fri, 14 Dec 2012, 11:00h, building E1 4 (MPI-Inf), room 0.24

Object class recognition is an active topic in computer vision still presenting many challenges. In most approaches, this task is addressed by supervised learning algorithms that need a large quantity of labels to perform well. This leads either to small datasets (< 10,000 images) that capture only a subset of the real-world class distribution (but with a controlled and verified labeling procedure), or to large datasets that are more representative but also add more label noise. Therefore, semi-supervised learning is a promising direction. It requires only few labels while simultaneously making use of the vast amount of images available today. In this thesis, we address object class recognition with semi-supervised learning. These algorithms depend on the underlying structure given by the data, the image description, and the similarity measure, and the quality of the labels. In the talk, I will illustrate this strong relationship between classification performance and the quality of the labels as well as the structure. Furthermore, I will show how we tackle these problems by active learning to get more representative labels and by several graph improvements.

November

Benny KNEISSL

Automated in silico Protein Modeling Strategies – Applications and Limits in G-Protein Coupled Receptor Modeling

(Advisor: Prof. Andreas Hildebrandt)

Fri, 30 Nov 2012, 11:00h, building E2 1, room 0.07

In the 21st century, computer-aided methods became a well-established tool in the pharmaceutical industry. Because many protein structures are still very difficult to determine experimentally, various in silico methods have been developed to predict their three-dimensional structure from the corresponding sequence. Unfortunately, the modeling process has often to be adapted to the target protein or the automatically generated models have to be manually refined afterwards by the users

based on their expert knowledge to finally achieve reasonable results. The aim of this thesis is to explore the applicability of automated in silico modeling strategies by means of G-protein coupled receptors (GPCRs). First, we analyze to which extent available protein structure prediction methods can be particularly tailored to the automated GPCR modelling case. Second, we develop our own approach and demonstrate that we obtain improved models compared to other state-of-the-art modeling tools. More important, our method does not rely on manual interactions by the user during the modeling process and is thus generally applicable for all GPCRs. Furthermore, we present a new sequence based method to predict structural distortions from ideal α -helical geometry. This method also exceeds the prediction accuracy of comparable approaches.

Pavel EMELIANENKO

Harnessing the Power of GPUs for Problem in real Algebraic Geometry

(Advisor: Prof. Kurt Mehlhorn)

Fri, 9 Nov 2012, 17:00h, building E1 4 (MPI-INF), room 0.24

This thesis presents novel parallel algorithms to leverage the power of GPUs (Graphics Processing Units) for exact computations with polynomials having large integer coefficients. The significance of such computations, especially in real algebraic geometry, is hard to undermine. On massively-parallel architectures such as GPU, the degree of data-level parallelism exposed by an algorithm is the main performance factor. We attain high efficiency through the use of structured matrix theory to assist the realization of relevant operations on polynomials on the graphics hardware. A detailed complexity analysis, assuming the PRAM model, also confirms that our approach achieves a substantially better parallel complexity in comparison to classical algorithms used for symbolic computations. Aside from the theoretical considerations, a large portion of this work is dedicated to the actual algorithm development and optimization techniques where we pay close attention to the specifics of the graphics hardware. As a byproduct of this work, we have developed high-throughput modular arithmetic which we expect to be useful for other GPU applications, in particular, open-key cryptography. We further discuss the algorithms for the solution of a system of polynomial equations, topology computation of algebraic curves and curve visualization which can profit to the full extent from the GPU acceleration. Extensive benchmarking on a real data demonstrates the superiority of our algorithms over several state-of-the-art approaches available to date. This thesis is written in English.

Andreas BAAK

Retrieval-based Approaches for Tracking and Reconstruction Human Motions

(Advisor: Prof. Meinard Müller)

Fri, 9 Nov 2012, 14:00h, building E1 4 (MPI-INF), room 0.19

In this thesis, we introduce novel approaches for increasing the stability, accuracy, and efficiency of marker-less human motion tracking and 3D human pose reconstruction. As one common underlying concept, the presented approaches contain a retrieval component making use of previously recorded marker-based motion capture (mocap) data. In particular, we contribute to three different areas dealing with various types of sensors including video cameras, optical mocap systems, inertial sensors, and depth cameras. Firstly, we introduce content-based retrieval techniques for automatically segmenting and annotating mocap data that is originally provided in form of unstructured data collections. Secondly, we show how such robust annotation procedures can be used to support and stabilize video-based marker-less motion tracking. Thirdly, we develop algorithms for reconstructing human motions from noisy depth sensor data in real-time. In all these contributions, a particular focus is put on efficiency issues in order to keep down the run time.

Peter Matthias GROSCHE

Signal Processing Methods for Beat Tracking, Music Segmentation and Audio Retrieval

(Advisor: Prof. Meinard Müller)

Fri, 9 Nov 2012, 12:00h, building E1 4 (MPI-INF), room 0.19

The goal of music information retrieval (MIR) is to develop novel strategies and techniques for organizing, exploring, accessing, and understanding music data in an efficient manner. The conversion of waveform-based audio data into semantically meaningful feature representations by the use of digital signal processing techniques is at the center of MIR and constitutes a difficult field of research because of the complexity and diversity of music signals. In this thesis, we introduce novel signal processing methods that allow for extracting musically meaningful information from audio signals. As main strategy, we exploit musical knowledge about the signals' properties to derive feature representations that show a significant degree of robustness against musical variations but still exhibit a high musical expressiveness. We apply this general strategy to three different areas of MIR: Firstly, we introduce novel techniques for extracting tempo and beat information, where we particularly consider challenging music with changing tempo and soft note onsets. Secondly, we present novel algorithms for the automated segmentation and analysis of folk song field recordings, where one has to cope with significant fluctuations in intonation and tempo as well as recording artifacts. Thirdly, we explore a cross-version approach to content-based music retrieval based on the query-by-example paradigm. In all three areas, we focus on application scenarios where strong musical variations make the extraction of musically meaningful information a challenging task.

Verena KONZ

Automated Methods for Audio – Based Music Analysis with Applications to Musicology

(Advisor: Prof. Meinard Müller)

Fri, 9 Nov 2012, 10:00h, building E1 4 (MPI-INF), room 0.19

This thesis contributes to bridging the gap between music information retrieval (MIR) and musicology. We present several automated methods for music analysis, which are motivated by concrete application scenarios being of central importance in musicology. In this context, the automated music analysis is performed on the basis of audio material. Here, one reason is that for a given piece of music usually many different recorded performances exist. The availability of multiple versions of a piece of music is exploited in this thesis to stabilize analysis results. We show how the presented automated methods open up new possibilities for supporting

musicologists in their work. Furthermore, we introduce novel interdisciplinary concepts which facilitate the collaboration between computer scientists and musicologists. Based on these concepts, we demonstrate how MIR researchers and musicologists may greatly benefit from each other in an interdisciplinary collaboration. Firstly, we present a fully automatic approach for the extraction of tempo parameters from audio recordings and show to which extent this approach may support musicologists in analyzing recorded performances. Secondly, we introduce novel user interfaces which are aimed at encouraging the exchange between computer science and musicology. In this context, we indicate the potential of computer-based methods in music education by testing and evaluating a novel MIR user interface at the University of Music Saarbrücken. Furthermore, we show how a novel multi-perspective user interface allows for interactively viewing and evaluating version-dependent analysis results and opens up new possibilities for interdisciplinary collaborations. Thirdly, we present a cross-version approach for harmonic analysis of audio recordings and demonstrate how this approach enables musicologists to explore harmonic structures even across large music corpora. Here, one simple yet important conceptual contribution is to convert the physical time axis of an audio recording into a performance-independent musical time axis given in bars.

Patrick WISCHNEWSKI

Efficient Reasoning Procedures for Complex First-Order Theories

(Advisor: Prof. Christoph Weidenbach)

Tues, 6 Nov 2012, 10:00h, building E1 5 (MPI-SWS), room HS 002

The complexity of a set of first-order formulas results from the size of the set and the complexity of the problem described by its formulas.

Decision Procedures for Ontologies

This thesis presents new superposition based decision procedures for large sets of formulas. The sets of formulas may contain expressive constructs like transitivity and equality. The procedures decide the consistency of knowledge bases, called ontologies, that consist of several million formulas and answer complex queries with respect to these ontologies. They are the first superposition based reasoning procedures for ontologies that are at the same time efficient, sound, and complete. The procedures are evaluated using the well-known ontologies YAGO, SUMO and CYC. The results of the experiments, which are presented in this thesis, show that these procedures decide the consistency of all three above-mentioned ontologies and usually answer queries within a few seconds.

Reductions for General Automated Theorem Proving

Sophisticated reductions are important in order to obtain efficient reasoning procedures for complex, particularly undecidable problems because they restrict the search space of theorem proving procedures. In this thesis, I have developed a new powerful reduction rule. This rule enables superposition based reasoning procedures to find proofs in sets of complex formulas. In addition, it increases the number of problems for which superposition is a decision procedure.

Hans-Jörg PETER

A Uniform Approach to the Complexity and Analysis of Succinct Systems

(Advisor: Prof. Bernd Finkbeiner)

Fri, 2 Nov 2012, 16:00h, building E1 7, room 0.01

This thesis provides a unifying view on the succinctness of systems:

the capability of a modeling formalism to describe the behavior of a system of exponential size using a polynomial syntax.

The key theoretical contribution is the introduction of sequential circuit machines as a new universal computation model that focuses on succinctness as the central aspect.

The thesis demonstrates that many well-known modeling formalisms such as communicating state machines, linear-time temporal logic, or timed automata exhibit an immediate connection to this machine model.

Once a (syntactic) connection is established, many complexity bounds for resource-restricted sequential circuit machines can be transferred to a certain formalism in a uniform manner.

As a consequence, besides a far-reaching unification of independent lines of research, we are also able to provide matching complexity bounds for various analysis problems, whose complexities were not known so far.

For example, we establish matching lower and upper bounds of the small witness problem and several variants of the bounded synthesis problem for timed automata, a particularly important succinct modeling formalism.

Also for timed automata, our complexity-theoretic analysis leads to the identification of tractable fragments of the timed synthesis problem under partial observability.

Specifically, we identify timed controller synthesis based on discrete or template-based controllers to be equivalent to model checking.

Based on this discovery, we develop a new model checking-based algorithm to efficiently find feasible template instantiations.

From a more practical perspective, this thesis also studies the preservation of succinctness in analysis algorithms using symbolic data structures.

While efficient techniques exist for specific forms of succinctness considered in isolation, we present a general approach based on abstraction refinement to combine off-the-shelf symbolic data structures.

In particular, for handling the combination of concurrency and quantitative timing behavior, we report on the tool Synthia which combines binary decision diagrams with difference bound matrices.

October

Sebastian ALTMAYER

Analysis of Preemptively Scheduled Hard Real-time Systems

(Advisor: Prof. Reinhard Wilhelm)

Thurs, 25 Oct 2012, 14:15h, building E1 1, room 407

As timing is a major property of hard real-time, proving timing correctness is of utter importance. A static timing analysis derives upper bounds on the execution time of tasks, a scheduling analysis uses these bounds and checks if each task meets its timing constraints. In preemptively scheduled systems with caches, this interface between timing analysis and scheduling analysis must be considered outdated. On a

context switch, a preempting task may evict cached data of a preempted task that need to be reloaded again after preemption. The additional execution time due to these reloads, called cache-related preemption delay (CRPD), may substantially prolong a task's execution time and strongly influence the system's performance. In this thesis, we present a formal definition of the cache-related preemption delay and determine the applicability and the limitations of a separate CRPD computation. To bound the CRPD based on the analysis of the preempted task, we introduce the concept of definitely cached useful cache blocks. This new concept eliminates substantial pessimism with respect to former analyses by considering the over-approximation of a preceding timing analysis. We consider the impact of the preempting task to further refine the CRPD bounds. To this end, we present the notion of resilience. The resilience of a cache block is a measure for the amount of disturbance of a preempting task a cache block of the preempted task may survive. Based on these CRPD bounds, we show how to correctly account for the CRPD in the schedulability analysis for fixed-priority preemptive systems and present new CRPD-aware response time analyses: ECB-Union and Multiset approaches.

Linlin LI

Computational Modeling of Lexical Ambiguity

(Advisor: Dr. Caroline Sporleder)

Wed, 10 Oct 2012, 15:00h, C7 4, conference room

Lexical ambiguity is a frequent phenomenon that can occur not only for words but also on the phrase level. Natural language processing systems need to efficiently deal with these ambiguities in various tasks, however, we often encounter such system failures in real applications. This thesis aims to study various lexical ambiguity phenomena from a computational modeling perspective and advance the state-of-the-art research.

Andreas BROSCART

Efficient Query Processing and Index Tuning using Proximity Scores

(Advisor: Prof. Gerhard Weikum)

Tues, 9 Oct 2012, 11:00h, E1 4 (MPI-Inf), room 0.24

In the presence of growing data, the need for efficient query processing under result quality and index size control becomes more and more a challenge to search engines. We show how to use proximity scores to make query processing effective and efficient with focus on either of the optimization goals.

More precisely, we make the following contributions:

- * We present a comprehensive comparative analysis of proximity score models and a rigorous analysis of the potential of phrases and adapt a leading proximity score model for XML data

- * We discuss the feasibility of all presented proximity score models for top-k query processing and present a novel index combining a content and proximity score that helps to accelerate top-k query processing and improves result quality.

- * We present a novel, distributed index tuning framework for term and term pair index lists that optimizes pruning parameters by means of well-defined optimization criteria under disk space constraints. Indexes can be tuned with emphasis on efficiency or effectiveness: the resulting indexes yield fast processing at high result quality.

- * We show that pruned index lists processed with a merge join outperform top-k query processing with unpruned lists at a high result quality.

- * Moreover, we present a hybrid index structure for improved cold cache run times.

September

Laura TOLOSI

Finding Regions of Abberant DNA Copy Number Associated with Tumor Phenotype

(Advisor: Prof. Thomas Lengauer)

Thurs, 27 Sept 2012, 10:00h, E1 4 (MPI-Inf), room 0.24

DNA copy number alterations are a hallmark of cancer. Understanding their role in tumor progression can help improve diagnosis, prognosis and therapy selection for cancer patients. High-resolution, genome-wide measurements of DNA copy number changes for large cohorts of tumors are currently available, owing to technologies like microarray-based array comparative hybridization (arrayCGH). In this thesis, we present a computational pipeline for statistical analysis of tumor cohorts, which can help extract relevant patterns of copy number aberrations and infer their association with various phenotypical indicators.

The main challenges are the instability of classification models due to the high dimensionality of the arrays compared to the small number of tumor samples, as well as the large correlations between copy number estimates measured at neighboring loci. We show that the feature ranking given by several widely-used methods for feature selection is biased due to the large correlations between features. In order to correct for the bias and instability of the feature ranking, we introduce methods for consensus segmentation of the set of arrays. We present three algorithms for consensus segmentation, which are based on identifying recurrent DNA breakpoints or DNA regions of constant copy number profile. The segmentation constitutes the basis for computing a set of super-features, corresponding to the regions. We use the super-features for supervised classification and we compare the models to baseline models trained on probe data. We validated the methods by training models for prediction of the phenotype of breast cancers and neuroblastoma tumors. We show that the multivariate segmentation affords higher model stability, in general improves prediction accuracy and facilitates model interpretation. One of our most important biological results refers to the classification of neuroblastoma tumors. We show that patients belonging to different age subgroups are characterized by distinct copy number patterns, with largest discrepancy when the subgroups are defined as older or younger than 16-18 months. We thereby confirm the recommendation for a higher age cutoff than 12 months (current clinical practice) for differential diagnosis of neuroblastoma.

Markus PISTER
Timing Model Derivation – Pipeline Analyzer Generation from Hardware Description Languages

(Advisor: Prof. Reinhard Wilhelm)

Wed, 19 Sept 2012, 14:30h, E1 7, room 0.01 (ground floor, Cluster building)

Safety-critical systems are forced to finish their execution within strict deadlines so that worst-case execution time (WCET) guarantees are a crucial part of their verification. Timing models of the analyzed hardware form the basis for static analysis-based approaches like the aiT WCET analyzer. Currently, timing models are hand-crafted based on frequently incorrect documentation causing the process to be error-prone and time-consuming. This thesis bridges the gap between automatic hardware synthesis and WCET analysis development by introducing a process for the derivation of timing models from VHDL specifications. We propose a set of transformations and abstractions to reduce the hardware design's complexity enabling the generation of efficient and provably correct WCET analyzers. They employ an abstract interpretation-based simulation of program executions based on a defined abstract simulation semantics. We have defined workflow patterns showing how to gradually apply the derivation process to VHDL models, thereby removing timing-irrelevant constructs. Interval property checking is used to validate the transformations. A further contribution of this thesis is the implementation of a tool set that realizes the introduced derivation process and shows its applicability to non-trivial industrial designs in experimental evaluations.

Influences on design choices to the quality of the derived timing model are presented building an informal predictability notion for VHDL.

Stefan POPOV

Algorithms and Data Structures for Interactive Ray Tracing on Commodity Hardware

(Advisor: Prof. Philipp Slusallek)

Tues, 18 Sept 2012, 11:00h, in The Visualization Center, DFKI (ground floor, new building)

Rendering methods based on ray tracing provide high image realism, but have been historically regarded as offline only. This has changed in the past decade, due to significant advances in the construction and traversal performance of acceleration structures and the efficient use of data-parallel processing. Today, all major graphics companies offer real-time ray tracing solutions. Our work, which we present in this talk, has contributed to this development with some key insights.

In the first part of the talk we focus on the GPU as a means to bring real-time ray tracing to desktop computers. There, we will briefly discuss our GPU friendly KD-tree and BVH traversal algorithms.

We will then address the limited support of dynamic scenes in previous work, by presenting two novel parallel-friendly construction algorithms for KD-trees and BVHs. By approximating the cost function, we accelerate construction by up to an order of magnitude (especially for BVHs), at the expense of only tiny degradation to traversal performance.

Finally, we address the topic of creating the "perfect" acceleration structure. We show how to develop a polynomial time non-greedy BVH construction algorithm. We then modify it to produce a new type of acceleration structure that inherits both the high performance of KD-trees and the small size of BVHs.

August

Andrey SHADRIN

Mixed Low- and High Level Programming Language Semantics and Automated Verification of a Small Hypervisor

(Advisor: Prof. Wolfgang Paul)

Mon, 27 August 2012, 15:00h, in building E1 7, room 001

Hypervisors are system software programs that virtualize the architecture they run on and are usually implemented in a mix of (macro) assembly and a high-level language like C. To verify such a software, assembly parts as well as C parts should be verified, where reasoning about those parts is done in different semantics. At the end, both semantics should be brought together in an overall correctness theorem of such a software program. The formal integration of correctness results accomplished in distinct semantics is challenging but inevitable for systems verification. This thesis is split into two parts. In the first one, we will present the mixed semantics of C and macro assembly. This semantics can handle mixed-language implementations where the execution context is changed by an external function call from assembly to C and vice versa. Also, we state a step-by-step simulation theorem between mixed programs and the compiled and assembled code. In the second part, we present the correctness of a small hypervisor, called Baby Hypervisor (BHV), described by the mixed semantics. BHV virtualizes a 32-bit RISC architecture. The BHV functional verification was shown using Microsoft's VCC, an automatic verifier for C with contracts. BHV C portions were verified with VCC. For making macro assembly feasible with VCC the original macro assembly is translated to C code simulating processor (simulation approach).

Alekh JINDAL

OctopusDB: Flexible and Scalable Storage Management for Arbitrary Database Engines

(Advisor: Prof. Jens Dittrich)

Fri, 24 August 2012, 14:30, in building E1 1, room 3.06

We live in a dynamic age with the economy, the technology, and the people around us changing faster than ever before. Consequently, the data management needs in our modern world are much different than those envisioned by the early database inventors in the 70s. Today, enterprises face the challenge of managing ever-growing dataset sizes with dynamically changing query workloads. The current practice to deal with changing query workloads is to have a different specialized product for each workload type, e.g. row stores for OLTP workload, column stores for OLAP workload, streaming systems for streaming workload, and scan-oriented systems for shared query processing. However, this has the additional penalty of managing a zoo of data managing systems in the first place, which is tedious, expensive, as well as counter-productive for modern enterprises. This thesis presents an alternative approach to supporting several query workloads in a data

managing system. We observe that each specialized database product has a different data store, indicating that different query workloads work well with different data layouts. Thus, a key requirement for supporting several query workloads is to support several data layouts. Therefore, in this thesis, we study ways to inject different data layouts into existing (and familiar) data managing systems. The goal is to develop a flexible storage layer which can support several query workloads in a single data managing system. We present a set of non-invasive techniques, coined Trojan Techniques, to inject different data layouts into a data managing system. Trojan Techniques are non-invasive in the sense that they do not make heavy untenable changes to the system and yet they bring significant improvements in query performance. Our experimental results show that Trojan Techniques can improve the performance of Hadoop MapReduce by a factor of up to 18, and that of a top-notch commercial database product by a factor of up to 17.

Dimitar DENEV

Models and Methods for Web Archive Crawling

(Advisor: Prof. Gerhard Weikum)

Mon, 20 August 2012, 15:00h, in building E1 4, room 024

Web archives offer a rich and plentiful source of information to researchers, analysts, and legal experts. For this purpose, they gather Web sites as the sites change over time. In order to keep up to high standards of data quality, Web archives have to collect all versions of the Web sites. Due to limited resources and technical constraints this is not possible. Therefore, Web archives consist of versions archived at various time points without guarantee for mutual consistency. This thesis presents a model for assessing the data quality in Web archives as well as a family of crawling strategies yielding high-quality captures. We distinguish between single-visit crawling strategies for exploratory and visit-revisit crawling strategies for evidentiary purposes. Single-visit strategies download every page exactly once aiming for an “undistorted” capture of the ever-changing Web. We express the quality of such the resulting capture with the “blur” quality measure. In contrast, visit-revisit strategies download every page twice. The initial downloads of all pages form the visit phase of the crawling strategy. The second downloads are grouped together in the revisit phase. These two phases enable us to check which pages changed during the crawling process. Thus, we can identify the pages that are consistent with each other. The quality of the visit-revisit captures is expressed by the “coherence” measure. Quality-conscious strategies are based on predictions of the change behaviour of individual pages. We model the Web site dynamics by Poisson processes with page specific change rates. Furthermore, we show that these rates can be statistically predicted. Finally, we propose visualization techniques for exploring the quality of the resulting Web archives. A fully functional prototype demonstrates the practical viability of our approach.

Piotr DIDYK

Perceptual Display: Exceeding Display Limitations by Exploiting the Human Visual System

(Advisor: Prof. Hans-Peter Seidel)

Mon, 20 August 2012, 16:30h, in building E1 4 (MPI-Inf), room 019

Existing displays have a number of limitations, which make it difficult to realistically reproduce real-world appearance; discrete pixels are used to represent images, which are refreshed only a limited number of times per second, the output luminance range is much smaller than in the real world, and only two dimensions are available to reproduce a three-dimensional scene. On the other hand, the human visual system has its own limitations, as those imposed by the density of photoreceptors, imperfections in the eye optics, or the limited ability to discern high-frequency information. In this talk, I will show that taking these limitations into account and using perceptual effects, which very often are not measurable physically, allow us to design methods which can overcome the physical limitations of display devices in order to enhance apparent image qualities. More precisely, I will discuss how high quality frames can be interleaved with low quality frames improving the sharpness of rendered sequence. Further, I will present an optimization technique which produces frames that shown in a rapid succession lead to an apparent increase in spatial resolution, and, finally, I will talk about the role of perception in context of stereovision.

July

Paul LIBBRECHT

Authoring of Semantic Mathematical Content for Learning on the Web

(Advisor: Prof. Jörg Siekmann)

Mon, 30 July 2012, 15:15h, in building E1 3, HS 001

This thesis explores how a mathematician can create content which lets students learn via ActiveMath. This software runs on the web, and employs semantics mathematical technologies and artificial intelligence to support the student. The thesis describes not only a tool to edit content but also studies the workflows of authors when creating content.

An important authoring paradigm introduced here is called WYCIWYG – What You Check is What You Get. This places the target learning environment as the central concern. Indeed, the learning environment, which is the assembly of content and software, represents the final product of the authoring work. This thesis exploits and develops classical paradigms: copy-and-paste, search, and collaborative authoring. To support the authoring workflows, several software pieces have been implemented and tested: within an editing tool and the ActiveMath software. The following work describes their usages, their technical facets, and case studies of how current authors have used them in their day-to-day activities.

Mahmoud FOUZ

Randomized Rumor Spreading in Social Networks & Complete Graphs

(Advisor: Prof. Markus Bläser)

Mon, 16 July 2012, 9:00h, in building E1 4 (MPI-Inf), room 024

This thesis deals with two rumor spreading problems. In the first part, we study the rumor spreading problem on social networks modeled by preferential attachment

graphs. We consider the push-pull strategy of Karp et al., where in each round, each vertex chooses a random neighbor and exchanges information with it. We prove the following. The push-pull strategy delivers a message to all nodes within $\Theta(\log n)$ rounds with high probability, where n is the number of nodes in the graph.

The best known bound so far was $O(\log^2 n)$ by Chierichetti, Lattanzi, and Panconesi [ICALP 2009]. If we slightly modify the protocol so that contacts are chosen uniformly from all neighbors but the one contacted in the previous round, then this time reduces to $\Theta(\log n / \log \log n)$. This is asymptotically optimal since it matches the diameter of the graph. In an asynchronous version of the protocol, the running time is shown to be even $O(\log n)$.

In the second part, we consider the rumor spreading problem on the complete graph. We propose a new push protocol that achieves an asymptotically optimal time of $(1 + o(1)) \log n$. It needs only $O(n f(n))$ calls, where $f(n) = \omega(1)$ can be arbitrary. The protocol is also robust against random node failures. We further extend it to deal with adversarial node failures efficiently

Yana MILEVA

Mining the Evolution of Software Component Usage

(Advisor: Prof. Andreas Zeller)

Mon, 9 July 2012, 16:00h, in building E1 1, room 407

The main focus of this thesis is the analysis of the evolution of software components. In order to track the evolution of software components, one needs to collect the evolution information of each component. This information is being stored in the version control system (VCS) of the project—the repository of the history of events happening throughout the project's lifetime. By using software archive mining techniques one can extract and leverage this information. The main contribution of this thesis is the introduction of evolution usage trends and evolution change patterns. The raw information about the occurrences of each component is being stored in the VCS of the project and by organizing it in evolution trends and patterns, we are able to draw conclusions and issue recommendations concerning each individual component and the project as a whole.

Evolution Trends An evolution trend is a way to track the evolution of a software component throughout the span of the project. The trend shows the increases and decreases in the usage of a specific component, which can be indicative of the quality of this component. AKTARI is a tool, presented in this thesis, that is based on such evolution trends and can be used by the software developers to observe and draw conclusions about the behavior of their project.

Evolution Patterns We call an evolution pattern, the pattern of a frequently occurring code change throughout the span of the project. Those frequently occurring changes are project-specific and are explanatory of the way the project evolves. Each such change evolution pattern contains in itself the specific way “things are done” in the project and as such can serve for defect detection and defect prevention. This technique is being implemented as a basis for the LAMARCK tool, presented in this thesis.

Pascal GWOSDEK

Hardware Accelerated Algorithms in Visual Computing

(Advisor: Prof. Joachim Weickert)

Wed, 11 July 2012, 14:15h, in building E1 1, room 407

This talk presents parallel algorithms for visual computing on graphics cards, and evaluates them with respect to their speed, scalability, and quality of the results. This includes homogeneous and anisotropic diffusion processes, diffusion image inpainting, optic flow, and halftoning.

We compare different solvers for homogeneous diffusion and present a novel 'extended' box filter. Moreover, we suggest to use the fast explicit diffusion scheme (FED) as an efficient and flexible solver for nonlinear and in particular for anisotropic parabolic diffusion problems on graphics hardware. For elliptic diffusion-like processes, we recommend to use cascadic FED or fast Jacobi schemes. Our presented optic flow algorithm represents the fastest method among the 30 most accurate techniques. A novel halftoning scheme produces state-of-the-art techniques, and has many applications in image processing and computer graphics.

June

Stephan WILHELM

Symbolic Representations in WCET Analysis

(Advisor: Prof. Reinhard Wilhelm)

Mon, 4 June 2012, 16:00h, in building E1 1, room 407

Reliable task-level execution time information is indispensable for validating the correct operation of safety-critical embedded real-time systems. Static worst-case execution time (WCET) analysis is a method that computes safe upper bounds of the execution time of single uninterrupted tasks. The method is based on abstract interpretation and involves abstract hardware models that capture the timing behavior of the processor on which the tasks run. For complex processors, task-level execution time bounds are obtained by a state space exploration which involves the abstract model and the program. Partial state space exploration is not sound. A full exploration can become too expensive. Symbolic state space exploration methods using binary decision diagrams (BDDs) are known to provide efficient means for covering large state spaces.

This work presents a symbolic method for the efficient state space exploration of abstract pipeline models in the context of static WCET analysis. The method has been implemented for the Infineon TriCore 1 which is a real-life processor of medium complexity. Experimental results on a set of benchmarks and an automotive industry application demonstrate that the approach improves the scalability of static WCET analysis while maintaining soundness.

April

Tom CRECELIUS

Socially Enhanced Search and Explorations in Social Tagging Networks

Social tagging networks have become highly popular for publishing and searching contents. Users in such networks can review, rate and comment on contents, or annotate them with keywords (social tags) to give short but exact text representations of even non-textual contents. Moreover, there is an inherent support for interactions and relationships among users and, thus, naturally form groups of friends or of common interests.

This dissertation addresses in this area three research topics, namely:

- (i) A comprehensive framework for search in social tagging networks and efficient top-k result retrieval.
- (ii) A new approach of solving the APSD Problem for dynamically update friendship relations in large user networks.
- (iii) Countering cheating in P2P authority computations over social networks.

Weiwei SUN

Learning Chinese Language Structures with Multiple Views

(Advisor: Prof. Hans Uszkoreit)

Tues, 17 April 2012, 16:00h, in building B3 1, room 0.11

Motivated by the inadequacy of single view approaches in many areas in NLP, this thesis studies multi-view Chinese language processing, including word segmentation, part-of-speech tagging, syntactic parsing and semantic role labelling. We consider three situations of multiple views in statistical NLP:

- (1) Heterogeneous computational models have been designed for a given problem;
- (2) Heterogeneous annotation data is available to train systems;
- (3) Supervised and unsupervised machine learning techniques are applicable.

In practice, we study general approaches for model ensemble, annotation ensemble and learning ensemble problems in Chinese language processing. First, we comparatively analyze successful single view approaches for Chinese lexical, syntactic and semantic processing. Our analysis highlights the diversity between heterogeneous systems built on different views, and motivates us to improve the state-of-the-art by combining or integrating heterogeneous approaches. Second, we study the annotation ensemble problem, i.e. learning from multiple data sets under different annotation standards. We propose a series of generalized stacking models to effectively utilize heterogeneous labelled data to reduce approximation errors for word segmentation and parsing. Finally, we are concerned with bridging the gap between unsupervised and supervised learning paradigms. We introduce feature induction solutions that harvest useful linguistic knowledge from large-scale unlabelled data and effectively use them as new features to enhance discriminative learning based systems. Our empirical study indicates that multi-view processing is beneficial, especially when diverse single view approaches have complementary strengths, and our stacking-inspired ensemble approaches are very effective to explore multiple views, at least for learning Chinese language structures.

March

Christian SCHMALTZ

Compression, Pose Tracking, and Halftoning

(Advisor: Prof. Joachim Weickert)

23 March 2012, 16:15h, in building E1 7, room 0.01

In this thesis, we discuss image compression, pose tracking, and halftoning. Although these areas seem to be unrelated at first glance, they can be connected through video coding as application scenario. Our first contribution is an image compression algorithm based on a rectangular subdivision scheme which stores only a small subsets of the image points. From these points, the remained of the image is reconstructed using partial differential equations. Afterwards, we present a pose tracking algorithm that is able to follow the 3-D position and orientation of multiple objects simultaneously. The algorithm can deal with noisy sequences, and naturally handles both occlusions between different objects, as well as occlusions occurring in kinematic chains. Our third contribution is a halftoning algorithm based on electrostatic principles, which can easily be adjusted to different settings through a number of extensions. Examples include modifications to handle varying dot sizes or hatching. In the final part of the thesis, we show how to combine our image compression, pose tracking, and halftoning algorithms to novel video compression codecs. In each of these four topics, our algorithms yield excellent results that outperform those of other state-of-the-art algorithms.

Anna Katharina DEHOF

Novel Approaches for Bond Assignment and NMR Shift Prediction

(Advisor: Prof. Hans-Peter Lenhof)

Thu, 15 Mar 2012, 15:00 h, in building E2 1, room 001

Methoden des molekularen Modellierens gehören zu den Grundpfeilern moderner biologischer und pharmazeutischer Forschung. Akkurate Modellierung-Methoden erfordern jedoch enormen Rechenaufwand, weshalb üblicherweise verschiedene Näherungsverfahren eingesetzt werden. Im Promotionsvortrag werden solche im Rahmen der Promotion entwickelten Näherungen für verschiedene Probleme aus der strukturbasierten Bioinformatik vorgestellt. Eine fundamentale Näherung der molekularen Physik ist die Einteilung chemischer Bindungen in wenige Klassen, meist in Form ganzzahliger Bindungsordnungen. In vielen Datensätzen ist diese Information nicht enthalten und eine automatische Zuweisung ist hochgradig schwierig. Für diese Problemstellung wird die BOA Constructor-Methode vorgestellt, die sowohl mit uneindeutigen Lösungen umgehen kann als auch vom Benutzer leicht erweitert werden kann. In umfangreichen Tests zeigen wir, dass unsere Methode dem bisherigen Stand der Forschung überlegen ist. Neben klassischen Anwendungen liefern Bindungsordnungen wertvolle Informationen für die statistische Vorhersage molekularer Eigenschaften wie z.B. der chemischen

Verschiebung von Proteinatomen. Mit der von uns entwickelten NightShift-Pipeline wird ein Verfahren zur automatischen Generierung von Vorhersagemodellen präsentiert, wie z.B. dem Spinster-Modell, das den bisherigen manuell entwickelten Verfahren überlegen ist. Die Kombination mit BOA Constructor führt zum sogenannten Liops-Modell, welches als erstes Modell die effiziente Berücksichtigung des Einflusses von nicht-Proteinatomen erlaubt.

February

Hennig Lars ZIMMER

Correspondence Problems in Computer Vision

(Advisor: Prof. Joachim Weickert)

Fri, 10 Feb 2012, 16:15 h, in building E1 1, room 3.06

Correspondence problems like optic flow belong to the fundamental problems in computer vision. Here, one aims at finding correspondences between the pixels in two (or more) images. The correspondences are described by a displacement vector field that is often found by minimising an energy (cost) function. In this thesis, we present several contributions to the energy-based solution of correspondence problems:

(i) We start by developing a robust data term with a high degree of invariance under illumination changes. Then, we design an anisotropic smoothness term that works complementary to the data term, thereby avoiding undesirable interference. Additionally, we propose a simple method for determining the optimal balance between the two terms.

(ii) When discretising image derivatives that occur in our continuous models, we show that adapting one-sided upwind discretisations from the field of hyperbolic differential equations can be beneficial. To ensure a fast solution of the nonlinear system of equations that arises when minimising the energy, we use the recent fast explicit diffusion (FED) solver in an explicit gradient descent scheme.

(iii) Finally, we present a novel application of modern optic flow methods where we align exposure series used in high dynamic range (HDR) imaging. Furthermore, we show how the alignment information can be used in a joint super-resolution and HDR method.

Shady ELBASSUONI

Effective Searching of RDF Knowledge Bases

(Advisor: Prof. Gerhard Weikum)

Mon, 6 Feb 2012, 15:00 h, in building E1 4 (MPI), room 024

RDF data has become a vital source of information for many applications. In this thesis, we present a set of models and algorithms to effectively search large RDF knowledge bases. These knowledge bases contain a large set of subject-predicate-object (SPO) triples where subjects and objects are entities and predicates express relationships between them. Searching such knowledge bases can be done using the W3C-endorsed SPARQL language or by similarly designed triple-pattern search. However, the exact-match semantics of triple-pattern search might fall short of satisfying the users needs by returning too many or too few results. Thus, IR-style searching and ranking techniques are crucial.

This thesis develops models and algorithms to enhance triple-pattern search. We propose a keyword extension to triple-pattern search that allows users to augment triple-pattern queries with keyword conditions. To improve the recall of triple-pattern search, we present a framework to automatically reformulate triple-pattern queries in such a way that the intention of the original user query is preserved while returning a sufficient number of ranked results. For efficient query processing, we present a set of top-k query processing algorithms and for ease of use, we develop methods for plain keyword search over RDF knowledge bases. Finally, we propose a set of techniques to diversify query results and we present several methods to allow users to interactively explore RDF knowledge bases to find additional contextual information about their query results.

Ulan DEGENBAEV

Formal Specification of the x86 Instruction Set Architecture

(Advisor: Prof. Wolfgang Paul)

Mon, 6 Feb 2012, 14:00 h, in building E1 3, room HS 003

In this thesis we specify the x86 instruction set architecture (ISA) by developing an abstract machine that models the behaviour of a modern computer with multiple x86 processors. Our model enables reasoning about low-level system software by providing formal interpretation of thousand pages of the processor vendor documentation written in informal prose.

We show how to reduce the problem of ISA formalization to two simpler problems: memory model specification and instruction semantics specification. We solve the former problem by extending the classical Total Store Ordering memory model with caches, translation-lookaside buffers, memory fences, locks, and other features of the x86 processor. For the latter problem we design a new domain-specific language which makes instruction semantics specification readable and compact.

January

Tim SCHWARTZ

The Always Best Positioned Paradigm for Mobile Indoor Applications

(Advisor: Prof. Wolfgang Wahlster)

Tues, 31 Jan 2012, 10:00 h, in the new DFKI building D3 4, VisRoom (-1.63)

In this dissertation, methods for personal positioning in outdoor and indoor environments are investigated. The Always Best Positioned paradigm, which has the goal of providing a preferably consistent self-positioning, will be defined.

Furthermore, the localization toolkit LOCATO will be presented, which allows to easily realize positioning systems that follow the paradigm. New algorithms were

developed, which particularly address the robustness of positioning systems with respect to the Always Best Positioned paradigm. With the help of this toolkit, three example positioning-systems were implemented, each designed for different applications and requirements: a low-cost system, which can be used in conjunction with user-adaptive public displays, a so-called opportunistic system, which enables positioning with room-level accuracy in any building that provides a WiFi infrastructure, and a high-accuracy system for instrumented environments, which works with active RFID tags and infrared beacons. Furthermore, a new and unique evaluation-method for positioning systems is presented, which uses step-accurate natural walking-traces as ground truth. Finally, six location based services will be presented, which were realized either with the tools provided by LOCATO or with one of the example positioning-systems.

Mark KAMINSKI

Incremental Decision Procedures for Modal Logics with Nominals and Eventualities

(Advisor: Prof. Gert Smolka)

Tues, 24 Jan 2012, 16:15 h, in building E1 3, HS 002

This thesis contributes to the study of incremental decision procedures for modal logics with nominals and eventualities. Eventualities are constructs that allow to reason about the reflexive-transitive closure of relations. Eventualities are an essential feature of temporal logics and propositional dynamic logic (PDL). Nominals extend modal logics with the possibility to reason about state equality. Modal logics with nominals are often called hybrid logics. Incremental procedures are procedures that can potentially solve a problem by performing only the reasoning steps needed for the problem in the underlying calculus.

We begin by introducing a class of syntactic models called demos and showing how demos can be used for obtaining nonincremental but worst-case optimal decision procedures for extensions of PDL with nominals, converse and difference modalities. We show that in the absence of nominals, such nonincremental procedures can be refined into incremental demo search procedures, obtaining a worst-case optimal decision procedure for modal logic with eventualities. We then develop the first incremental decision procedure for hybrid logic with eventualities, which we eventually extend to deal with hybrid PDL.

Sathish Chandra PAMMI

Synthesis of Listener Vocalizations - Towards Interactive Speech Synthesis

(Advisor: Prof. Hans Uszkoreit)

Mon, 16 Jan 2012, 16:00 h, building D3 2 (DFKI), Reuse seminar room

Spoken and multi-modal dialogue systems start to use listener vocalizations, such as uh-huh and mm-hm, for natural interaction. Generation of listener vocalizations is one of the major objectives of emotionally colored conversational speech synthesis. Success in this endeavor depends on the answers to three questions: Where to synthesize a listener vocalization? What meaning should be conveyed through the synthesized vocalization? And, how to realize an appropriate listener vocalization with the intended meaning? This thesis addresses the latter question. The investigation starts with proposing a three-stage approach: (i) data collection, (ii) annotation, and (iii) realization. The first stage presents a method to collect natural listener vocalizations from German and British English professional actors in a recording studio. In the second stage, we explore a methodology for annotating listener vocalizations -- meaning and behavior (form) annotation. The third stage proposes a realization strategy that uses unit selection and signal modification techniques to generate appropriate listener vocalizations upon user requests. Finally, we evaluate naturalness and appropriateness of synthesized vocalizations using perception studies. The work is implemented in the open source MARY text-to-speech framework, and it is integrated into the SEMAINE project's Sensitive Artificial Listener (SAL) demonstrator.

Catalin HRITCU

Union, Intersection, and Refinement Types and Reasoning About Type Disjointness for Security Protocol Analysis

(Advisor: Prof. Michael Backes)

Tues, 10 Jan 2012, 14:00h, building E1 1, R 4.07

In this thesis we present two new type systems for verifying the security of cryptographic protocol models expressed in a spi-calculus and, respectively, of protocol implementations expressed in a concurrent lambda calculus. The two type systems combine prior work on refinement types with union and intersection types and with the novel ability to reason statically about the disjointness of types. The increased expressivity enables the analysis of important protocol classes that were previously out of scope for the type-based analyses of cryptographic protocols. In particular, our type systems can statically analyze protocols that are based on zero-knowledge proofs, even in scenarios when certain protocol participants are compromised. The analysis is scalable and provides security proofs for an unbounded number of protocol executions. The two type systems come with mechanized proofs of correctness and efficient implementations.

2012

December

Rawia AWADALLAH

Methods for Constructing an Opinion Network Politically Controversial Topics

(Advisor: Prof. Gerhard Weikum)

Fri, 21 Dec 2012, 12:00h, building E1 4 (MPI-Inf), room 024

The US presidential race, the re-election of President Hugo Chavez, and the economic crisis in Greece and other European countries are some of the controversial topics being played on the news every day. To understand the landscape of opinions on political controversies, it would be helpful to know which politician or other stakeholder takes which position – support or opposition - on specific aspects of these topics. The work described in this thesis aims to automatically derive a map of the opinions-people network from news and other Web documents. The focus is on acquiring opinions held by various stakeholders on politically controversial topics. This opinions-people network serves as a knowledge-base of opinions in the form of <opinion holder><opinion><topic> triples. Our system to build this knowledge-base makes use of on-line news sources in order to extract opinions from text snippets. These sources come with a set of unique challenges. For example, processing text snippets involves not just identifying the topic and the opinion, but also attributing that opinion to a specific opinion holder. This requires making use of deep parsing and analyzing the parse tree. Moreover, in order to ensure uniformity, both the topic as well the opinion holder should be mapped to canonical strings, and the topics should also be organized into a hierarchy. Our system relies on two main components: i) acquiring opinions which uses a combination of techniques to extract opinions from on-line news sources, and ii) organizing topics which crawls and extracts debates from on-line sources, and organizes these debates in a hierarchy of political controversial topics. We present systematic evaluations of the different components of our system, and show their high accuracies. We also present some of the different kinds of applications that require political analysis. We present some application requires political analysis such as identifying flip-floppers, political bias, and dissenters. Such applications can make use of the knowledge-base of opinions.

Ernst Moritz HAHN

Model Checking Stochastic Hybrid Systems

(Advisor: Prof. Holger Hermanns)

Fri, 21 Dec 2012, 9:00h, building E1 7, room 001

The interplay of random phenomena with discrete-continuous dynamics deserves increased attention in many systems of growing importance. Their verification needs to consider both stochastic behaviour and hybrid dynamics. In the verification of classical hybrid systems, one is often interested in deciding whether unsafe system states can be reached. In the stochastic setting, we ask instead whether the probability of reaching particular states is bound by a given threshold. In this thesis, we consider stochastic hybrid systems and develop a general abstraction framework for deciding such problems. This gives rise to the first mechanisable technique that can, in practice, formally verify safety properties of systems which feature all the relevant aspects of nondeterminism, general continuous-time dynamics, and probabilistic behaviour. Being based tools for classical hybrid systems, future improvements in the effectivity of such tools directly carry over to improvements in the effectivity of our technique.

We extend the method in several directions: First, we discuss how we can handle continuous probability distributions. We then consider systems which we are in partial control of. Next, we consider systems in which probabilities are parametric, to analyse entire system families at once. Afterwards, we consider systems equipped with rewards, modelling costs or bonuses. Finally, we consider all orthogonal combinations of the extensions to the core model.

Ndapandula NAKASHOLE

Automatic Extraction of Facts, Relations, and Entities for Web-Scale Knowledge Base Population

(Advisor: Prof. Gerhard Weikum)

Thur, 20 Dec 2012, 15:15h, building E1 4 (MPI-Inf), room 024

Equipping machines with knowledge, through the construction of machine-readable knowledge bases, presents a key asset for semantic search, machine translation, question answering, and other formidable challenges in artificial intelligence. However, human knowledge predominantly resides in books and other natural language text forms. This means that knowledge bases must be extracted and synthesized from natural language text.

When the source of text is the Web, extraction methods must cope with ambiguity, noise, scale, and updates. The goal of this dissertation is to develop knowledge base population methods that address the afore mentioned characteristics of Web text. The dissertation makes three contributions. The first contribution is a method for mining high-quality facts at scale, through distributed constraint reasoning and a pattern representation model that is robust against noisy patterns. The second contribution is a method for mining a large comprehensive collection of relation types beyond those commonly found in existing knowledge bases. The third contribution is a method for extracting facts from dynamic Web sources such as news articles and social media where one of the key challenges is the constant emergence of new entities. All methods have been evaluated through experiments involving Web-scale text collections.

Christian HAHN

A Platform-Independent Domain –Specific Modeling Language for Multiagent Systems

(Advisor: Prof. Jörg Siekmann)

Thur, 20 Dec 2012, 14:00h, building D3 4 (DFKI, VisCenter), room -1.63 NB

Associated with the increasing acceptance of agent-based computing as a novel software engineering paradigm, recently a lot of research addresses the development of suitable techniques to support the agent-oriented software development. The state-of-the-art in agent-based software development is to (i) design the agent systems basing on an agent-based methodology and (ii) take the resulting design artifact as a base to manually implement the agent system using existing agent-oriented programming languages or general purpose languages like Java. Apart from failures made when manually transform an abstract specification into a concrete implementation, the gap between design and implementation may also result in the divergence of design and implementation. The framework discussed in this dissertation presents a platform-independent domain-specific modeling language for MASs called DSML4MAS that allows modeling agent systems in a platform-independent and graphical manner. Apart from the abstract design,

DSML4MAS also allows to automatically (i) check the generated design artifacts against a formal semantic specification to guarantee the well-formedness of the design and (ii) translate the abstract specification into a concrete implementation. Taking both together, DSML4MAS ensures that for any well-formed design, an associated implementation will be generated closing the gap between design and code.

Victor ALVAREZ

Selected Topics in Algorithmic Geometry

(Advisor: Prof. Raimund Seidel)

Thur, 20 Dec 2012, 11:00h, building E1 7, room 0.01

Let P be a set of n points on the plane with no three points on a line. A crossing-free structure on P is a straight-edge plane graph whose vertex set is P . In this thesis we consider problems of two different topics in the area of algorithmic geometry: Geometry using Steiner points, and counting algorithms. These topics have certain crossing-free structures on P as our primary objects of study. Our results can roughly be described as follows:

i) Given a k -coloring of P , with $k \geq 3$ colors, we will show how to construct a set of Steiner points $S = S(P)$ such that a k -colored quadrangulation can always be constructed on $(P \cup S)$. The bound we show of $|S|$ significantly improves on previously known results.

ii) We also show how to construct a set $S = S(P)$ of Steiner points such that a triangulation of $(P \cup S)$ having all its vertices of even (odd) degree can always be constructed. We show that $|S| \leq n/3 + c$, where c is a constant. We also look at other variants of this problem.

iii) With respect to counting algorithms, we show new algorithms for counting triangulations, pseudo-triangulations, crossing-free matchings and crossing-free spanning cycles on P . Our algorithms are simple and allow good analysis of their running times. These algorithms significantly improve over previously known results. We also show an algorithm that counts triangulations approximately, and a hardness result of a particular instance of the problem of counting triangulations exactly.

iv) We show experiments comparing our algorithms for counting triangulations with another well-known algorithm that is supposed to be very fast in practice.

Christoph ENDRES

PRESTK: Situation –Aware Presentation of Messages and Infotainment Content for Drivers

(Advisor: Prof. Wolfgang Wahlster)

Wed, 19 Dec 2012, 14:30h, building E1 1, room 407

The amount of in-car information systems has dramatically increased over the last few years. These potentially mutually independent information systems presenting information to the driver increase the risk of driver distraction.

In a first step, orchestrating these information systems using techniques from scheduling and presentation planning avoid conflicts when competing for scarce resources such as screen space. In a second step, the cognitive capacity of the driver as another scarce resource has to be considered. For the first step, an algorithm fulfilling the requirements of this situation is presented and evaluated.

For the second step, I define the concept of System Situation Awareness (SSA) as an extension of Endsley's Situation Awareness (SA) model. I claim that not only the driver needs to know what is happening in his environment, but also the system, e.g., the car. In order to achieve SSA, two paths of research have to be followed:

(1) Assessment of cognitive load of the driver in an unobtrusive way. I propose to estimate this value using a model based on environmental data.

(2) Developing model of cognitive complexity induced by messages presented by the system. Three experiments support the claims I make in my conceptual contribution to this field.

A prototypical implementation of the situation-aware presentation management toolkit PRESTK is presented and shown in two demonstrators.

Manuel GORIUS

Adaptive Delay-constrained Internet Media Transport

(Advisor: Prof. Thorsten Herfet)

Wed, 19 Dec 2012, 13:30h, building C6 3, room 9.05

Reliable transport layer Internet protocols do not satisfy the requirements of packetized, real-time multimedia streams. The available thesis motivates and defines predictable reliability as a novel, capacity-approaching transport paradigm, supporting an application-specific level of reliability under a strict delay constraint. This paradigm is being implemented into a new protocol design -- the Predictably Reliable Real-time Transport protocol (PRRT).

In order to predictably achieve the desired level of reliability, proactive and reactive error control must be optimized under the application's delay constraint. Hence, predictably reliable error control relies on stochastic modeling of the protocol response to the modeled packet loss behavior of the network path. The result of the joined modeling is periodically evaluated by a reliability control policy that validates the protocol configuration under the application constraints and under consideration of the available network bandwidth. The adaptation of the protocol parameters is formulated into a combinatorial optimization problem that is solved by a fast search algorithm incorporating explicit knowledge about the search space. Experimental evaluation of PRRT in real Internet scenarios demonstrates that predictably reliable transport meets the strict QoS constraints of high-quality, audio-visual streaming applications.

David GÜNTHER

Topological Analysis of Discrete Scalar Data

(Advisor: Dr. Tino Weinkauff)

Tues, 18 Dec 2012, 10:15h, building E1 4 (MPI-Inf), room 0.19

This thesis presents a novel computational framework that allows for a robust extraction and quantification of the Morse-Smale complex of a scalar field given on a 2- or 3-dimensional manifold. The proposed framework is based on Forman's discrete Morse theory, which guarantees the topological consistency of the computed complex. Using a graph theoretical formulation of this theory, we present an algorithmic library that computes the Morse-Smale complex combinatorially with

an optimal complexity of $O(n^2)$ and efficiently creates a multi-level representation of it. We explore the discrete nature of this complex, and relate it to the smooth counterpart. It is often necessary to estimate the feature strength of the individual components of the Morse-Smale complex – the critical points and separatrices. To do so, we propose a novel output-sensitive strategy to compute the persistence of the critical points. We also extend this well-founded concept to separatrices by introducing a novel measure of feature strength called separatrix persistence. We evaluate the applicability of our methods in a wide variety of application areas ranging from computer graphics to planetary science to computer and electron tomography.

Marc SCHLICKLING

Timing Model Derivation – Static Analysis of Hardware Description Languages

(Advisor: Prof. Reinhard Wilhelm)

Mon, 17 Dec 2012, 8:30h, building E1 7, room 0.01

Safety-critical hard real-time systems are subject to strict timing constraints. In order to derive guarantees on the timing behavior, the worst-case execution time (WCET) of each task comprising the system has to be known. The aiT tool has been developed for computing safe upper bounds on the WCET of a task. Its computation is mainly based on abstract interpretation of timing models of the processor and its periphery. These models are currently hand-crafted by human experts, which is a time-consuming and error-prone process. Modern processors are automatically synthesized from formal hardware specifications. Besides the processor's functional behavior, also timing aspects are included in these descriptions. A methodology to derive sound timing models using hardware specifications is described within this thesis. To ease the process of timing model derivation, the methodology is embedded into a sound framework.

A key part of this framework are static analyses on hardware specifications. This thesis presents an analysis framework that is build on the theory of abstract interpretation allowing use of classical program analyses on hardware description languages. Its suitability to automate parts of the derivation methodology is shown by different analyses. Practical experiments demonstrate the applicability of the approach to derive timing models. Also the soundness of the analyses and the analyses' results is proven.

Sandra EBERT

Semi-Supervised Learning for Image Classification

(Advisor: Prof. Bernt Schiele)

Fri, 14 Dec 2012, 11:00h, building E1 4 (MPI-Inf), room 0.24

Object class recognition is an active topic in computer vision still presenting many challenges. In most approaches, this task is addressed by supervised learning algorithms that need a large quantity of labels to perform well. This leads either to small datasets ($< 10,000$ images) that capture only a subset of the real-world class distribution (but with a controlled and verified labeling procedure), or to large datasets that are more representative but also add more label noise. Therefore, semi-supervised learning is a promising direction. It requires only few labels while simultaneously making use of the vast amount of images available today. In this thesis, we address object class recognition with semi-supervised learning. These algorithms depend on the underlying structure given by the data, the image description, and the similarity measure, and the quality of the labels. In the talk, I will illustrate this strong relationship between classification performance and the quality of the labels as well as the structure. Furthermore, I will show how we tackle these problems by active learning to get more representative labels and by several graph improvements.

November

Benny KNEISSL

Automated in silico Protein Modeling Strategies – Applications and Limits in G-Protein Coupled Receptor Modeling

(Advisor: Prof. Andreas Hildebrandt)

Fri, 30 Nov 2012, 11:00h, building E2 1, room 0.07

In the 21st century, computer-aided methods became a well-established tool in the pharmaceutical industry. Because many protein structures are still very difficult to determine experimentally, various in silico methods have been developed to predict their three-dimensional structure from the corresponding sequence. Unfortunately, the modeling process has often to be adapted to the target protein or the automatically generated models have to be manually refined afterwards by the users based on their expert knowledge to finally achieve reasonable results. The aim of this thesis is to explore the applicability of automated in silico modeling strategies by means of G-protein coupled receptors (GPCRs). First, we analyze to which extent available protein structure prediction methods can be particularly tailored to the automated GPCR modelling case. Second, we develop our own approach and demonstrate that we obtain improved models compared to other state-of-the-art modeling tools. More important, our method does not rely on manual interactions by the user during the modeling process and is thus generally applicable for all GPCRs. Furthermore, we present a new sequence based method to predict structural distortions from ideal α -helical geometry. This method also exceeds the prediction accuracy of comparable approaches.

Pavel EMELIANENKO

Harnessing the Power of GPUs for Problem in real Algebraic Geometry

(Advisor: Prof. Kurt Mehlhorn)

Fri, 9 Nov 2012, 17:00h, building E1 4 (MPI-INF), room 0.24

This thesis presents novel parallel algorithms to leverage the power of GPUs (Graphics Processing Units) for exact computations with polynomials having large integer coefficients. The significance of such computations, especially in real algebraic geometry, is hard to undermine. On massively-parallel architectures such as GPU, the degree of data-level parallelism exposed by an algorithm is the main performance factor. We attain high efficiency through the use of structured matrix theory to assist the realization of relevant operations on polynomials on the graphics

hardware. A detailed complexity analysis, assuming the PRAM model, also confirms that our approach achieves a substantially better parallel complexity in comparison to classical algorithms used for symbolic computations. Aside from the theoretical considerations, a large portion of this work is dedicated to the actual algorithm development and optimization techniques where we pay close attention to the specifics of the graphics hardware. As a byproduct of this work, we have developed high-throughput modular arithmetic which we expect to be useful for other GPU applications, in particular, open-key cryptography. We further discuss the algorithms for the solution of a system of polynomial equations, topology computation of algebraic curves and curve visualization which can profit to the full extent from the GPU acceleration. Extensive benchmarking on a real data demonstrates the superiority of our algorithms over several state-of-the-art approaches available to date. This thesis is written in English.

Andreas BAAK

Retrieval-based Approaches for Tracking and Reconstruction Human Motions

(Advisor: Prof. Meinard Müller)

Fri, 9 Nov 2012, 14:00h, building E1 4 (MPI-INF), room 0.19

In this thesis, we introduce novel approaches for increasing the stability, accuracy, and efficiency of marker-less human motion tracking and 3D human pose reconstruction. As one common underlying concept, the presented approaches contain a retrieval component making use of previously recorded marker-based motion capture (mocap) data. In particular, we contribute to three different areas dealing with various types of sensors including video cameras, optical mocap systems, inertial sensors, and depth cameras. Firstly, we introduce content-based retrieval techniques for automatically segmenting and annotating mocap data that is originally provided in form of unstructured data collections. Secondly, we show how such robust annotation procedures can be used to support and stabilize video-based marker-less motion tracking. Thirdly, we develop algorithms for reconstructing human motions from noisy depth sensor data in real-time. In all these contributions, a particular focus is put on efficiency issues in order to keep down the run time.

Peter Matthias GROSCHE

Signal Processing Methods for Beat Tracking, Music Segmentation and Audio Retrieval

(Advisor: Prof. Meinard Müller)

Fri, 9 Nov 2012, 12:00h, building E1 4 (MPI-INF), room 0.19

The goal of music information retrieval (MIR) is to develop novel strategies and techniques for organizing, exploring, accessing, and understanding music data in an efficient manner. The conversion of waveform-based audio data into semantically meaningful feature representations by the use of digital signal processing techniques is at the center of MIR and constitutes a difficult field of research because of the complexity and diversity of music signals. In this thesis, we introduce novel signal processing methods that allow for extracting musically meaningful information from audio signals. As main strategy, we exploit musical knowledge about the signals' properties to derive feature representations that show a significant degree of robustness against musical variations but still exhibit a high musical expressiveness. We apply this general strategy to three different areas of MIR: Firstly, we introduce novel techniques for extracting tempo and beat information, where we particularly consider challenging music with changing tempo and soft note onsets. Secondly, we present novel algorithms for the automated segmentation and analysis of folk song field recordings, where one has to cope with significant fluctuations in intonation and tempo as well as recording artifacts. Thirdly, we explore a cross-version approach to content-based music retrieval based on the query-by-example paradigm. In all three areas, we focus on application scenarios where strong musical variations make the extraction of musically meaningful information a challenging task.

Verena KONZ

Automated Methods for Audio – Based Music Analysis with Applications to Musicology

(Advisor: Prof. Meinard Müller)

Fri, 9 Nov 2012, 10:00h, building E1 4 (MPI-INF), room 0.19

This thesis contributes to bridging the gap between music information retrieval (MIR) and musicology. We present several automated methods for music analysis, which are motivated by concrete application scenarios being of central importance in musicology. In this context, the automated music analysis is performed on the basis of audio material. Here, one reason is that for a given piece of music usually many different recorded performances exist. The availability of multiple versions of a piece of music is exploited in this thesis to stabilize analysis results. We show how the presented automated methods open up new possibilities for supporting musicologists in their work. Furthermore, we introduce novel interdisciplinary concepts which facilitate the collaboration between computer scientists and musicologists. Based on these concepts, we demonstrate how MIR researchers and musicologists may greatly benefit from each other in an interdisciplinary collaboration. Firstly, we present a fully automatic approach for the extraction of tempo parameters from audio recordings and show to which extent this approach may support musicologists in analyzing recorded performances. Secondly, we introduce novel user interfaces which are aimed at encouraging the exchange between computer science and musicology. In this context, we indicate the potential of computer-based methods in music education by testing and evaluating a novel MIR user interface at the University of Music Saarbrücken. Furthermore, we show how a novel multi-perspective user interface allows for interactively viewing and evaluating version-dependent analysis results and opens up new possibilities for interdisciplinary collaborations. Thirdly, we present a cross-version approach for harmonic analysis of audio recordings and demonstrate how this approach enables musicologists to explore harmonic structures even across large music corpora. Here, one simple yet important conceptual contribution is to convert the physical time axis of an audio recording into a performance-independent musical time axis given in bars.

Patrick WISCHNEWSKI

Efficient Reasoning Procedures for Complex First-Order Theories

(Advisor: Prof. Christoph Weidenbach)

Tues, 6 Nov 2012, 10:00h, building E1 5 (MPI-SWS), room HS 002

The complexity of a set of first-order formulas results from the size of the set and the complexity of the problem described by its formulas.

Decision Procedures for Ontologies

This thesis presents new superposition based decision procedures for large sets of formulas. The sets of formulas may contain expressive constructs like transitivity and equality. The procedures decide the consistency of knowledge bases, called ontologies, that consist of several million formulas and answer complex queries with respect to these ontologies. They are the first superposition based reasoning procedures for ontologies that are at the same time efficient, sound, and complete. The procedures are evaluated using the well-known ontologies YAGO, SUMO and CYC. The results of the experiments, which are presented in this thesis, show that these procedures decide the consistency of all three above-mentioned ontologies and usually answer queries within a few seconds.

Reductions for General Automated Theorem Proving

Sophisticated reductions are important in order to obtain efficient reasoning procedures for complex, particularly undecidable problems because they restrict the search space of theorem proving procedures. In this thesis, I have developed a new powerful reduction rule. This rule enables superposition based reasoning procedures to find proofs in sets of complex formulas. In addition, it increases the number of problems for which superposition is a decision procedure.

Hans-Jörg PETER

A Uniform Approach to the Complexity and Analysis of Succinct Systems

(Advisor: Prof. Bernd Finkbeiner)

Fri, 2 Nov 2012, 16:00h, building E1 7, room 0.01

This thesis provides a unifying view on the succinctness of systems:

the capability of a modeling formalism to describe the behavior of a system of exponential size using a polynomial syntax.

The key theoretical contribution is the introduction of sequential circuit machines as a new universal computation model that focuses on succinctness as the central aspect.

The thesis demonstrates that many well-known modeling formalisms such as communicating state machines, linear-time temporal logic, or timed automata exhibit an immediate connection to this machine model.

Once a (syntactic) connection is established, many complexity bounds for resource-restricted sequential circuit machines can be transferred to a certain formalism in a uniform manner.

As a consequence, besides a far-reaching unification of independent lines of research, we are also able to provide matching complexity bounds for various analysis problems, whose complexities were not known so far.

For example, we establish matching lower and upper bounds of the small witness problem and several variants of the bounded synthesis problem for timed automata, a particularly important succinct modeling formalism.

Also for timed automata, our complexity-theoretic analysis leads to the identification of tractable fragments of the timed synthesis problem under partial observability. Specifically, we identify timed controller synthesis based on discrete or template-based controllers to be equivalent to model checking.

Based on this discovery, we develop a new model checking-based algorithm to efficiently find feasible template instantiations.

From a more practical perspective, this thesis also studies the preservation of succinctness in analysis algorithms using symbolic data structures.

While efficient techniques exist for specific forms of succinctness considered in isolation, we present a general approach based on abstraction refinement to combine off-the-shelf symbolic data structures.

In particular, for handling the combination of concurrency and quantitative timing behavior, we report on the tool Synthia which combines binary decision diagrams with difference bound matrices.

October

Sebastian ALTMEYER

Analysis of Preemptively Scheduled Hard Real-time Systems

(Advisor: Prof. Reinhard Wilhelm)

Thurs, 25 Oct 2012, 14:15h, building E1 1, room 407

As timing is a major property of hard real-time, proving timing correctness is of utter importance. A static timing analysis derives upper bounds on the execution time of tasks, a scheduling analysis uses these bounds and checks if each task meets its timing constraints. In preemptively scheduled systems with caches, this interface between timing analysis and scheduling analysis must be considered outdated. On a context switch, a preempting task may evict cached data of a preempted task that need to be reloaded again after preemption. The additional execution time due to these reloads, called cache-related preemption delay (CRPD), may substantially prolong a task's execution time and strongly influence the system's performance. In this thesis, we present a formal definition of the cache-related preemption delay and determine the applicability and the limitations of a separate CRPD computation. To bound the CRPD based on the analysis of the preempted task, we introduce the concept of definitely cached useful cache blocks. This new concept eliminates substantial pessimism with respect to former analyses by considering the over-approximation of a preceding timing analysis. We consider the impact of the preempting task to further refine the CRPD bounds. To this end, we present the notion of resilience. The resilience of a cache block is a measure for the amount of disturbance of a preempting task a cache block of the preempted task may survive. Based on these CRPD bounds, we show how to correctly account for the CRPD in the schedulability analysis for fixed-priority preemptive systems and present new CRPD-aware response time analyses: ECB-Union and Multiset approaches.

Linlin LI

Computational Modeling of Lexical Ambiguity

(Advisor: Dr. Caroline Sporleder)

Wed, 10 Oct 2012, 15:00h, C7 4, conference room

Lexical ambiguity is a frequent phenomenon that can occur not only for words but

also on the phrase level. Natural language processing systems need to efficiently deal with these ambiguities in various tasks, however, we often encounter such system failures in real applications. This thesis aims to study various lexical ambiguity phenomena from a computational modeling perspective and advance the state-of-the-art research.

Andreas BROSCART

Efficient Query Processing and Index Tuning using Proximity Scores

(Advisor: Prof. Gerhard Weikum)

Tues, 9 Oct 2012, 11:00h, E1 4 (MPI-Inf), room 0.24

In the presence of growing data, the need for efficient query processing under result quality and index size control becomes more and more a challenge to search engines. We show how to use proximity scores to make query processing effective and efficient with focus on either of the optimization goals.

More precisely, we make the following contributions:

* We present a comprehensive comparative analysis of proximity score models and a rigorous analysis of the potential of phrases and adapt a leading proximity score model for XML data

* We discuss the feasibility of all presented proximity score models for top-k query processing and present a novel index combining a content and proximity score that helps to accelerate top-k query processing and improves result quality.

* We present a novel, distributed index tuning framework for term and term pair index lists that optimizes pruning parameters by means of well-defined optimization criteria under disk space constraints. Indexes can be tuned with emphasis on efficiency or effectiveness: the resulting indexes yield fast processing at high result quality.

* We show that pruned index lists processed with a merge join outperform top-k query processing with unpruned lists at a high result quality.

* Moreover, we present a hybrid index structure for improved cold cache run times.

September

Laura TOLOSI

Finding Regions of Abberant DNA Copy Number Associated with Tumor Phenotype

(Advisor: Prof. Thomas Lengauer)

Thurs, 27 Sept 2012, 10:00h, E1 4 (MPI-Inf), room 0.24

DNA copy number alterations are a hallmark of cancer. Understanding their role in tumor progression can help improve diagnosis, prognosis and therapy selection for cancer patients. High-resolution, genome-wide measurements of DNA copy number changes for large cohorts of tumors are currently available, owing to technologies like microarray-based array comparative hybridization (arrayCGH). In this thesis, we present a computational pipeline for statistical analysis of tumor cohorts, which can help extract relevant patterns of copy number aberrations and infer their association with various phenotypical indicators.

The main challenges are the instability of classification models due to the high dimensionality of the arrays compared to the small number of tumor samples, as well as the large correlations between copy number estimates measured at neighboring loci. We show that the feature ranking given by several widely-used methods for feature selection is biased due to the large correlations between features. In order to correct for the bias and instability of the feature ranking, we introduce methods for consensus segmentation of the set of arrays. We present three algorithms for consensus segmentation, which are based on identifying recurrent DNA breakpoints or DNA regions of constant copy number profile. The segmentation constitutes the basis for computing a set of super-features, corresponding to the regions. We use the super-features for supervised classification and we compare the models to baseline models trained on probe data. We validated the methods by training models for prediction of the phenotype of breast cancers and neuroblastoma tumors. We show that the multivariate segmentation affords higher model stability, in general improves prediction accuracy and facilitates model interpretation. One of our most important biological results refers to the classification of neuroblastoma tumors. We show that patients belonging to different age subgroups are characterized by distinct copy number patterns, with largest discrepancy when the subgroups are defined as older or younger than 16-18 months. We thereby confirm the recommendation for a higher age cutoff than 12 months (current clinical practice) for differential diagnosis of neuroblastoma.

Markus PISTER

Timing Model Derivation – Pipeline Analyzer Generation from Hardware Description Languages

(Advisor: Prof. Reinhard Wilhelm)

Wed, 19 Sept 2012, 14:30h, E1 7, room 0.01 (ground floor, Cluster building)

Safety-critical systems are forced to finish their execution within strict deadlines so that worst-case execution time (WCET) guarantees are a crucial part of their verification. Timing models of the analyzed hardware form the basis for static analysis-based approaches like the aiT WCET analyzer. Currently, timing models are hand-crafted based on frequently incorrect documentation causing the process to be error-prone and time-consuming. This thesis bridges the gap between automatic hardware synthesis and WCET analysis development by introducing a process for the derivation of timing models from VHDL specifications. We propose a set of transformations and abstractions to reduce the hardware design's complexity enabling the generation of efficient and provably correct WCET analyzers. They employ an abstract interpretation-based simulation of program executions based on a defined abstract simulation semantics. We have defined workflow patterns showing how to gradually apply the derivation process to VHDL models, thereby removing timing-irrelevant constructs. Interval property checking is used to validate the transformations. A further contribution of this thesis is the implementation of a tool set that realizes the introduced derivation process and shows its applicability to non-trivial industrial designs in experimental evaluations.

Influences on design choices to the quality of the derived timing model are

presented building an informal predictability notion for VHDL.

Stefan POPOV

Algorithms and Data Structures for Interactive Ray Tracing on Commodity Hardware

(Advisor: Prof. Philipp Slusallek)

Tues, 18 Sept 2012, 11:00h, in The Visualization Center, DFKI (ground floor, new building)

Rendering methods based on ray tracing provide high image realism, but have been historically regarded as offline only. This has changed in the past decade, due to significant advances in the construction and traversal performance of acceleration structures and the efficient use of data-parallel processing. Today, all major graphics companies offer real-time ray tracing solutions. Our work, which we present in this talk, has contributed to this development with some key insights.

In the first part of the talk we focus on the GPU as a means to bring real-time ray tracing to desktop computers. There, we will briefly discuss our GPU friendly KD-tree and BVH traversal algorithms.

We will then address the limited support of dynamic scenes in previous work, by presenting two novel parallel-friendly construction algorithms for KD-trees and BVHs. By approximating the cost function, we accelerate construction by up to an order of magnitude (especially for BVHs), at the expense of only tiny degradation to traversal performance.

Finally, we address the topic of creating the "perfect" acceleration structure. We show how to develop a polynomial time non-greedy BVH construction algorithm. We then modify it to produce a new type of acceleration structure that inherits both the high performance of KD-trees and the small size of BVHs.

August

Andrey SHADRIN

Mixed Low- and High Level Programming Language Semantics and Automated Verification of a Small Hypervisor

(Advisor: Prof. Wolfgang Paul)

Mon, 27 August 2012, 15:00h, in building E1 7, room 001

Hypervisors are system software programs that virtualize the architecture they run on and are usually implemented in a mix of (macro) assembly and a high-level language like C. To verify such a software, assembly parts as well as C parts should be verified, where reasoning about those parts is done in different semantics. At the end, both semantics should be brought together in an overall correctness theorem of such a software program. The formal integration of correctness results accomplished in distinct semantics is challenging but inevitable for systems verification. This thesis is split into two parts. In the first one, we will present the mixed semantics of C and macro assembly. This semantics can handle mixed-language implementations where the execution context is changed by an external function call from assembly to C and vice versa. Also, we state a step-by-step simulation theorem between mixed programs and the compiled and assembled code. In the second part, we present the correctness of a small hypervisor, called Baby Hypervisor (BHV), described by the mixed semantics. BHV virtualizes a 32-bit RISC architecture. The BHV functional verification was shown using Microsoft's VCC, an automatic verifier for C with contracts. BHV C portions were verified with VCC. For making macro assembly feasible with VCC the original macro assembly is translated to C code simulating processor (simulation approach).

Alekh JINDAL

OctopusDB: Flexible and Scalable Storage Management for Arbitrary Database Engines

(Advisor: Prof. Jens Dittrich)

Fri, 24 August 2012, 14:30, in building E1 1, room 3.06

We live in a dynamic age with the economy, the technology, and the people around us changing faster than ever before. Consequently, the data management needs in our modern world are much different than those envisioned by the early database inventors in the 70s. Today, enterprises face the challenge of managing ever-growing dataset sizes with dynamically changing query workloads. The current practice to deal with changing query workloads is to have a different specialized product for each workload type, e.g. row stores for OLTP workload, column stores for OLAP workload, streaming systems for streaming workload, and scan-oriented systems for shared query processing. However, this has the additional penalty of managing a zoo of data managing systems in the first place, which is tedious, expensive, as well as counter-productive for modern enterprises. This thesis presents an alternative approach to supporting several query workloads in a data managing system. We observe that each specialized database product has a different data store, indicating that different query workloads work well with different data layouts. Thus, a key requirement for supporting several query workloads is to support several data layouts. Therefore, in this thesis, we study ways to inject different data layouts into existing (and familiar) data managing systems. The goal is to develop a flexible storage layer which can support several query workloads in a single data managing system. We present a set of non-invasive techniques, coined Trojan Techniques, to inject different data layouts into a data managing system. Trojan Techniques are non-invasive in the sense that they do not make heavy untenable changes to the system and yet they bring significant improvements in query performance. Our experimental results show that Trojan Techniques can improve the performance of Hadoop MapReduce by a factor of up to 18, and that of a top-notch commercial database product by a factor of up to 17.

Dimitar DENEV

Models and Methods for Web Archive Crawling

(Advisor: Prof. Gerhard Weikum)

Mon, 20 August 2012, 15:00h, in building E1 4, room 024

Web archives offer a rich and plentiful source of information to researchers, analysts, and legal experts. For this purpose, they gather Web sites as the sites change over time. In order to keep up to high standards of data quality, Web archives have to collect all versions of the Web sites. Due to limited resources and technical

constraints this is not possible. Therefore, Web archives consist of versions archived at various time points without guarantee for mutual consistency. This thesis presents a model for assessing the data quality in Web archives as well as a family of crawling strategies yielding high-quality captures. We distinguish between single-visit crawling strategies for exploratory and visit-revisit crawling strategies for evidentiary purposes. Single-visit strategies download every page exactly once aiming for an “undistorted” capture of the ever-changing Web. We express the quality of such the resulting capture with the “blur” quality measure. In contrast, visit-revisit strategies download every page twice. The initial downloads of all pages form the visit phase of the crawling strategy. The second downloads are grouped together in the revisit phase. These two phases enable us to check which pages changed during the crawling process. Thus, we can identify the pages that are consistent with each other. The quality of the visit-revisit captures is expressed by the “coherence” measure. Quality-conscious strategies are based on predictions of the change behaviour of individual pages. We model the Web site dynamics by Poisson processes with page specific change rates. Furthermore, we show that these rates can be statistically predicted. Finally, we propose visualization techniques for exploring the quality of the resulting Web archives. A fully functional prototype demonstrates the practical viability of our approach.

Piotr DIDYK

Perceptual Display: Exceeding Display Limitations by Exploiting the Human Visual System

(Advisor: Prof. Hans-Peter Seidel)

Mon, 20 August 2012, 16:30h, in building E1 4 (MPI-Inf), room 019

Existing displays have a number of limitations, which make it difficult to realistically reproduce real-world appearance; discrete pixels are used to represent images, which are refreshed only a limited number of times per second, the output luminance range is much smaller than in the real world, and only two dimensions are available to reproduce a three-dimensional scene. On the other hand, the human visual system has its own limitations, as those imposed by the density of photoreceptors, imperfections in the eye optics, or the limited ability to discern high-frequency information. In this talk, I will show that taking these limitations into account and using perceptual effects, which very often are not measurable physically, allow us to design methods which can overcome the physical limitations of display devices in order to enhance apparent image qualities. More precisely, I will discuss how high quality frames can be interleaved with low quality frames improving the sharpness of rendered sequence. Further, I will present an optimization technique which produces frames that shown in a rapid succession lead to an apparent increase in spatial resolution, and, finally, I will talk about the role of perception in context of stereovision.

July

Paul LIBBRECHT

Authoring of Semantic Mathematical Content for Learning on the Web

(Advisor: Prof. Jörg Siekmann)

Mon, 30 July 2012, 15:15h, in building E1 3, HS 001

This thesis explores how a mathematician can create content which lets students learn via ActiveMath. This software runs on the web, and employs semantics mathematical technologies and artificial intelligence to support the student. The thesis describes not only a tool to edit content but also studies the workflows of authors when creating content.

An important authoring paradigm introduced here is called WYCIWYG – What You Check is What You Get. This places the target learning environment as the central concern. Indeed, the learning environment, which is the assembly of content and software, represents the final product of the authoring work. This thesis exploits and develops classical paradigms: copy-and-paste, search, and collaborative authoring. To support the authoring workflows, several software pieces have been implemented and tested: within an editing tool and the ActiveMath software. The following work describes their usages, their technical facets, and case studies of how current authors have used them in their day-to-day activities.

Mahmoud FOUZ

Randomized Rumor Spreading in Social Networks & Complete Graphs

(Advisor: Prof. Markus Bläser)

Mon, 16 July 2012, 9:00h, in building E1 4 (MPI-Inf), room 024

This thesis deals with two rumor spreading problems. In the first part, we study the rumor spreading problem on social networks modeled by preferential attachment graphs. We consider the push-pull strategy of Karp et al., where in each round, each vertex chooses a random neighbor and exchanges information with it. We prove the following. The push-pull strategy delivers a message to all nodes within $\Theta(\log n)$ rounds with high probability, where n is the number of nodes in the graph.

The best known bound so far was $O(\log^2 n)$ by Chierichetti, Lattanzi, and Panconesi [ICALP 2009]. If we slightly modify the protocol so that contacts are chosen uniformly from all neighbors but the one contacted in the previous round, then this time reduces to $\Theta(\log n / \log \log n)$. This is asymptotically optimal since it matches the diameter of the graph. In an asynchronous version of the protocol, the running time is shown to be even $O(\log n)$.

In the second part, we consider the rumor spreading problem on the complete graph. We propose a new push protocol that achieves an asymptotically optimal time of $(1 + o(1)) \log n$. It needs only $O(n f(n))$ calls, where $f(n) = \omega(1)$ can be arbitrary. The protocol is also robust against random node failures. We further extend it to deal with adversarial node failures efficiently

Yana MILEVA

Mining the Evolution of Software Component Usage

(Advisor: Prof. Andreas Zeller)

Mon, 9 July 2012, 16:00h, in building E1 1, room 407

The main focus of this thesis is the analysis of the evolution of software components. In order to track the evolution of software components, one needs to collect the evolution information of each component. This information is being stored in the version control system (VCS) of the project—the repository of the history of events happening throughout the project's lifetime. By using software archive mining techniques one can extract and leverage this information. The main contribution of this thesis is the introduction of evolution usage trends and evolution change patterns. The raw information about the occurrences of each component is being stored in the VCS of the project and by organizing it in evolution trends and patterns, we are able to draw conclusions and issue recommendations concerning each individual component and the project as a whole.

Evolution Trends An evolution trend is a way to track the evolution of a software component throughout the span of the project. The trend shows the increases and decreases in the usage of a specific component, which can be indicative of the quality of this component. AKTARI is a tool, presented in this thesis, that is based on such evolution trends and can be used by the software developers to observe and draw conclusions about the behavior of their project.

Evolution Patterns We call an evolution pattern, the pattern of a frequently occurring code change throughout the span of the project. Those frequently occurring changes are project-specific and are explanatory of the way the project evolves. Each such change evolution pattern contains in itself the specific way “things are done” in the project and as such can serve for defect detection and defect prevention. This technique is being implemented as a basis for the LAMARCK tool, presented in this thesis.

Pascal GWOSDEK

Hardware Accelerated Algorithms in Visual Computing

(Advisor: Prof. Joachim Weickert)

Wed, 11 July 2012, 14:15h, in building E1 1, room 407

This talk presents parallel algorithms for visual computing on graphics cards, and evaluates them with respect to their speed, scalability, and quality of the results. This includes homogeneous and anisotropic diffusion processes, diffusion image inpainting, optic flow, and halftoning.

We compare different solvers for homogeneous diffusion and present a novel 'extended' box filter. Moreover, we suggest to use the fast explicit diffusion scheme (FED) as an efficient and flexible solver for nonlinear and in particular for anisotropic parabolic diffusion problems on graphics hardware. For elliptic diffusion-like processes, we recommend to use cascadic FED or fast Jacobi schemes. Our presented optic flow algorithm represents the fastest method among the 30 most accurate techniques. A novel halftoning scheme produces state-of-the-art techniques, and has many applications in image processing and computer graphics.

June

Stephan WILHELM

Symbolic Representations in WCET Analysis

(Advisor: Prof. Reinhard Wilhelm)

Mon, 4 June 2012, 16:00h, in building E1 1, room 407

Reliable task-level execution time information is indispensable for validating the correct operation of safety-critical embedded real-time systems. Static worst-case execution time (WCET) analysis is a method that computes safe upper bounds of the execution time of single uninterrupted tasks. The method is based on abstract interpretation and involves abstract hardware models that capture the timing behavior of the processor on which the tasks run. For complex processors, task-level execution time bounds are obtained by a state space exploration which involves the abstract model and the program. Partial state space exploration is not sound. A full exploration can become too expensive. Symbolic state space exploration methods using binary decision diagrams (BDDs) are known to provide efficient means for covering large state spaces.

This work presents a symbolic method for the efficient state space exploration of abstract pipeline models in the context of static WCET analysis. The method has been implemented for the Infineon TriCore 1 which is a real-life processor of medium complexity. Experimental results on a set of benchmarks and an automotive industry application demonstrate that the approach improves the scalability of static WCET analysis while maintaining soundness.

April

Tom CRECELIUS

Socially Enhanced Search and Explorations in Social Tagging Networks

(Advisor: Prof. Gerhard Weikum)

Mon, 23 April 2012, 16:30h, in building E1 4 (MPI-Inf), room 024

Social tagging networks have become highly popular for publishing and searching contents. Users in such networks can review, rate and comment on contents, or annotate them with keywords (social tags) to give short but exact text representations of even non-textual contents. Moreover, there is an inherent support for interactions and relationships among users and, thus, naturally form groups of friends or of common interests.

This dissertation addresses in this area three research topics, namely:

- (i) A comprehensive framework for search in social tagging networks and efficient top-k result retrieval.
- (ii) A new approach of solving the APSD Problem for dynamically update friendship relations in large user networks.
- (iii) Countering cheating in P2P authority computations over social networks.

Weiwei SUN

Learning Chinese Language Structures with Multiple Views

(Advisor: Prof. Hans Uszkoreit)

Tues, 17 April 2012, 16:00h, in building B3 1, room 0.11

Motivated by the inadequacy of single view approaches in many areas in NLP, this

thesis studies multi-view Chinese language processing, including word segmentation, part-of-speech tagging, syntactic parsing and semantic role labelling. We consider three situations of multiple views in statistical NLP:

- (1) Heterogeneous computational models have been designed for a given problem;
- (2) Heterogeneous annotation data is available to train systems;
- (3) Supervised and unsupervised machine learning techniques are applicable.

In practice, we study general approaches for model ensemble, annotation ensemble and learning ensemble problems in Chinese language processing. First, we comparatively analyze successful single view approaches for Chinese lexical, syntactic and semantic processing. Our analysis highlights the diversity between heterogeneous systems built on different views, and motivates us to improve the state-of-the-art by combining or integrating heterogeneous approaches. Second, we study the annotation ensemble problem, i.e. learning from multiple data sets under different annotation standards. We propose a series of generalized stacking models to effectively utilize heterogeneous labelled data to reduce approximation errors for word segmentation and parsing. Finally, we are concerned with bridging the gap between unsupervised and supervised learning paradigms. We introduce feature induction solutions that harvest useful linguistic knowledge from large-scale unlabelled data and effectively use them as new features to enhance discriminative learning based systems. Our empirical study indicates that multi-view processing is beneficial, especially when diverse single view approaches have complementary strengths, and our stacking-inspired ensemble approaches are very effective to explore multiple views, at least for learning Chinese language structures.

March

Christian SCHMALTZ

Compression, Pose Tracking, and Halftoning

(Advisor: Prof. Joachim Weickert)

23 March 2012, 16:15h, in building E1 7, room 0.01

In this thesis, we discuss image compression, pose tracking, and halftoning. Although these areas seem to be unrelated at first glance, they can be connected through video coding as application scenario. Our first contribution is an image compression algorithm based on a rectangular subdivision scheme which stores only a small subsets of the image points. From these points, the remained of the image is reconstructed using partial differential equations. Afterwards, we present a pose tracking algorithm that is able to follow the 3-D position and orientation of multiple objects simultaneously. The algorithm can deal with noisy sequences, and naturally handles both occlusions between different objects, as well as occlusions occurring in kinematic chains. Our third contribution is a halftoning algorithm based on electrostatic principles, which can easily be adjusted to different settings through a number of extensions. Examples include modifications to handle varying dot sizes or hatching. In the final part of the thesis, we show how to combine our image compression, pose tracking, and halftoning algorithms to novel video compression codecs. In each of these four topics, our algorithms yield excellent results that outperform those of other state-of-the-art algorithms.

Anna Katharina DEHOF

Novel Approaches for Bond Assignment and NMR Shift Prediction

(Advisor: Prof. Hans-Peter Lenhof)

Thu, 15 Mar 2012, 15:00 h, in building E2 1, room 001

Methoden des molekularen Modellierens gehören zu den Grundpfeilern moderner biologischer und pharmazeutischer Forschung. Akkurate Modelling-Methoden erfordern jedoch enormen Rechenaufwand, weshalb üblicherweise verschiedene Näherungsverfahren eingesetzt werden. Im Promotionsvortrag werden solche im Rahmen der Promotion entwickelten Näherungen für verschiedene Probleme aus der strukturbasierten Bioinformatik vorgestellt. Eine fundamentale Näherung der molekularen Physik ist die Einteilung chemischer Bindungen in wenige Klassen, meist in Form ganzzahliger Bindungsordnungen. In vielen Datensätzen ist diese Information nicht enthalten und eine automatische Zuweisung ist hochgradig schwierig. Für diese Problemstellung wird die BOA Constructor-Methode vorgestellt, die sowohl mit uneindeutigen Lösungen umgehen kann als auch vom Benutzer leicht erweitert werden kann. In umfangreichen Tests zeigen wir, dass unsere Methode dem bisherigen Stand der Forschung überlegen ist. Neben klassischen Anwendungen liefern Bindungsordnungen wertvolle Informationen für die statistische Vorhersage molekularer Eigenschaften wie z.B. der chemischen Verschiebung von Proteinatomen. Mit der von uns entwickelten NightShift-Pipeline wird ein Verfahren zur automatischen Generierung von Vorhersagemodellen präsentiert, wie z.B. dem Spinster-Modell, das den bisherigen manuell entwickelten Verfahren überlegen ist. Die Kombination mit BOA Constructor führt zum sogenannten Liops-Modell, welches als erstes Modell die effiziente Berücksichtigung des Einflusses von nicht-Proteinatomen erlaubt.

February

Hennig Lars ZIMMER

Correspondence Problems in Computer Vision

(Advisor: Prof. Joachim Weickert)

Fri, 10 Feb 2012, 16:15 h, in building E1 1, room 3.06

Correspondence problems like optic flow belong to the fundamental problems in computer vision. Here, one aims at finding correspondences between the pixels in two (or more) images. The correspondences are described by a displacement vector field that is often found by minimising an energy (cost) function. In this thesis, we present several contributions to the energy-based solution of correspondence problems:

(i) We start by developing a robust data term with a high degree of invariance under

illumination changes. Then, we design an anisotropic smoothness term that works complementary to the data term, thereby avoiding undesirable interference. Additionally, we propose a simple method for determining the optimal balance between the two terms.

(ii) When discretising image derivatives that occur in our continuous models, we show that adapting one-sided upwind discretisations from the field of hyperbolic differential equations can be beneficial. To ensure a fast solution of the nonlinear system of equations that arises when minimising the energy, we use the recent fast explicit diffusion (FED) solver in an explicit gradient descent scheme.

(iii) Finally, we present a novel application of modern optic flow methods where we align exposure series used in high dynamic range (HDR) imaging. Furthermore, we show how the alignment information can be used in a joint super-resolution and HDR method.

Shady ELBASSUONI

Effective Searching of RDF Knowledge Bases

(Advisor: Prof. Gerhard Weikum)

Mon, 6 Feb 2012, 15:00 h, in building E1 4 (MPI), room 024

RDF data has become a vital source of information for many applications. In this thesis, we present a set of models and algorithms to effectively search large RDF knowledge bases. These knowledge bases contain a large set of subject-predicate-object (SPO) triples where subjects and objects are entities and predicates express relationships between them. Searching such knowledge bases can be done using the W3C-endorsed SPARQL language or by similarly designed triple-pattern search. However, the exact-match semantics of triple-pattern search might fall short of satisfying the users needs by returning too many or too few results. Thus, IR-style searching and ranking techniques are crucial.

This thesis develops models and algorithms to enhance triple-pattern search. We propose a keyword extension to triple-pattern search that allows users to augment triple-pattern queries with keyword conditions. To improve the recall of triple-pattern search, we present a framework to automatically reformulate triple-pattern queries in such a way that the intention of the original user query is preserved while returning a sufficient number of ranked results. For efficient query processing, we present a set of top-k query processing algorithms and for ease of use, we develop methods for plain keyword search over RDF knowledge bases. Finally, we propose a set of techniques to diversify query results and we present several methods to allow users to interactively explore RDF knowledge bases to find additional contextual information about their query results.

Ulan DEGENBAEV

Formal Specification of the x86 Instruction Set Architecture

(Advisor: Prof. Wolfgang Paul)

Mon, 6 Feb 2012, 14:00 h, in building E1 3, room HS 003

In this thesis we specify the x86 instruction set architecture (ISA) by developing an abstract machine that models the behaviour of a modern computer with multiple x86 processors. Our model enables reasoning about low-level system software by providing formal interpretation of thousand pages of the processor vendor documentation written in informal prose.

We show how to reduce the problem of ISA formalization to two simpler problems: memory model specification and instruction semantics specification. We solve the former problem by extending the classical Total Store Ordering memory model with caches, translation-lookaside buffers, memory fences, locks, and other features of the x86 processor. For the latter problem we design a new domain-specific language which makes instruction semantics specification readable and compact.

January

Tim SCHWARTZ

The Always Best Positioned Paradigm for Mobile Indoor Applications

(Advisor: Prof. Wolfgang Wahlster)

Tues, 31 Jan 2012, 10:00 h, in the new DFKI building D3 4, VisRoom (-1.63)

In this dissertation, methods for personal positioning in outdoor and indoor environments are investigated. The Always Best Positioned paradigm, which has the goal of providing a preferably consistent self-positioning, will be defined.

Furthermore, the localization toolkit LOCATO will be presented, which allows to easily realize positioning systems that follow the paradigm. New algorithms were developed, which particularly address the robustness of positioning systems with respect to the Always Best Positioned paradigm. With the help of this toolkit, three example positioning-systems were implemented, each designed for different applications and requirements: a low-cost system, which can be used in conjunction with user-adaptive public displays, a so-called opportunistic system, which enables positioning with room-level accuracy in any building that provides a WiFi infrastructure, and a high-accuracy system for instrumented environments, which works with active RFID tags and infrared beacons. Furthermore, a new and unique evaluation-method for positioning systems is presented, which uses step-accurate natural walking-traces as ground truth. Finally, six location based services will be presented, which were realized either with the tools provided by LOCATO or with one of the example positioning-systems.

Mark KAMINSKI

Incremental Decision Procedures for Modal Logics with Nominals and Eventualities

(Advisor: Prof. Gert Smolka)

Tues, 24 Jan 2012, 16:15 h, in building E1 3, HS 002

This thesis contributes to the study of incremental decision procedures for modal logics with nominals and eventualities. Eventualities are constructs that allow to

reason about the reflexive-transitive closure of relations. Eventualities are an essential feature of temporal logics and propositional dynamic logic (PDL). Nominals extend modal logics with the possibility to reason about state equality. Modal logics with nominals are often called hybrid logics. Incremental procedures are procedures that can potentially solve a problem by performing only the reasoning steps needed for the problem in the underlying calculus.

We begin by introducing a class of syntactic models called demos and showing how demos can be used for obtaining nonincremental but worst-case optimal decision procedures for extensions of PDL with nominals, converse and difference modalities. We show that in the absence of nominals, such nonincremental procedures can be refined into incremental demo search procedures, obtaining a worst-case optimal decision procedure for modal logic with eventualities. We then develop the first incremental decision procedure for hybrid logic with eventualities, which we eventually extend to deal with hybrid PDL.

Sathish Chandra PAMMI

Synthesis of Listener Vocalizations - Towards Interactive Speech Synthesis

(Advisor: Prof. Hans Uszkoreit)

Mon, 16 Jan 2012, 16:00 h, building D3 2 (DFKI), Reuse seminar room

Spoken and multi-modal dialogue systems start to use listener vocalizations, such as uh-huh and mm-hm, for natural interaction. Generation of listener vocalizations is one of the major objectives of emotionally colored conversational speech synthesis. Success in this endeavor depends on the answers to three questions: Where to synthesize a listener vocalization? What meaning should be conveyed through the synthesized vocalization? And, how to realize an appropriate listener vocalization with the intended meaning? This thesis addresses the latter question. The investigation starts with proposing a three-stage approach: (i) data collection, (ii) annotation, and (iii) realization. The first stage presents a method to collect natural listener vocalizations from German and British English professional actors in a recording studio. In the second stage, we explore a methodology for annotating listener vocalizations -- meaning and behavior (form) annotation. The third stage proposes a realization strategy that uses unit selection and signal modification techniques to generate appropriate listener vocalizations upon user requests. Finally, we evaluate naturalness and appropriateness of synthesized vocalizations using perception studies. The work is implemented in the open source MARY text-to-speech framework, and it is integrated into the SEMAINE project's Sensitive Artificial Listener (SAL) demonstrator.

Catalin HRITCU

Union, Intersection, and Refinement Types and Reasoning About Type Disjointness for Security Protocol Analysis

(Advisor: Prof. Michael Backes)

Tues, 10 Jan 2012, 14:00h, building E1 1, R 4.07

In this thesis we present two new type systems for verifying the security of cryptographic protocol models expressed in a spi-calculus and, respectively, of protocol implementations expressed in a concurrent lambda calculus. The two type systems combine prior work on refinement types with union and intersection types and with the novel ability to reason statically about the disjointness of types. The increased expressivity enables the analysis of important protocol classes that were previously out of scope for the type-based analyses of cryptographic protocols. In particular, our type systems can statically analyze protocols that are based on zero-knowledge proofs, even in scenarios when certain protocol participants are compromised. The analysis is scalable and provides security proofs for an unbounded number of protocol executions. The two type systems come with mechanized proofs of correctness and efficient implementations.