SAARLAND UNIVERSITY

GRADUATE SCHOOL OF COMPUTER SCIENCE

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

SAARLAND UNIVERSITY
COMPUTER SCIENCE

max planck institut informatik

Max Planck Institute for Software Systems

German Research Center for Artificial Intelligence

2015

December

Bojan PEPIKJ
Richer Object Representations for Object Class Detection in Challenging Real-World Images
(Advisor: Prof. Bernt Schiele)
Mon, 21.12.2015, 15:00h, building E1 4, room 0.24

Object class detection in real world images has been a synonym for object localization for the longest time. State-of-the-art detection methods, inspired by renowned detection benchmarks, typically target 2D bounding box localization of objects. At the same time, due to the rapid technological and scientific advances, high-level vision applications, aiming at understanding the visual world as a whole, are coming into the focus. The diversity of the visual world challenges these applications in terms of representational complexity, robust inference and training data. As objects play a central role in any vision system, it has been argued that richer object representations, providing higher level of detail than modern detection methods, are a promising direction towards understanding visual scenes. Besides bridging the gap between object class detection and high-level tasks, richer object representations also lead to more natural object descriptions, bringing computer vision closer to human perception. Inspired by these prospects, this thesis explores four different directions towards richer object representations, namely, 3D object representations, fine-grained representations, occlusion representations, as well as understanding convnet representations.
Moreover, this thesis illustrates that richer object representations can facilitate high-level applications, providing detailed and natural object descriptions. In addition, the presented representations attain high performance rates, at least on par or often superior to state-of-the-art methods.

Zhongjie WANG
Pattern Search for Visualization
(Advisor: Prof. Hans-Peter Seidel)
Fri, 18.12.2015, 15:00h, building E1 4, room 0.19

The main topic of this thesis is pattern search in data sets for the purpose of visual data analysis. By giving a reference pattern, pattern search aims to discover similar occurrences in a data set with invariance to translation, rotation and scaling. To address this problem, we developed algorithms dealing with different types of data: scalar fields, vector fields, and line fields. For scalar fields, we use the SIFT algorithm (Scale-Invariant Feature Transform) to find a sparse sampling of prominent features in the data with invariance to translation, rotation, and scaling. Then, the user can define a pattern as a set of SIFT features by e.g. brushing a region of interest. Finally, we locate and rank matching patterns in the entire data set. Due to the sparsity and accuracy of SIFT features, we achieve fast and memory-saving pattern query in large scale scalar fields.
For vector fields, we propose a hashing strategy in scale space to accelerate the convolution-based pattern query. We encode the local flow behavior in scale space using a sequence of hierarchical base descriptors, which are pre-computed and hashed into a number of hash tables. This ensures a fast fetching of similar occurrences in the flow and requires only a constant number of table lookups.
For line fields, we present a stream line segmentation algorithm to split long stream lines into globally-consistent segments, which provides similar segmentations for similar flow structures. It gives the benefit of isolating a pattern from long and dense stream lines, so that our patterns can be defined sparsely and have a significant extent, i.e., they are integration-based and not local. This allows for a greater flexibility in defining features of interest. For user-defined patterns of curve segments, our algorithm finds similar ones that are invariant to similarity transformations.
Additionally, we present a method for shape recovery from multiple views. This semi-automatic method fits a template mesh to high-resolution normal data. In contrast to existing 3D reconstruction approaches, we accelerate the data acquisition time by omitting the structured light scanning step of obtaining low frequency 3D information.

Stefan SCHUH
Understanding Fundamental Database Operations on Modern Hardware
(Advisor: Prof. Jens Dittrich)
Fri, 18.12.2015, 11:00h, building E1 1, room 407

We live in an interconnected digital society, where many companies like e.g. Google, Facebook, and Twitter gather and manage tremendous amounts of data every day.Several different architectures have evolved to cope with these vast amount of data over the years.
Traditionally, mainframes were used to handle huge amounts of data. However, the mainframe has to renew itself to allow for modern data analytics to be efficient and affordable. Advances in the main memory capacity led to the development of in-memory databases architectures, run on many-core non-uniform memory access (NUMA) machines that can handle terabytes of data on a single machine.
As another architecture Google developed MapReduce, a distributed framework for data processing on hundreds or even thousands of commodity machines, to handle

data that cannot be stored or processed by a single machine, even if it has a capacity in the range of terabytes.

In this talk I will focus on relational equi-join algorithms on modern many-core NUMA servers with large main memory capacities and introduce our own variants of those join algorithms.

Sven BUGIEL
Establishing Mandatory Access Control on Android OS
(Advisor: Prof. Michael Backes)
Mon, 14.12.2015, 14:00h, building E1 7, room 0.01

Common characteristic of all mobile operating systems for smart devices is an extensive middleware that provides a feature-rich API for the onboard sensors and user's data (e.g., contacts). To effectively protect the device's integrity, the user's privacy, and to ensure non-interference between mutually distrusting apps, it is imperative that the middleware enforces rigid security and privacy policies. This thesis presents a line of work that integrates mandatory access control (MAC) mechanisms into the middleware of the popular, open source Android OS. While our early work established a basic understanding for the integration of enforcement hooks and targeted very specific use-cases, such as multi-personna phones, our most recent works adopt important lessons learned and design patterns from established MAC architectures on commodity systems and intertwine them with the particular security requirements of mobile OS architectures like Android. Our most recent work also complemented the Android IPC mechanism with provisioning of better provenance information on the origins of IPC communication. Such information is a crucial building block for any access control mechanism on Android. Lastly, this dissertation outlines further directions of ongoing and future research on access control on modern mobile operating systems.

Mohamed Amir YOSEF
U-AIDA: A Customizable System for Named Entity Recognition, Classification, and Disambiguation
(Advisor: Prof. Gerhard Weikum)
Fri, 11.12.2015, 15:00h, building E1 5, room 0.29

Recognizing and disambiguating entities such as people, organizations, events or places in natural language text are essential steps for many linguistic tasks such as information extraction and text categorization. A variety of named entity disambiguation methods have been proposed, but most of them focus on Wikipedia as a sole knowledge resource. This focus does not fit all application scenarios, and customization to the respective application domain is crucial.
This dissertation addresses the problem of building an easily customizable system for named entity disambiguation. The first contribution is the development of a universal and flexible architecture that supports plugging in different knowledge resources. The second contribution is utilizing the flexible architecture to develop two domain-specific disambiguation systems. The third contribution is the design of a complete pipeline for building disambiguation systems for languages other than English that have poor annotated resources such as Arabic. The fourth contribution is a novel approach that performs fine-grained type classification of names in natural language text.

Simon OLBERDING
Fabricating Custom-Shaped Thin-Film Interactive Surfaces
(Advisor: Dr. Jürgen Steimle)
Thu, 10.12.2015, 10:00h, building E1 7, room 0.01

Novel possibilities in digital fabrication empower end-users to design and fabricate custom objects and surfaces. But in order to equip these with interactivity, the user has to fall back on existing display and touch surfaces, which come in a predefined resolution and shape. This constrains the shape and consequently the application areas of the interactive device. This work presents three approaches, based on printed electronics, for designing and fabricating interactive surfaces with an individual shape and functionality on flexible materials.
The first contribution is a design and fabrication approach for custom and flexible touch-displays. We present display primitives to support their digital design. In order for the end-user to physically instantiate the design, we show two practicable printing processes for electroluminescent displays. Additionally, a technical solution for adding touch sensitivity is provided. In the second contribution, we extend this work with a novel design and fabrication approach for interactive folded 3D objects. In an integrated digital design environment, the user annotates 3D models with user interface elements. Based on the design, folding and printing patterns including interactive elements are automatically generated. The third contribution focuses on manual forming. We show a flexible multi-touch sensor foil, which can be cut using physical tools while remaining functional. We contribute novel patterns for the internal wiring of the sensor to support a variety of shapes.

Ahmed ELHAYEK
Marker-less Motion Capture in General Scenes with Sparse Multi-camera Setups
(Advisor: Prof. Christian Theobalt)
Wed, 9.12.2015, 10:00h, building E1 47, room 0.19

Human motion-capture from videos is one of the fundamental problems in computer vision and computer graphics. Its applications can be found in a wide range of industries. Even with all the developments in the past years, industry and academia alike still rely on complex and expensive marker-based systems. Many state-of-the-art marker-less motion-capture methods come close to the performance of marker-based algorithms, but only when recording in highly controlled studio environments with exactly synchronized, static and sufficiently many cameras. While relative to marker-based systems, this yields an easier apparatus with a reduced setup time, the hurdles towards practical application are still large and the costs are considerable. By being constrained to a controlled studio, marker-less methods fail to fully play out their advantage of being able to capture scenes without actively modifying them.
In the area of marker-less human motion-capture, this thesis proposes several novel algorithms for simplifying the motion-capture to be applicable in new general outdoor

scenes. The first is an optical multi-video synchronization method which achieves subframe accuracy in general scenes. In this step, the synchronization parameters of multiple videos are estimated. Then, we propose a spatio-temporal motion-capture method which uses the synchronization parameters for accurate motion-capture with unsynchronized cameras. Afterwards, we propose a motion capture method that works with moving cameras, where multiple people are tracked even in front of cluttered and dynamic backgrounds with potentially moving cameras. Finally, we reduce the number of cameras employed by proposing a novel motion-capture method which uses as few as two cameras to capture high-quality motion in general environments, even outdoors. The methods proposed in this thesis can be adopted in many practical applications to achieve similar performance as complex motion-capture studios with a few consumer-grade cameras, such as mobile phones or GoPros, even for uncontrolled outdoor scenes.

Quan NGUYEN
Effizienz und Ergonomie von Multitouch-Interaktion
(Advisor: Prof. Michael Kipp, now Uni Augsburg)
Mon, 7.12.2015, 10:30h, building D3 2, room -1.63 (DFKI VisCenter)

Die vorliegende Arbeit beschäftigt sich mit Grundfragen der Effektivität, Effizienz und Zufriedenheit von Multitouch-Interaktionen. Mithilfe einer Multitouch-Steuerung für 3D- Animation konnte gezeigt werden, dass selbst unerfahrene Multitouch-Nutzer in der Lage sind, hoch komplexe Aufgaben koordiniert und effizient zu lösen. Ein neu entwickeltes Koordinationsmaß bestätigt, dass Nutzer den Vorteil eines Multitouch nutzen, indem sie koordiniert mehrere Finger gleichzeitig für 3D-Animationen in Echtzeit einsetzen.
In drei weiteren Studien zu zentralen Multitouch-Interaktionstechniken konnte gezeigt werden, dass
die Originalformulierung von Fitts' Gesetz nicht ausreicht, um die Effizienz von Multitouch-Interaktionen adäquat zu bewerten und zu analysieren. Fitts' Gesetz ist ein Modell zur Vorhersage und Analyse von Interaktionszeiten und beinhaltet ursprünglich nur die Distanz der Interaktionsbewegung und die Zielgröße. Diese Arbeit zeigt, dass Vorhersagen mit Fitts' Gesetz bessere Ergebnisse liefern, wenn sie neben diesen beiden Faktoren auch Bewegungsrichtung, Startpunkt der Bewegung und Neigung des Multitouch- Display berücksichtigen.
Die Ergebnisse dieser Arbeit liefern Anhaltspunkte, um effiziente und benutzerfreundliche Interaktionstechniken zu entwickeln. Zudem könnten sie eingesetzt werden, um Analysen von Intertaktionstechniken für Multitouch teilautomatisch durchzuführen.

November

Peter BACKES
Cluster Abstraction of Graph Transformation Systems
(Advisor: Prof. Reinhard Wilhelm)
Fri, 13.11.2015, 14:00h, building E1 7, room 0.01

This dissertation explores the problem of analyzing the reachable graphs of graph transformation systems. Such systems rewrite graphs according to subgraph replacement rules; we allow negative application conditions to be specified in addition. This problem is hard because the number of reachable graphs is potentially unbounded.
We use abstract interpretation to compute a finite, overapproximated representation of the reachable graphs. The main idea is the notion of a cluster: We abstract the graph locally for each of its nodes such that we obtain a bounded cluster with the node and its immediate neighborhood. Then, we eliminate duplicate clusters such that we obtain a bounded abstraction for the entire graph. We lift concrete rule application to this abstraction, eventually obtaining an overapproximation of all reachable graphs.
We present ASTRA, an implementation of cluster abstraction, and the merge protocol from car platooning, our main test case. This protocol enables autonomous cars to form and merge platoons consisting of a leader car and several followers, such that the leader controls speed and lane. The abstraction does well with the merge protocol, and also manages to analyze several other standard case studies from the literature, as well as test cases automatically generated from a higher-level formalism.

Fatemeh TORABI ASR
An Information Theoretic Approach to Production and Comprehension of Discourse Markers
(Advisor: Dr. Vera Demberg)
Tues, 10.11.2015, 10:00h, building C7 4, room 1.17

Discourse relations are the building blocks of a coherent text. The most important linguistic elements for constructing these relations are discourse markers. The presence of a discourse marker between two discourse segments provides information on the inferences that need to be made for interpretation of the two segments as a whole (e.g., because marks a reason). This thesis presents a new framework for studying human communication at the level of discourse by adapting ideas from information theory. A discourse marker is viewed as a symbol with a measurable amount of relational information. This information is communicated by the writer of a text to guide the reader towards the right semantic decoding. To examine the information theoretic account of discourse markers, we conduct empirical corpus-based investigations, offline crowd-sourced studies and online laboratory experiments. The thesis contributes to computational linguistics by proposing a quantitative meaning representation for discourse markers and showing its advantages over the classic descriptive approaches. For the first time, we show that readers are very sensitive to the fine-grained information encoded in a discourse marker obtained from its natural usage and that writers use explicit marking for less expected relations in terms of linguistic and cognitive predictability. These findings open new directions for implementation of advanced natural language processing systems.

Yongjie YANG
A Deep Exploration of the Complexity Border of Strategic Voting Problems

(Advisor: Prof. Jiong Guo - now ShanDong University)
Mon, 02.11.2015, 16:00h, building E1 7, room 0.01

Voting has found applications in a variety of areas, including multiagent systems, expert recommendation systems, political elections, protein recognition, etc. A major concern in a voting is that there may be a powerful agent (e.g., the chairman or a millionaire) who has an incentive to manipulate the voting. A prominent way to address this issue is to use computational complexity as a barrier. The key point is that if it is NP-hard for the powerful agent to figure out how to exactly manipulate the voting in order to reach his goal, the agent may give up his manipulative plan. Fortunately, researchers have proved that many strategic votingproblems are NP-hard in general (where the domain of the preferences of the voters is not restricted). Unfortunately, many NP-hardness results turned out to be polynomial-time solvable in restricted elections, such as in single-peaked elections. Since in many real-world applications, the preferences of the voters fulfill some combinatorial properties, these polynomial-time solvability results in restricted elections bring back the possibility of a powerful agent to manipulate the voting in practice. This thesis is aimed to pinpoint the (parameterized) complexity border between very restricted elections (e.g., single-peaked elections) and general elections for numerous strategic voting problems. We derived NP-hardness, polynomial-time solvability, W-hardness and fixed-parameter tractability results for the problems studied.

October

Martin SUDA
Resolution-based methods for linear temporal reasoning
(Advisor: Prof. Christoph Weidenbach - co-tutelle with Prague)
Fri, 16.10.2015, 10:00h, building E1 5, room 002

The aim of this thesis is to explore the potential of resolution-based methods for linear temporal reasoning. On the abstract level, this means to develop new algorithms for automated reasoning about properties of systems which evolve in time. More concretely, we will:
1) show how to adapt the superposition framework to proving theorems in propositional Linear Temporal Logic (LTL),
2) use a connection between superposition and the CDCL calculus of modern SAT solvers to come up with an efficient LTL prover,
3) specialize the previous to reachability properties and discover a close connection to Property Directed Reachability (PDR), an algorithm recently developed for model checking of hardware circuits,
4) further improve PDR by providing a new technique for enhancing clause propagation phase of the algorithm, and
5) adapt PDR to automated planning by replacing the SAT solver inside with a planning-specific procedure.
We implemented the proposed ideas and provide experimental results which demonstrate their practical potential on representative benchmark sets. Our system LS4 is shown to be the strongest LTL prover currently publicly available. The mentioned enhancement of PDR substantially improves the performance of our implementation of the algorithm for hardware model checking in the multi-property setting. Finally, our planner PDRplan has been compared with the state-of-the-art planners on the benchmarks from the International Planning Competition with very promising results.

September

Ahmad BARGHASH
Identifying biological associations from high-throughput datasets
(Advisor: Prof. Volkhard Helms)
Wed, 30.09.2015, 10:00h, building E2 1 (CBI), room 0.01

The omnipresence of huge High-throughput biological datasets creates the urgent need for efficient and robust computational approaches to handle and analyze such databases and to identify informative associations. In the first part of this thesis, we analyzed amino acid datasets of membrane transporters from different organisms for the purpose of transferring functional annotations of the transporters across species. Our second computational approach used several outlier detection algorithms for detecting sample and gene outliers in expression/methylation datasets. In the next step, we used published expression and methylation datasets from GEO to analyse and confirm possible tumor markers for HCC, liver diseases, and breast cancer. The final part of this thesis dealt with large scale exon expression, methylation, and chromatin modification datasets for 11 different developmental stages from the Human Epigenome Atlas. In this genome wide analysis we aimed to identify cases of differential exon usage in different dataset. The approaches presented in this thesis may advance the current stages of tumor marker identification.

Tomasz TYLENDA
Methods and Tools for Summarization of Entities and Facts in Knowledge Bases
(Advisor: Prof. Gerhard Weikum)
Mon, 28.09.2015, 11:00h, building E1 4, MPII, room 0.24

Knowledge bases have become key assets for search and analytics over large document corpora. They are used in applications ranging from highly specialized tasks in bioinformatics to general purpose search engines. The large amount of structured knowledge they contain calls for effective summarization and ranking methods.
The goal of this dissertation is to develop methods for automatic summarization of entities in knowledge bases, which also involves augmenting them with information about the importance of particular facts on entities of interest. We make two main contributions.
First, we develop a method to generate a summary of information about an entity using the type information contained in a knowledge base. We call such a summary a semantic snippet. Our method relies on having importance information about types, which is external to the knowledge base. We show that such information can

be obtained using human computing methods, such as Amazon Mechanical Turk, or extracted from the edit history of encyclopedic articles in Wikipedia.

Our second contribution is linking facts to their occurrences in supplementary documents. Information retrieval on text uses the frequency of terms in a document to judge their importance. Such an approach, while natural, is difficult for facts extracted from text. This is because information extraction is only concerned with finding any occurrence of a fact. To overcome this limitation we propose linking known facts with all their occurrences in a process we call fact spotting. We develop two solutions to this problem and evaluate them on a real world corpus of biographical documents.

Oliver DEMETZ
Feature Invariance versus Change Estimation in Variational Motion Estimation
(Advisor: Prof. Joachim Weickert)
Fri, 25 Sept 2015, 15:15h, building E1 7 (MMCI), room 0.01

The robust estimation of correspondences in image sequences belongs to the fundamental problems in computer vision. One of the big practical challenges are appearance changes of objects, because the traditional brightness constancy assumption only holds under idealised conditions. Two chapters of this thesis approach this problem in the context of dense optic flow estimation. We contribute two solutions that follow very different strategies. First, we consider invariances, i.e. features of the input images that remain unaffected under certain changes in illumination and appearance. We present a systematic overview of available invariances and introduce our complete rank transform. Our prototypical variational framework can incorporate any of the discussed invariant features and facilitates a fair comparison. Our second contribution is motivated by the insight that challenging appearance changes such as drop shadows often are local phenomena, but invariances operate globally. Thus, we develop a method that integrates illumination changes explicitly into our model and compensates the constancy assumption locally for them. The third contribution of this thesis considers the scale space behaviour of variational optic flow methods. The existence of such a behaviour in the regularisation parameter is unquestionable, however, there is no through analysis in this direction up to now. We close this gap as the third contribution of this thesis.

Tim DAHMEN
Tomographic Reconstruction of Combined Tilt- and Focal Series in Scanning Transmission Electron Microscopy
(Advisor: Prof. Philipp Slusallek)
Thu, 24 Sept 2015, 10:00h, building D3 4 (DFKI), Visenter, NB -1.63

In this thesis, we investigate tomography using scanning transmission electron microscopy (STEM) with limited depth of field. A combined tilt- and focal series (CTFS) is proposed as a new recording scheme for high-angle annular dark-field STEM tomography. Hereby, three-dimensional (3D) data is acquired by mechanically tilting the specimen and at each tilt direction recording a series of images with different focal depth (focal series).

In a theoretic part, the STEM transform is introduced as a generalization of the well-known ray transform for parallel illumination that considers the convergent shape of an electron beam in aberration corrected STEM. The STEM transform is investigated analytically and it is shown that it is (1) a linear convolution and (2) self-adjoint.

Additionally, we introduce an iterative algorithm to solve the problem of tomographic reconstruction for data recorded with this new scheme. The algorithm uses the Kaczmarz method to solve the system Ax=b in a least-squares sense. By experimental evaluation of the algorithm we show that the method significantly reduces the artifacts known as axial elongation, i.e. leads to a more isotropic resolution than pure tilt series based approaches at the same tilt range as well as pure focal series methods.

Michael KARL
Optimized Wide Area Media Transport Strategies
(Advisor: Prof. Thorsten Herfet)
Wed, 30 Sept 2015, 13:15, building E1 1, room 407

Modern networks are increasingly used for a multitude of different application scenarios, which primarily include multimedia transmission with a strict delivery time constraint. This kind of content has far more complex requirements than traditional applications, such as file transfer via download or sending an email. This thesis motivates the segmentation of default IP-based communication paths for optimized network and transport utilization concerning retransmission error coding schemes and strict delivery time limits. Theoretical considerations for general channel coding approaches with no time limitation show, that the network load measurably decreases in case the transmission is fully split. Time-restricted transmissions with packet-retransmission error correction schemes require more complex mechanisms for the time budget distribution and retransmission characteristics. The development of an upper bound for the number of segments established on a transmission path with hybrid error correction and a time constraint represents a significant contribution of this thesis. Metrics are presented that enable a reliable identification of network segments with a high optimization potential. Mathematical programming and graph theoretical methods demonstrate their versatility. Experimental measurements and simulations verify that network path segmentation leads to a significant reduction of the network load.

Manuel LAMOTTE-SCHUBERT
Automatic Authorization Analysis
(Advisor: Prof. Christoph Weidenbach)
Fri, 18 Sept 2015, 15:00, building E1 4, room 0.24

Today many companies use an ERP (Enterprise Resource Planning) system such as the SAP system to run their daily business ranging from financial issues down to the actual control of a production line. These systems are very complex from the view of administration of authorizations and include a high potential for errors. The administrators need support to verify their decisions on changes in the authorization setup of such systems and also assistance to implement planned changes error-free.

First-order theorem proving is a reliable and correct method to offer this support to administrators at work. But it needs on the one hand a corresponding formalization of an SAP ERP system instance in first-order logic, and on the other hand, a sound and terminating calculus that can deal with the complexity of such systems. Since first-order logic is well-known to be undecidable in general, current research deals with the challenge of finding suitable and decidable sub-classes of first-order logic which are then usable for the mapping of such systems.

The thesis presents a (general) new decidable first-order clause class, named BDI (Bounded Depth Increase), which naturally arose from the requirement to assist administrators at work with real-world authorization structures in the SAP system and the automatic proof of properties in these systems. The new class strictly includes known classes such as PVD.

The arity of function and predicate symbols as well as the shape of atoms is not restricted in BDI as it is for many other classes. Instead the shape of "cycles" in resolution inferences is restricted such that the depth of generated clauses may increase but is still finitely bound.

The thesis shows that the Hyper-resolution calculus modulo redundancy elimination terminateson BDI clause sets. Further, it employs this result to the Ordered Resolution calculus which is also proved to terminate on BDI, and thus yielding a more efficient decision procedure which is able to solve real-world SAP authorization instances.

The test of conditions of BDI has been implemented into the state-of-the art theorem prover SPASS in order to be able to detect decidability for any given problem automatically. The implementation has been applied to the problems of the TPTP Library in order to detect potential new decidable problems satisfying the BDI class properties and further to find non-terminating problems "close" to the BDI class having only a few clauses which are responsible for the undecidability of the problem.

Christine RIZKALLAH
Verification of Program Computations
(Advisor: Prof. Kurt Mehlhorn)
Fri, 18 Sept 2015, 9:00, building E1 4, room 0.24

Formal verification of complex algorithms is challenging. Verifying their implementations in reasonable time is infeasible using current verification tools and usually involves intricate mathematical theorems. Certifying algorithms compute in addition to each output a witness certifying that the output is correct. A checker for such a witness is usually much simpler than the original algorithm -- yet it is all the user has to trust. The verification of checkers is feasible with current tools and leads to computations that can be completely trusted.

In this talk, we describe a framework to seamlessly verify certifying computations. We demonstrate the effectiveness of our approach by presenting the verification of typical examples of the industrial-level and widespread algorithmic library LEDA. We present and compare two alternative methods for verifying the C implementation of the checkers.

Moreover, we present work that was done during an internship at NICTA. This work contributes to building a feasible framework for verifying efficient file systems code. As opposed to the algorithmic problems we address, file systems code is mostly straightforward and hence a good candidate for automation.

August

Fidaa ABED
Coordinating Selfish Players in Scheduling Games
(Advisor: Prof. Kurt Mehlhorn)
Tues, 18 Aug 2015, 11:00, building E1 5, room 0.29

We investigate coordination mechanisms that schedule n jobs on m unrelated machines. The objective is to minimize the makespan. It was raised as an open question whether it is possible to design a coordination mechanism that has constant price of anarchy using preemption. We give a negative answer. Next we introduce multi-job players that control a set of jobs, with the aim of minimizing the sum of the completion times of theirs jobs. In this setting, previous mechanisms designed for players with single jobs are inadequate, e.g., having large price of anarchy, or not guaranteeing pure Nash equilibria. To meet this challenge, we design three mechanisms that induce pure Nash equilibria while guaranteeing relatively small price of anarchy.

Then we consider multi-job players where each player's objective is to minimize the weighted sum of completion time of her jobs, while the social cost is the sum of players' costs. We first prove that if machines order jobs according to Smith-rule, then the coordination ratio is at most 4, moreover this is best possible among non-preemptive policies. Then we design a preemptive policy, {\em externality} that has coordination ratio 2.618, and complement this result by proving that this ratio is best possible even if we allow for randomization or full information. An interesting consequence of our results is that an ε-local optima of $R|\,|\sum w_iC_i$ for the jump neighborhood can be found in polynomial time and is within a factor of 2.618 of the optimal solution.

Xiaoqi CAO
Semantic Search and Composition in Unstructured Peer-to-Peer Networks
(Advisor: Dr. Matthias Klusch)
Mon, 17 Aug 2015, 10:00, building D3 2, VIS-Center NB -1.63

This dissertation focuses on several research questions in the area of semantic search and composition in unstructured peer-to-peer (P2P) networks. Going beyond the state of the art, the proposed semantic-based search strategy S2P2P offers a novel path-suggestion based query routing mechanism, providing a reasonable tradeoff between search performance and network traffic overhead.

In addition, the first semantic-based data replication scheme DSDR is proposed. It enables peers to use semantic information to select replica numbers and target peers to address predicted future demands. With DSDR, k-random search can achieve better precision and recall than it can with a near-optimal non-semantic replication strategy. Further, this thesis introduces a functional automatic semantic service composition method, SPSC. Distinctively, it enables peers to jointly compose

complex workflows with high cumulative recall but low network traffic overhead, using heuristic-based bidirectional chaining and service memorization mechanisms. Its query branching method helps to handle dead-ends in a pruned search space. SPSC is proved to be sound and a lower bound of its completeness is given. Finally, this thesis presents iRep3D for semantic-index based 3D scene selection in P2P search. Its efficient retrieval scales to answer hybrid queries involving conceptual, functional and geometric aspects. iRep3D outperforms previous representative efforts in terms of search precision and efficiency.

Fahimeh RAMEZANI
Application of Multiplicative Weights Update Method in Algorithmic Game Theory
(Advisor: Prof. Kurt Mehlhorn)
Mon, 17 Aug 2015, 11:00, building E1 7, room 0.01

In this thesis, we apply the Multiplicative Weights Update Method (MWUM) to the design of approximation algorithms for some optimization problems in game-theoretic settings. Lavi and Swamy introduced a randomized mechanism for combinatorial auctions that uses an approximation algorithm for the underlying optimization problem, so-called social welfare maximization and converts the approximation algorithm to a randomized mechanism that is truthful-in-expectation, which means each player maximizes its expected utility by telling the truth. The mechanism is powerful, but unlikely to be efficient in practice, because it uses the Ellipsoid method.
We follow the general scheme suggested by Lavi and Swamy and replace the Ellipsoid method with MWUM. This results in a faster and simpler approximately truthful-in-expectation mechanism. We also extend their assumption regarding the existence of an exact solution for the LP-relaxation of social welfare maximization. We assume that there exists an approximation algorithm for the LP and establish a new randomized approximation mechanism.
We also consider the problem of computing an approximate saddle point, or equivalently equilibrium, for a convex-concave functions $F: X \times Y \to \RR$, where X and Y are convex sets of arbitrary dimensions. Our main contribution is the design of a randomized algorithm for computing an $\eps$-approximation saddle point for F. Our algorithm is based on combining a technique developed by Grigoriadis and Khachiyan, which is a randomized variant of Brown's fictitious play, with the recent results on random sampling from convex sets.

Mohamed HAMED Ali Fahmy
Integrative computational approaches for studying stem cell differentiation and complex diseases
(Advisor: Prof. Volkhard Helms)
Mon, 10 Aug 2015, 12:00, building E2 1, room 0.01

The biological functions of the molecular components (genes, proteins, miRNAs, siRNAs,..etc) of biological cells and mutations/perturbations thereof are tightly connected with cellular malfunctions and disease pathways. These molecular elements interact with each other forming a complex interwoven regulatory machinery that governs, on one hand, regular cellular pathways, and on the other hand, their dysregulation or malfunction in pathological processes. Therefore, revealing these critical molecular interactions in complex living systems is being considered as one of the major goals of current systems biology.
In this dissertation, we introduce practical computational approaches implemented as freely available software tools to integrate heterogeneous sources of large-scale genomic data and unravel the combinatorial regulatory interactions between different molecular elements. We employed these approaches to investigate the molecular mechanisms of cellular differentiation (namely hematopoiesis) as an example for biological processes and human breast cancer and diabetes as examples for complex diseases. The provided topological and functional analyses of our approaches as validated on cellular differentiation and complex diseases promotes them as reliable systems biology tools for researchers across the life science communities.

July

Jochen FREY
ASaP – Integrationsplattform für Smart Services in Intelligenten Umgebungen
(Advisor: Prof. Wolfgang Wahlster)
Tue, 21 July 2015, 14:00, building D3 2, room -2.17 (Reuse)

Die Entwicklung von mit vernetzten Sensoren und Aktuatoren ausgestatteten Wohnungen, sogenannten Smart Homes, ist ein wichtiger Zukunftstrend in Deutschland. Laut aktuellen Studien wird die Zahl der Smart Homes in Deutschland bis spätestens 2020 die Millionenmarke erreicht haben. Das übergeordnete Forschungsgebiet von Smart Homes sind Intelligente Umgebungen, welche im Allgemeinen alle räumlichen Umgebungen umfassen, die dem Benutzer eine Form von intelligenter Assistenz und Unterstützung bereitstellen. Die vorliegende Arbeit beschäftigt sich mit der Fragestellung, wie neuartige digitale Mehrwertdienste in Form von sogenannten Smart Services innerhalb einer Intelligenten Umgebungen geschaffen und nahtlos darin integriert werden können. Die Aufgabenstellung beinhaltet zwei grundlegende Herausforderungen:
- Der Kunde als Bewohner eines zukünftigen Smart Homes soll durch ein durchgängiges Marktplatzkonzept dazu befähigt werden, auf seine Situation passende Smart Services zu finden und auf einer zentralen Plattform zu installieren. Bisherige Ansätze fokussieren oftmals auf einen einzigen expliziten Anwendungsfall. Eine allgemeine und dienstübergreifende Interoperabilität der Services ist somit nicht gewährleistet.
- Den Dienstanbietern und Entwicklern müssen Werkzeuge und Entwicklungsumgebungen an die Hand gegeben werden, mit denen Smart Services schnell und effizient konzipiert, umgesetzt und auf einem zentralen Marktplatz verteilt werden können.
In der vorliegenden Arbeit wird ein allgemeines Integrationskonzept vorgestellt, welches sowohl die Ansätze und Konzepte des bestehenden Industriestandards Universal Remote Console als auch die zugrunde liegende Referenzarchitektur der Smart Service Welt berücksichtigt.

Aufbauend auf einer einheitlichen semantischen Modellierung einer Intelligenten Umgebung wird neben einem Konzept zur Realisierung von personalisierten Benutzerschnittstellen auch ein auf einem Multiagentensystem aufbauende Plattform zur dienstübergreifenden Integration von Smart Services vorgestellt. Die Schwerpunkte der Arbeit liegen auf der Konzeption, dem Entwurf und der Realisierung einer offenen, hochflexiblen, erweiterbaren und auf Standards basierenden Softwarearchitektur für Smart Services im Kontext von Smart Home Technologien.

Ali POURMIRI
Random Walk-Based Algorithms on Networks
(Advisor: Prof. Kurt Mehlhorn)
Fri, 3 July 2015, 9:30, building E1 7, room 0.01

The \emph{Push} and \emph{Push-Pull} protocols are basic randomized protocols for information dissemination on networks. We generalize the Push and Push-Pull protocols so that each node, in every round, contacts a specified number of neighbors to exchange the rumor, where the number assigned to each node is chosen independently according to a probability distribution R. We then provide both lower and upper bounds on the rumor spreading time on a complete network with n nodes depending on statistical properties of R such as the mean or the variance. For instance, we show that if R is power law distribution with exponent $\beta\in(2,3)$, then the rumor spreads in $O(\log n\log n)$ rounds with high probability. Random k-trees and random k-Apollonian networks are random graph models that are suitable to model poorly connected, small-world and scale free networks. We analyze the behavior of the standard Push-Pull protocol on these networks and show that the Push-Pull protocol propagates a rumor from a randomly chosen informed node to almost all nodes in at most $\log^2 n$ rounds with high probability. In addition to that we derive a lower bound of $n^{\Omega(1)}$ for the runtime of the protocol to inform all nodes of random k-trees. Besides the rumor spreading protocols, we also propose a new random walk-based algorithm for sequentially allocating n balls into n bins that are organized as a d-regular graph with n nodes, say G, where $d\ge 3$ can be any integer. Provided G has a sufficiently large girth, we establish an upper bound for the maximum number of balls at any bin after allocating n balls by the algorithm.

Willem HAGEMANN
Symbolic Orthogonal Projections: A New Polyhedral Presentation for Reachability Analysis of Hybrid Systems
(Advisor: Prof. Christoph Weidenbach)
Thu, 2 July 2015, 11:00, building E1 4, room 0.24

My thesis deals with the reachability analysis of hybrid systems with continuous dynamics defined by linear differential equations with inputs. Special importance is given to the treatment of the necessary geometrical operations. After a short introduction into reachability analysis of hybrid systems, I will discuss a new representation class for convex polyhedra, the symbolic orthogonal projections (sops). Fundamental geometrical operations, like affine transformations, intersections, Minkowski sums, and convex hulls, can be performed by simple block matrix operations on the representation. Compared to traditional representations, like half-space representations, vertex representations, or representations by support functions, this is a unique feature of sops. Linear programming helps us to evaluate sops. In the second part of my talk, I will discuss the application of sops in reachability analysis of hybrid systems. It turns out that sops are well-suited for the discrete computations. Thereupon, I demonstrate also the applicability of sops for the continuous computation of reachability analysis. Additionally, I present experimental results which show that sops allow an efficient and accurate computation of the reachable states.

June

Esfandiar MOHAMMADI
Abstracting Cryptographic Protocols
(Advisor: Prof. Michael Backes)
Mon, 29 June 2015, 10:30, building E1 7, room 0.01

Cryptographic protocols can be used to harden the security of IT systems, but only if these protocols are carefully woven into the system. Thus, a security analysis of the system overall is necessary to ensure that the cryptographic protocols are effectively applied. The classical proof technique for computational security proofs — reduction proofs — are, due to their complexity, hard to automatically derive, already for medium-sized IT systems. Hence, there is a successful line of research about abstracting cryptographic protocols.
In this talk, we consider two kinds of abstractions: i) symbolic abstractions, where formal rules characterize the attacker's capabilities and the functionality of the cryptographic operations, and ii) ideal functionalities, where all honest parties are replaced by one single incorruptible machine. For symbolic abstractions we study their accuracy (called computational soundness), i.e., under which conditions these abstractions capture all cryptographic attacks. We establish a computational soundness result for malleable zero-knowledge proofs, and we establish two results for re-using existing computational soundness proofs, in particular for obtaining strong equivalence properties.
Additionally, we devise an ideal functionality for the most-widely used anonymous communication protocol Tor and (using the UC framework) prove its accuracy. For improving Tor's performance, we moreover propose a novel, provably secure key-exchange protocol.

Pedro FONSECA
Effective Testing for Concurrency Bugs
(Advisor: Prof. Rodrigo Rodrigues)
Wed, 24 June 2015, 13:00, building E1 5, room 0.29

In the current multi-core era, concurrency bugs are a serious threat to software reliability. As hardware becomes more parallel, concurrent programming will become increasingly pervasive. However, correct concurrent programming is known to be

extremely challenging for developers and can easily lead to the introduction of concurrency bugs. This talk addresses this challenge by proposing novel techniques to help developers expose and detect concurrency bugs.

First, I will briefly present the results of a bug study that analyzed the external and internal effects of real-world concurrency bugs. This study revealed that a significant fraction of concurrency bugs qualify as semantic or latent bugs, which are two particularly challenging classes of concurrency bugs. Based on the insights from the study, I will present a concurrency bug detector, PIKE, that analyzes the behavior of program executions to infer whether concurrency bugs have been triggered during a concurrent execution. In addition, I will present the design of a testing tool, SKI, that allows developers to test operating system kernels for concurrency bugs in a practical manner. SKI bridges the gap between user-mode testing and kernel-mode testing by enabling the systematic exploration of the kernel thread interleaving space. Both PIKE and SKI have been shown to be effective at finding concurrency bugs.

Radu CURTICAPEAN
The simple, little and slow things count
(Advisor: Prof. Markus Bläser)
Mon, 22 June 2015, 14:30, building E1 7, room 0.01

The study of counting problems has become a classical subfield of computational complexity theory since Valiant's seminal paper introduced the complexity class #P and proved that counting perfect matchings is complete for this class. Since then, counting problems have been refined along various dimensions, including approximate counting, counting on restricted graph classes, and counting modulo a fixed number.

In this talk, we consider one of the most recent such refinements, namely parameterized counting complexity, as introduced by Flum and Grohe. Here, we ask questions of the following kind: On inputs x with a parameter k, can we solve a given problem in time f(k)*poly(n) for a function f that depends only on k? If we can, we call the problem fixed-parameter tractable (fpt). Otherwise, we try to prove its #W[1]-hardness, which is the parameterized analogue of #P-hardness, which in turn is the counting analogue of NP-hardness.

In the first part of this talk, we parameterize counting perfect matchings (PerfMatch) by structural parameters of the input graph G. For instance, PerfMatch is known to be fpt when parameterized by the genus of G. We introduce a new tractable structural parameter, namely, the size of an excluded single-crossing minor in G. We complement this by showing that PerfMatch is #W[1]-hard when parameterized by the size of an arbitrary excluded minor.

Then we turn our attention to counting general subgraphs H other than perfect matchings in a host graph G. Instead of imposing structural parameters on G, we parameterize by the size of H, giving rise to the problems #Sub(C) for fixed graph classes C: For inputs H and G, where H is from C, we wish to count H-copies in G. Here, C could be the class of matchings, cycles, paths, or any other graph class. We give a full dichotomy for these problems: Either #Sub(C) has a polynomial-time algorithm or it is #W[1]-complete. Assuming that FPT is not equal to #W[1], we can thus precisely say for which graph classes C the subgraph counting problem #Sub(C) admits polynomial-time algorithms.

Finally, we prove conditional lower bounds for counting problems under the counting exponential-time hypothesis #ETH. This hypothesis, introduced by Dell et al., asserts that the satisfying assignments to a given formula in 3-CNF cannot be counted in time 2^o(n). Building upon this, we introduce a general technique that allows to derive tight lower bounds for other counting problems, such as PerfMatch, the Tutte polynomial, the matching polynomial, and various other counting problems.

Chuong H. NGUYEN
Data-driven Approaches for Interactive Appearance Editing
(Advisor: Prof. Hans-Peter Seidel)
Mon, 22 June 2015, 9:00, building E1 4, room 0.19

This thesis proposes several techniques for interactive editing of digital content and fast rendering of virtual 3D scenes. Editing of digital content - such as images or 3D scenes- is difficult, requires artistic talent and technical expertise. To alleviate these difficulties, we exploit data-driven approaches that use the easily accessible Internet data (e. g., images, videos, materials) to develop new tools for digital content manipulation. Our proposed techniques allow casual users to achieve high-quality editing by interactively exploring the manipulations without the need to understand the underlying physical models of appearance.

First, the thesis presents a fast algorithm for realistic image synthesis of virtual 3D scenes. This serves as the core framework for a new method that allows artists to fine tune the appearance of a rendered 3D scene. Here, artists directly paint the final appearance and the system automatically solves for the material parameters that best match the desired look. Along this line, an example-based material assignment approach is proposed, where the 3D models of a virtual scene can be materialized" simply by giving a guidance source (image/video). Next, the thesis proposes shape and color subspaces of an object that are learned from a collection of exemplar images. These subspaces can be used to constrain image manipulations to valid shapes and colors, or provide suggestions for manipulations. Finally, data-driven color manifolds which contain colors of a specific context are proposed. Such color manifolds can be used to improve color picking performance, color stylization, compression or white balancing.

Adrian NEUMANN
On Efficiency and Reliability in Computer Science
(Advisor: Prof. Kurt Mehlhorn)
Fri, 19 June 2015, 15:00, building E1 4, room 0.24

Efficiency of algorithms and robustness against mistakes in their implementation or uncertainties in their input has always been of central interest in computer science. This thesis presents results for a number of problems related to this topic. Certifying algorithms enable reliable implementations by providing a certificate with their answer. A simple program can check the answers using the certificates. If the the checker accepts, the answer of the complex program is correct. We present a novel certifying algorithm for 3-edge-connectivity as well as a simplified certifying algorithm for 3-vertex-connectivity. Storing the state of computations, so called checkpointing,

also helps with reliability since we can recover from errors without having to restart the computation.

In this thesis we show how to do checkpointing with bounded memory and present several strategies to minimize the worst-case recomputation. In theory, the input for problems is accurate and well-defined. However, in practice it often contains uncertainties necessitating robust solutions. We consider a robust variant of the well-known k-median problem, where the clients are grouped into sets. We want to minimize the connection cost of the expensive group. This solution is robust against which group we actually need to serve. We show that this problem is hard to approximate, even on the line, and evaluate heuristic solutions.

Thomas BÜHLER
A flexible framework for solving constrained ratio problems in machine learning
(Advisor: Prof. Matthias Hein)
Wed, 17 June 2015, 14:00, building E1 1, room 4.07

The (constrained) optimization of a ratio of non-negative set functions is a problem appearing frequently in machine learning. As these problems are typically NP hard, the usual approach is to approximate them through convex or spectral relaxations. While these relaxations can be solved globally optimal, they are often too loose and thus produce suboptimal results. I present a flexible framework for solving such constrained fractional set programs (CFSP). The main idea is to transform the combinatorial problem into an equivalent unconstrained continuous problem. I will show that such a tight relaxation exists for every CFSP. It turns out that the tight relaxations can then be related to a certain type of nonlinear eigenproblem. I present a method to solve nonlinear eigenproblems and thus optimize the corresponding ratios of in general non-differentiable differences of convex functions. While the global optimality cannot be guaranteed, one can prove the convergence to a solution of the associated nonlinear eigenproblem. Moreover, in practice the loose spectral relaxations are outperformed by a large margin. Going over to constrained fractional set programs and the corresponding nonlinear eigenproblems leads to greater modelling flexibility, as will be demonstrated for several applications in data analysis, namely the optimization of balanced graph cuts, constrained local clustering, community detection via densest subgraphs and sparse principal component analysis.

Iliyan GEORGIEV
Path Sampling Techniques for Efficient Light Transport Simulation
(Advisor: Prof. Philipp Slusallek)
Wed, 10 June 2015, 11:00, building D3 2 (DFKI), room -2.17 (Reuse)

Reproducing the interactions between light and matter in a physically accurate way can significantly improve the realistic appearance of synthetic images, however such effects can be very computationally expensive to simulate. Pressed by strict requirements on image quality and visual realism, industrial applications have recently moved away from using legacy rasterization-based rendering solutions to fully embrace physically-based Monte Carlo methods. This dramatic shift has rekindled the interest in developing new and robust light transport simulation algorithms that can efficiently handle a wide range of scenes with complex materials and lighting -- a problem that we address in this thesis.

May

Florian DAIBER
Touching the 3rd Dimension – Interaction with Stereoscopic Data On and Above Interactive Surfaces
(Advisor: Prof. Antonio Krüger)
Mon, 4 May 2015, 14:00, building D3 2 (DFKI), room -2.17 (Reuse)

In recent years, interaction with three-dimensional (3D) data has become more and more popular. However, current 3D User Interfaces (3DUI), as for example provided by Virtual Reality (VR) systems, are very often expert systems with complex user interfaces and highinstrumentation. While stereoscopic displays allow users to perceive 3D content in an intuitive and natural way, interaction with stereoscopic data is still a challenging task, especially with objects that are displayed with different parallaxes. To overcome this interaction problem, multi-touch or depth sensing as commodity tracking technologies can be used. This thesis therefore investigates the challenges that occur when the flat world of surface computing meets the spatially complex 3D space.

The thesis contributes a number of interactive research prototypes, interaction techniques and insights from the corresponding studies. 3D interaction techniques will be designed and studied in canonical 3DUI tasks, which leads to the conclusion that significant differences arise when interacting with objects that are stereoscopically displayed at different parallaxes. The results give implications for the design of usable 3D interaction techniques that might enable VR in the living room or at public places. To sum up, this work contributes to a better understanding of stereoscopic 3D applications for end users and can be seen as a step towards a ubiquitous distribution of interactive 3D technologies and applications.

April

Mohamed ABDEL-MAKSOUD
Processor Pipelines in WCET Analysis
(Advisor: Prof. Reinhard Wilhelm)
Fri, 24 Apr 2015, 10:30, building E1 7, room 001

Due to their nature, hard real-time embedded systems (e.g. flight control systems) must be guaranteed to satisfy their time constraints under all operating conditions. The provision of such guarantee relies on safe and precise estimates of the worst-case execution time (WCET) of tasks. As the execution time depends on both the program and the architecture running it, the growing sophistication of architectures complicates the task of timing analyses. This work studies the impact of the design of the microprocessor's pipeline on the precision and efficiency of WCET analysis.

We study the influence of the design of the load-store unit (LSU) in a modern microprocessor, the PowerPC 7448, on WCET analysis. To this end, we introduce a simplified variant of the existing design of the LSU by reducing its queue sizes. The study contributes empirical evidence supporting the argument that micro-architectural innovations do not improve, and sometimes harm, a processor's worst-case timing behavior. Building on this evidence, we introduce a compiler optimization to reduce analysis time and memory consumption during the two most-computationally-demanding steps of WCET analysis. With our prototype implementation of the optimization, we observe an analysis speedup of around 635% at the cost of an increase in the WCET bound of 6%. Moreover, under a less precise yet significantly faster variant of the analysis, the WCET bound is decreased by 5% while the analysis is sped up by 350%.

Stephan SEUFERT
Algorithmic Building Blocks for Relationship Analysis over Large Graphs
(Advisor: Prof. Gerhard Weikum)
Mon, 20 Apr 2015, 10:00, building E1 4, room 024

Over the last decade, large-scale graph datasets with millions of vertices and edges have emerged in many diverse problem domains. Notable examples include online social networks, the Web graph, or knowledge graphs connecting semantically typed entities. An important problem in this setting lies in the analysis of the relationships between the contained vertices, in order to gain insights into the structure and dynamics of the modeled interactions.
In this work, we develop efficient and scalable algorithms for three important problems in relationship analysis and make the following contributions:
• We present the Ferrari index structure to quickly probe a graph for the existence of an (indirect) relationship between two designated query vertices, based on an adaptive compression of the transitive closure of the graph.
• In order to quickly assess the relationship strength for a given pair of vertices as well as computing the corresponding paths, we present the PathSketch index structure for the fast approximation of shortest paths in large graphs. Our work extends a previously proposed prototype in several ways, including efficient index construction, compact index size, and faster query processing.
• We present the Espresso algorithm for characterizing the relationship between two sets of entities in a knowledge graph. This algorithm is based on the identification of important events from the interaction history of the entities of interest. These events are subsequently expanded into coherent subgraphs, corresponding to characteristic topics describing the relationship.
We provide extensive experimental evaluations for each of the methods, demonstrating the efficiency of the individual algorithms as well as their usefulness for facilitating effective analysis of relationships in large graphs.

Pramod BHATOTIA
Incremental Parallel and Distributed Systems
(Advisor: Prof. Rodrigo Rodrigues)
Tues, 7 Apr 2015, 15:30, building E1 5, room 029

Many applications exhibit the property that small, localized changes to the input result in only minor changes to the output. Incremental computations seek to exploit this observation to speed-up successive runs of an application by tracking control and data dependencies, and re-executing only those parts of the computation that are affected by a given input change. To realize the benefits of incremental computation, researchers and practitioners are developing new systems where the programmer can provide efficient update mechanisms for changing application data. Unfortunately, most of the existing solutions are limiting because they either depart from existing programming models, or require programmers to devise an incremental update mechanism (or a dynamic algorithm) on a per-application basis.
The high-level goal of my thesis research is to enable practical, automatic, and efficient incremental computations in parallel and distributed systems. My approach neither requires a radical departure from current models of programming nor complex, application-specific dynamic algorithms.
In this talk, I will first present a high-level description of my thesis research. Thereafter, as a concrete example I will describe my recent project "iThreads". iThreads is a POSIX-compliant threading library to support parallel incremental computation targeting unmodified C/C++ multithreaded programs. The iThreads library can be used as a drop-in replacement for the pthreads library, making it trivial to obtain the benefits of incremental computation by a simple exchange of libraries linked, without even recompiling the application code. To achieve this result, we design our algorithms and implementation to operate at the compiled binary code level by leveraging operating system-specific mechanisms. Our design choice of iThreads tries to strike a balance between efficiency of incremental computation and transparency for unmodified C/C++ pthreads-based programs. Our evaluation on a multicore platform using benchmarks from the PARSEC and Phoenix suites shows significant performance improvements for the majority of applications.

March

Luis Francisco (Beta) ZILIANI
Interactive Typed Tactic Programming in the Coq Proof Assistant
(Advisor: Dr. Derek Dreyer)
Tues, 24 Mar 2015, 16:00, building E1 5, room 029

Proof assistants like Coq are now eminently effective for formalizing "research-grade" mathematics, verifying serious software systems, and, more broadly, enabling researchers to mechanize and breed confidence in their results. Nevertheless, the mechanization of substantial proofs typically requires a significant amount of manual effort. To alleviate this burden, proof assistants typically provide an additional language for tactic programming. Tactics support general-purpose scripting of automation routines, as well as fine control over the state of an interactive proof. However, for most existing tactic languages (e.g., ML, Ltac), the price to pay for this freedom is that the behavior of a tactic lacks any static specification within the base logic of the theorem prover (such as, in Coq, a type). As a result of being untyped, tactics are known to be difficult to compose, debug, and maintain.

In my thesis I present two different approaches to typed tactic programming in the context of Coq: Lemma Overloading and Mtac. Both approaches rely on the heavily on the unification algorithm of Coq, which is a complex mixture of different heuristics and has not been formally specified. Therefore, I also build and describe a new unification algorithm better suited for tactic programming in Coq. In the talk I will describe the high level achievements of my work and introduce in depth Mtac, a new programming language for tactic programming.

February

Martin SLAWSKI
Topics in learning sparse and low-rank models of non-negative data
(Advisor: Prof. Matthias Hein)
Wed, 25 Feb 2015, 10:00, building E1 1, room 407

Sparse and low-rank models are commonly used to perform dimensionality reduction, which is essential for obtaining compact and interpretable representations of high-dimensional data. In this thesis, we investigate aspects of sparse and low-rank modeling in conjunction with non-negative data or non-negativity constraints. The talk is focused on the problem of learning sparse non-negative representations. We present a statistical analysis of non-negative least squares regression (NNLS) in which we show that under certain conditions NNLS adapts to underlying sparsity similarly to methods explicitly promoting sparsity via some form of regularization (like the popular 'lasso'). At a theoretical level, this finding questions the widely held opinion that regularization is inevitable in high-dimensional inference. On the practical side, implications for problems in signal processing such as compressed sensing and spike train deconvolution are discussed.

Johannes HOFFART
Discovering and Disambiguating Named Entities in Text
(Advisor: Prof. Gerhard Weikum)
Thu, 12 Feb 2015, 10:30, building E1 5, room 0.29

Discovering entities such as people, organizations, songs, or places in natural language texts is a valuable asset for semantic search, machine translation, and information extraction. A key challenge is the ambiguity of entity names, requiring robust methods to disambiguate names to canonical entities registered in a knowledge base. Additionally, in this dynamic world, new entities are constantly emerging, and disambiguation methods need to cope with the resulting incompleteness of knowledge bases.
This dissertation develops methods to discover and disambiguate named entities, thus linking texts to knowledge bases. The first contribution is a robust disambiguation method using a graph algorithm that makes use of the coherence among entities in the input. The second contribution is a novel model to compute the coherence among entities that works especially well for lesser known entities and is applicable to newly emerging entities. The third contribution addresses the discovery of emerging entities by modeling the entities not present in the knowledge base in an explicit manner. Finally, two applications using the developed entity disambiguation methods are presented.

Rashid IBRAGIMOV
Exact and heuristic algorithms for network alignment using graph edit distance models
(Advisor: Prof. Jan Baumbach)
Thu, 5 Feb 2015, 8:30, building E1 7, room 0.01

In the thesis we study theoretical and practical questions of applying the graph edit distance (GED) model to the protein-protein interaction network alignment problem using topological information of graphs only. In order to understand what makes graph comparison using GED hard we consider one of the simplest formulations of GED with edge insertions and deletions on trees and a derived from it problem formulated as the Covering Tree with Stars problem. Further, we introduce two new problems that extend the notion of edge preservation, that is common for many classical problems on graphs, such as the graph isomorphism problem. With the input graphs being trees, we study how the complexity of the problems changes depending on the degree of relaxation of the edge preservation. For the practical application, we present fairly general and efficient heuristics for the global protein-protein interaction network alignment problem and evaluate them on real network data. Two of the three introduced techniques are connected with pairwise global network alignment. Here we find a one-to-one correspondence between the nodes of the input networks by exploiting a GED model that counts the number of deleted and inserted edges. We extend the proposed model to the case of multiple networks to minimize the sum of edit operations between every pair of the input networks.

Arnd HARTMANNS
On the Analysis of Stochastic Timed Systems
(Advisor: Prof. Holger Hermanns)
Wed, 4 Feb 2015, 16:15, building E1 7, room 0.01

The formal methods approach to develop reliable and efficient safety- or performance-critical systems is to construct mathematically well-defined models of such systems on which properties of interest, such as safety guarantees or performance requirements, can be verified automatically. In this thesis, we present ways to extend the reach of the exhaustive and statistical model checking techniques to verify reachability and reward-based properties of compositional behavioural models that support quantitative aspects such as real time and randomised decisions. In particular, we introduce two approaches that allow sound statistical model checking for the nondeterministic-randomised model of Markov decision processes, and investigate potential ways to apply similar methods to their real-time extension, probabilistic timed automata. The most general model we consider are stochastic timed automata. They allow nondeterministic choices as well delays that are nondeterministic or based on continuous probability distributions. We present the first exhaustive model checking algorithm that allows their analysis. All of these approaches have been implemented as part of the Modest Toolset, which

supports the construction and verification of models specified in the formal modelling language Modest. We conclude by applying this language and toolset to study novel distributed control strategies for photovoltaic microgenerators.