Yearbook 2014/2015



International Max Planck Research School for Computer Science

ORGANIZED BY THE MAX PLANCK INSTITUTES FOR INFORMATICS AND SOFTWARE SYSTEMS

IN COOPERATION WITH THE COMPUTER SCIENCE DEPARTMENT AT SAARLAND UNIVERSITY

International Max Planck Research School

for Computer Science



Yearbook 2014/2015



CONTENTS

ABOUT IMPRS-CS	5
Alumni: M.Sc	6
ALUMNI: PHD	31
Current Students: M.Sc	87
Current Students: PHD	88
IMPRS-CS STAFF	92
Awards	96

International Max Planck Research School for Computer Science



ABOUT IMPRS-CS

The International Max Planck Research School for Computer Science (IMPRS-CS) is a graduate program jointly run by the Max Planck Institute for Informatics (MPI-INF), the Max Planck Institute for Software Systems (MPI-SWS), and the Computer Science Department at Saarland University.

MPI-INF and MPI-SWS are among the more than 80 institutes run by the Max Planck Society. The MPIs are Germany's prime basic research facilities with world-class, foundational research in the fields of medicine, biology, chemistry, physics, technology, and the humanities. Since 1948, MPI researchers have won 17 Nobel prizes, which testifies to the quality and innovation of MPI research programs.

Educating and training junior scientists is of primary importance for the future of science, research, and innovation. The Max Planck Society, in cooperation with German universities, has launched the International Max Planck Research Schools (IMPRS) initiative.

Admitted students receive a first rate, research-oriented education in their chosen area of concentration. They enjoy close supervision by world-renowned scientists in a competitive, yet collaborative, environment – rich in interaction with other students, post-docs, and scientists. The program is fully funded.





Georgios ARVANITIDIS NATIONALITY: Greek DEPARTMENT: Machine Learning

THESIS TITLE:

Robust Principal Component Analysis Based on the Trimmed Component Wise Reconstruction Error

Abstract of Master's Thesis:

The Principal Component Analysis (PCA) is one of the most frequently used algorithms for data analysis. However, PCA is not robust to outliers and even a single outlier can already distort the principal components arbitrarily. This is the reason why robust PCA attracted much attention the last decades, which is defined as the problem of recovering the principal components of the uncontaminated data.

In this thesis we present a novel approach for robust PCA which is based on the trimmed component wise reconstruction error. Compared to the other existing methods, our approach has no free parameters, as well as does not depend on assumptions about the underlying noise distribution or the data distribution. We propose two algorithms *tcrpcaS* and *tcrpcaSt* and we show through the experiments with synthetic and real world datasets, that they perform better or they are at least competitive with current state-of-the-art and recently proposed algorithms.

International Max Planck Research School for Computer Science



Hafiz Ahmad Shahzad ASLAM

NATIONALITY: Pakistani DEPARTMENT: Software Engineering

THESIS TITLE:

Inter-Application Communication Testing of Android Applications Using Intent Fuzzing

Testing is a crucial stage in the software development process that is used to uncover bugs and potential security threats. If not conducted thoroughly, buggy software may cause erroneous, malicious and even harmful behavior. Unfortunately in most software systems, testing is either completely neglected or not thoroughly conducted. One such example is Google's popular mobile platform, Android OS, where inter-application communication is not properly tested. This is because of the difficulty which it possesses in the development overhead and the manual labour required by developers in setting up the testing environment. Consequently, the lack of Android application testing continues to cause Android users to experience erroneous behavior and sudden crashes, impacting user experience and potentially resulting in financial losses. When a caller application attempts to communicate with a potentially buggy application, the caller application will suffer functional errors or it may even potentially crash. Incidentally, the user will complain that the caller application is not providing the promised functionality, resulting in a devaluation of the application's user rating. Successive failures will no longer be considered as isolated events, potentially crippling developer credibility of the calling application.

In this thesis we present an automated tester for inter-application communication in Android applications. The approach used for testing is called Intent based Testing. Android applications are typically divided into multiple components that communicate via intents: messages passed through Android OS to coordinate operations between the different components. Intents are also used for inter-application communication, rendering them relevant for security. In this work, we designed and built a fully automated tool called IntentFuzzer, to test the stability of inter-application communication of Android applications



using intents. Firstly, it statically analyzes the application to generate intents. Next, it tests the inter-application communication by fuzzing them, that is, injecting random input values that uncover unwanted behavior. In this way, we are able to expose several new defects including potential security issues which we discuss briefly in the Evaluation section.



Darya DEDIK NATIONALITY: Ukraine DEPARTMENT: Databases and Information Systems

THESIS TITLE:

Robust Type Classification of Out of Knowledge Base Entities

ABSTRACT OF MASTER'S THESIS:

Assigning fine-grained semantic type labels to entity mentions in text has been studied intensively. Prior works on fine-grained named entity recognition and classification typically focused on unsupervised or weakly supervised learning techniques, as well as used handcrafted rules and manually defined patterns. Recent approaches exploit rich information from knowledge bases such as YAGO, Freebase, and WordNet. They mostly concentrate on linking noun-phrase mentions found in text to their corresponding Wikipedia articles or knowledge base entities. However, highly dynamic sources like news streams or social media are developing fast and provide great amount of unlinkable entities which are not prominent enough to have their own Wikipedia articles and do not have entries in knowledge bases yet. Such new entities are continuously emerging and contain great amount of information to explore. For instance, we may be interested to know which company appeared recently and offers better investment or career opportunities, which banks offer best interest rates, which blog posts advertise trending products, and so on. Constantly appearing information can be extremely useful for analysis of changes in industry, culture, sports, etc. Many natural language processing applications would benefit from the ability to detect and type such entities.

This thesis is focused on one of the least studied cases – assigning type labels to emerging entities with a very fine set of semantic types (505 classes). This task is challenging since emerging entities are not known in any knowledge base and do not have Wikipedia articles yet. Therefore, our only hope is to employ contextual cues to assign types to such entity mentions. Our method is based on HYENA, which is one of the most recent approaches in finegrained



named entity recognition and classification. HYENA is remarkably strong in typing non-emerging known entities. It exploits the YAGO knowledge base to retrieve known entity types, relations between known entities, fine-grained taxonomy and entity-type lists (gazetteers). In other words, HYENA has access to a large amount of information about the known entities, their types and mentions. In case of emerging entities, on the other hand, such information is not available. Emerging entities do not have types in the YAGO database and have no entries in gazetteers. We address this issue by proposing strong contextual features based on keyphrases, which are in some way representative for an entity, but are not available in the immediate surrounding of its mention.

Consider, for example, a web page of a scientific group and assume we would like to classify which mentions refer to researchers, students, developed software packages or projects. We could, of course, employ the standard HYENA for this task, but it would probably have difficulties classifying new students and projects, since they would be missing from a database. In contrast, if the web page contains hyper-links to the corresponding personal and project pages, we could follow these links and collect keyphrases for respective entity mentions. Therefore, we are able to extract more information and use it in classification. As shown in our experiments, this results in significant performance improvements.

We performed extensive experimental evaluation of the proposed approach in three distinct regimes of entity mentions classification: in knowledge base, simulated out of knowledge base and "real" out of knowledge base experiments, with the main focus on the latter setting. Our results show that keyphrases are promising features with high potential for fine-grained classification of the out of knowledge base entity mentions. Moreover, they can also offer a strong alternative to gazetteers in tagging known in knowledge base entities.



Mohamed GAD-ELRAB NATIONALITY: Egyptian DEPARTMENT: Databases and Information Systems

THESIS TITLE:

AIDArabic+ Named Entity Disambiguation for Arabic Text

Abstract of Master's Thesis:

Named Entity Disambiguation (NED) is the problem of mapping mentions of ambiguous names in a natural language text onto canonical entities such as people or places, registered in a knowledge base. Recent advances in this field enable semantically understanding content in different types of text. While the problem had been extensively studied for the English text, the support for other languages and, in particular, Arabic is still in its infancy. In addition, Arabic web content (e.g. in the social media) has been exponentially increasing over the last few years. Therefore, we see a great potential for endeavors that support entity-level analytics of these data. AIDArabic is the first work in the direction of using evidences from both English and Arabic Wikipedia to allow disambiguation of Arabic content to an automatically generated knowledge base from Wikipedia.

The contributions of this thesis are threefold: 1) We introduce EDRAK resource as an automatic augmentation for AIDArabic's entity catalog and disambiguation data components using information beyond manually crafted data in the Arabic Wikipedia. We build EDRAK by fusing external web resources and the output of machine translation and transliteration applied on the data extracted from the English Wikipedia. 2) We incorporate an Arabic-specific input pre-processing module into the disambiguation process to handle the complex features of Arabic text. 3) We automatically build a test corpus from other parallel English-Arabic corpus to overcome the absence of standard benchmarks for Arabic NED systems. We evaluated the data resource as well as the full pipeline using a mix of manual and automatic assessment. Our enrichment approaches in EDRAK are capable of expanding the disambiguation space from 143K entities, in the original AIDArabic, to 2.4 M entities. Moreover, the full disambiguation process is able to map 94.7% of the mentions to non-null entities with a precision of 73 %, compared to 87.2 % non-null mapping with only 69 % precision in the original AIDArabic.





Mayank GOYAL NATIONALITY: Indian DEPARTMENT: Modeling and Simulation

THESIS TITLE:

Lumping of Approximate Master Equations for Networks

Abstract of Master's Thesis:

Epidemic or information spreading phenomena in social networks are often modelled as stochastic processes, where each node is an agent (which can be in one of a few states) and edges represent the possible spreading channels. The topology of such an interaction network plays a very important role in the dynamics of the process. If the networks degree distribution is constant or Poissonian (Erdos-Renyi graphs), the Mean Field (MF) approach often approximates well the average population of agents in each state. This is not more than solving one differential equation (ODE) for each population type. However, if the underlying graph's degree distribution is closer to a heavy-tailed distribution (e.g. Power law), which is actually the case for most real world networks, the degree based mean field (DBMF) approach has to be used. This requires solving one ODE for each different degree and each population type of the network. A more accurate representation of the network is possible by including pairwise dynamical correlations (Pair Approximation). Pair Approximation (PA) approach takes into account the edges between a node and its neighbours. In PA, for a node of degree k we need to consider the probability of its neighbour being infected or susceptible. This requires solving three differential equations for each degree k and each population type.

In the current approaches, there is a problem of polynomial increase in the number of differential equations, both with the population size and the number of possible states, due to the increase of the maximum degree of the network and the number of transitions involved. In this thesis, we will attempt to ame-

14

liorate such an explosion by lumping the differential equations for several degrees, thus reducing their number and obtaining an approximation of the DBMF and PA solution. We propose four different techniques of lumping differential equations for DBMF and PA approaches. These techniques are based upon the probability distribution of the degrees in the network. Our aim is to investigate the performance and accuracy of these lumping techniques.





Ilya GRISHCHENKO NATIONALITY: Russian

THESIS TITLE:

Static Analysis of Android Applications

DEPARTMENT: Secure and Privacy-Preserving Systems

ABSTRACT OF MASTER'S THESIS:

Mobile and portable devices are machines that users carry with them everywhere, they can be seen as constant personal assistants of modern human life. Today the Android operating system for mobile devices is the most popular one and the number of users still grows: as of September 2013, 1 billion devices have been activated [Goob]. This makes the Android market attractive for developers willing to provide new functionality. As a consequence, 48 billion applications ("apps") have been installed from the Google Play store [BBC].

Apps often require user data in order to perform the intended activity. At the same time parts of this data can be treated as sensitive private information, for instance, authentication credentials for accessing the bank account. The most significant built-in security measure in Android, the permission system, provides only little control on how the app is using the supplied data.

In order to mitigate the threat mentioned above, the hidden unintended app activity, the recent research goes in three main directions: inline-reference monitoring modifies the app to make it safe according to user defined restrictions, dynamic analysis monitors the app execution in order to prevent undesired activity, and static analysis verifies the app properties from the app code prior to execution.

As we want to have provable security guarantees before we execute the app, we focus on static analysis. This thesis presents a novel static analysis technique based on Horn clause resolution. In particular, we propose the small-step concrete semantics for Android apps, we develop a new form of abstraction which

is supported by general theorem provers. Additionally, we have proved the soundness of our analysis technique.

We have developed a tool that takes the bytecode of the Android app and makes it accessible to the theorem prover. This enables the automated verification of a variety of security properties, for instance, whether a certain functionality is preceded by a particular one, for instance, whether the output of a bank transaction is secured before sending it to the bank, or on which values it operates, for instance, whether the IP-address of the bank is the only possible transaction destination.

A case study as well as a performance evaluation of our tool conclude this thesis.





Charles Darwis HARIMAN NATIONALITY: Indonesian DEPARTMENT: Databases and Information Systems

THESIS TITLE:

Part-Whole Commonsense Knowledge Harvesting from the Web

Abstract of Master's Thesis:

Commonsense knowledge about Part-Whole relations (e.g., webcam partOf notebook) is important for interpreting user input in web search and question answering, or for object detection in images. Prior work on knowledge base construction has compiled tens of thousands of Part-Whole assertions, but with substantial limitations: i) semantically different kinds of Part-Whole relations are conflated into a single generic relation, ii) the arguments of a Part-Whole assertion are merely words with ambiguous meaning, iii) the assertions are merely qualitative and lack cardinality information (e.g., a bird has two legs while a spider has eight).

This thesis presents a new method for automatically acquiring Part-Whole commonsense from Web contents at an unprecedented scale, yielding many millions of assertions, while specifically addressing the three shortcomings of prior work. Our method combines pattern-based information extraction methods with logical reasoning. We carefully distinguish different relations: physicalPartOf, memberOf, substanceOf. We consistently map the arguments of all assertions onto WordNet senses, eliminating the ambiguity of word-level assertions. We infer cardinalities for many assertions. The resulting Part-Whole commonsense knowledge base has very high quality and will be made publicly available.



Razvan Florin HULEA

NATIONALITY: Romanian DEPARTMENT: Computer Graphics

THESIS TITLE:

Compressed Vibration Modes for Deformable Objects

ABSTRACT OF MASTER'S THESIS:

In this work a new concept of compressed vibration modes for tetrahedral meshes will be introduced. These modes are in fact compressed eigenfunctions of the Hessian matrix of the deformation energy and, if assembled together, they form an orthonormal basis where each eigenfunction will have local support. The sparsity is achieved by adding a L1 regularizer to the objective function. The regularizer has shrinkage properties, such that a notable amount of the coefficients of the eigenfunctions will be set to zero, which will lead to localized results. Moreover, the support of the functions can be controlled by a parameter.

The variational problem is then solved using ADMM (Alternating Direction Method of Multipliers), a complex optimization routine. Even though ADMM was developed for convex optimization problems, the results will be trapped in local minima, since the objective function together with the constraints will lead to a non-linear optimization problem. Nevertheless, the results look intuitive for the human eye and can be achieved in only a few hundred iterations. Potential applications of this method can be adapted in areas such as interactive deformations or even clustering.





Edite KAPCARI NATIONALITY: Albanian DEPARTMENT: German Research Center for Artificial Intelligence

THESIS TITLE:

Parallel vs. Traditional Faceted Browsing: Comparative Studies and Proposed Enhancements

Abstract of Master's Thesis:

This thesis has two major purposes: (1) to investigate the strengths and limitations of the Faceted Browsing (FB) and the Parallel Faceted Browsing (PFB) facet navigation paradigms and (2) to propose a hybrid system compound of solutions to the found limitations in both paradigms. Faceted navigation is one of the most powerful mechanisms used nowadays in search interfaces. Confronted with thousands of available options, users find it difficult to make a choice that would successfully address their needs. FB assists the user in performing a simple and intuitive filtering of options but has the limitation of visualizing only one individual or set of results at a time. PFB on the other hand, circumvents this limitation by allowing the user to track, display and compare multiple search paths simultaneously in a single web-page. The main limitation of this paradigm was thought to reside in its learnability, due to the novel way the information is explored. A theoretical analysis of the types of choice that FB and PFB support helped us in understanding the potential users, in deriving some hypothesis which suggested empirical studies, and in evaluating the results. In addition to the interpretation of the empirical results obtained, we propose a set of design solutions addressing possible ways of combating the limitations that were found in these two paradigms. The results of experiments show that PFB is an important innovative type of navigation, which, if combined with other powerful mechanisms such as pivoting or critique-based recommender systems, can further exploit the benefits of the combined components and overcome some of their limitations. On the basis of these results and analyses, we conclude this work by formulating some recommendations for people who work with PFB.



Anna KHOREVA NATIONALITY: Russian DEPARTMENT: Computer Vision and Multimodal Computing

THESIS TITLE:

Video Segmentation with Graph Cuts

ABSTRACT OF MASTER'S THESIS:

Despite the recent progress, video segmentation research is currently limited by the lack of study of low-level features. The computational complexity of video data and variations of color, texture and motion over time introduce additional challenges to the task of video segmentation.

Spectral methods build the basis of many state-of-the-art image and video segmentation techniques. However, in contrast to image segmentation extension of spectral clustering to video is far less researched. Little attention has been paid to the effects of the 2-norm relaxation with different cut criteria applied to video segmentation and the tight 1-norm relaxation, which in practice yields better partitions than spectral clustering, has not been yet adopted to video processing.

In this work we contribute with an experimental analysis of low-level features used for video segmentation. Our results show that the features, integrated by local grouping cues, provide good performance with low error rate and are sufficient for obtaining high-quality segmentations.

In an effort to understand the benefits and drawbacks of recently developed spectral methods applied to video, we provide an extensive empirical comparison of the 1-norm and 2-norm relaxation techniques with different graph cut objective functions, examining their impact on segmentation performance. We show that applied to video the standard spectral method outperforms 1-spectral clustering and the optimal in the sense of balanced graph cuts solution is not always reached by relaxation techniques.



We propose to employ constrained spectral clustering to video segmentation, where must-link constraints are integrated into spectral framework via sparsification preserving all balanced cuts in the reduced graph. We present a novel methodology for discriminative learning of must-links from low-level features. The proposed method allows to reduce runtime and memory consumptions and improve the performance. The experimental results on the Berkeley motion segmentation dataset demonstrate the relevance and accuracy of our method as compared to other existing video segmentation algorithms.

International Max Planck Research School for Computer Science



Uzair MAHMOOD NATIONALITY: Pakistani DEPARTMENT: Cryptographic Systems

THESIS TITLE:

Ensuring Integrity of Recommendations in a Marketplace

Abstract of Master's Thesis:

Huge amounts of information available today have prompted the use of recommendations. These are made by recommendation systems that provide users with concise content they are expected to like. Much of these systems are centralized and a single provider makes use of system wide information to generate recommendations. These systems are very common such as the best selling and top rated products in Amazon and the top trending topics in Twitter. Marketplaces like Amazon make use of recommendations to suggest products to buyers, with buyers more inclined to purchase the products that are recommended. The marketplace scenario provides an incentive for sellers to collude with a marketplace to boost their product recommendations leading to increased sales and therefore, calls for an increased integrity check on recommendations.

We provide a protocol for a centralized marketplace e.g. Amazon that allows for verifiability of a recommendation list of top rated products that is generated using buyer feedback from valid transactions. We provide a solution even in the presence of collusion among seller(s), the marketplace and a subset of the buyers in such a scenario. Our protocol employs a novel combination of a token system based on accumulators that works in conjunction with the distributed Homomorphic MAC. Use of our protocol enables only valid buyers to leave a rating and allows verifiability of the recommendation list computation while reducing storage requirements for the sellers. We also provide an analysis of our protocol and the attacks it is resilient against.





Pedro Eduardo MERCADO LOPEZ

NATIONALITY: Mexican
DEPARTMENT: Computer Vision and Multimodal Computing

THESIS TITLE:

Clustering and Community Detection in Signed Networks

ABSTRACT OF MASTER'S THESIS:

We consider the problem of clustering and community detection in signed networks where one has both positive and negative relationships. For clustering in signed networks we analyze the signed ratio cut and introduce the signed Cheeger cut, which aim to find balanced cuts in signed networks. In particular, for the signed Cheeger cut we present the first analysis based on the Stochastic block model, giving sufficient conditions so that ground truth clusters are exactly recovered in expectation. Moreover, we show that under certain conditions, the state of the art does not recover ground truth clusters in expectation while the signed Cheeger cut does.

Further, we address the problem of community detection by introducing a novel extension of the maximum densest subgraph problem with the goal of finding communities with dense positive and (almost) no negative relationships.

In contrast to previous work, which claims that there is no clustering structure in online signed networks, our new methods are the first to find both clustering and community structure in all considered networks.

Mohamed OMRAN

NATIONALITY: Egyptian DEPARTMENT: Computer Vision and Multimodal Computing

THESIS TITLE:

Pedestrian Detection Meets Stuff

Abstract of Master's Thesis:

Object detection and semantic labelling are commonly treated as separate problems. We argue that such a separation is not justified, and present a unified architecture for semantic labelling and pedestrian detection. Furthermore, by feeding the semantic labelling information into the pedestrian detector, we are able to further improve detection quality, reaching top results on the challenging Caltech-USA dataset over a state of the art baseline.



Panagiotis MANDROS NATIONALITY: Greek DEPARTMENT: Exploratory Data Analysis

THESIS TITLE:

Information Theoretic Supervised Feature Selection for Continuous Data

Abstract of Master's Thesis:

Feature selection is a dimensionality reduction process that judiciously selects a subset of features from which to build a prediction model. In particular, it alleviates computational and high dimensionality problems by reducing the number of dimensions one has to consider. Recognizing the importance of feature selection to data analysis, in this thesis we aim to contribute with an Exploratory Data Analysis approach – the discipline of working directly with empirical data without making unnecessary assumptions (e.g. assumptions on data distributions).

With this goal in mind, we propose a parameter and assumption free supervised information-theoretic feature selection method for data with continuous attributes and discrete/nominal class label. The novelty of our method lies in working directly with continuous data by employing Cumulative Entropy, a new entropy measure designed specifically for continuous random variables. While our objective is to work directly with continuous data, we will show that for some computations we need to perform discretization — a process converting continuous data to discrete one. In such cases, we optimally, non-parametrically, and efficiently discretize our data with dynamic programming, completely avoiding naive discretizations commonly used in such cases. Our method aims at identifying higher order dependencies in the data, differentiating itself from methods that search only for pairwise. To evaluate the benefits of our method, we perform extensive experiments on both synthetic and real-world data sets. The results show that our method compares favorably to state-of-theart information-theoretic feature selection techniques.



Quynh Ngoc NGUYEN NATIONALITY: Vietnamese DEPARTMENT: Machine Learning

THESIS TITLE:

A Tensor Block Coordinate Ascent Framework for Hyper Graph Matching

Abstract of Master's Thesis:

Graph matching is a core problem of computer vision and pattern recognition. The corresponding quadratic assignment problem is NP-hard and, in the past, several methods have been developed for the second-order case. In recent years also third order approaches became popular as they allow for better integration of geometric information. For most of these higher-order algorithms no theoretical guarantees are known. In this thesis we propose a general framework for tensor block coordinate ascent methods for third-order graph matching. We first formulate the problem as maximization of a third-order polynomial function with respect to assignment constraints. Then, we prove the equivalence between the optimization of the original function and the optimization of its associated multilinear form over any constraint set by extending one result from the matrix case. Based on this, we propose two new algorithms which both come along with the guarantee of monotonic ascent in the objective on the set of discrete assignment matrices. In the experiments, we show that our algorithms outperform previous work both in terms of achieving better objective values and in terms of matching accuracy. This holds in particular for very challenging settings where one has a high number of outliers and other form of noise.



John Samy Selim TIAB

NATIONALITY: Egyptian DEPARTMENT: Computer Graphics

THESIS TITLE:

Design and Evaluation Techniques for Cuttable Multi-Touch Sensor Sheets

Abstract of Master's Thesis:

Touch interaction is one of the dominant interaction techniques used nowadays. However, touch interactive devices are fixed in shape and size. These devices do not allow users to customize their shapes according to their needs. On the other hand, we are used to tailor various materials to our needs using direct customization method like cutting. Carpenters cut wood, children have fun cutting paper and tailors customize fabrics to fit their customers. Therefore, why do not we use cutting as an ad hoc method for shape customization of touch interactive devices? The problem lies in the design of current devices and electronic components. They are fabricated using rigid and expensive material that does not tolerate cutting. Moreover, the internal circuitry design of these components is not robust to cuts. Printed electronics provides the material affordance needed to produce flexible, cheap and thin electronic components. These components are printed on cuttable substrates like paper and thin films. However, they use the traditional circuitry design, which is not robust against cutting. Therefore, we contribute novel resilient circuitry designs inspired form nature and network topologies to increase the robustness of multitouch sensor sheets against cuts. Cuttable multi-touch sensor sheets enable a wide variety of ubiquitous and rapid prototyping applications. Users are able to directly customize the sheets to different shapes and sizes in order to augment physical objects with touch interactivity or prototype touch interactive artifacts. In addition, we contribute a simulation framework that evaluates the robustness of the sensor sheets against cuts and compares the performance of the various designs proposed.



Mónica Paola VIDRIALES ESCOBAR

NATIONALITY: Egyptian DEPARTMENT: Telecommunications

THESIS TITLE:

Integration of Direct- and Feature-Based Methods for Correspondences Refinement in Structure-from-Motion

ABSTRACT OF MASTER'S THESIS:

Accurate feature matching across multiple frames is a key to achieve a highquality Structure from Motion (SfM) estimation. Feature matching is typically performed using feature-based or image-based methods. Feature-based methods exhibit a high robustness against image variations but may generate poorly localized features across distinct views and are prone to jitter. On the other hand, image-based techniques provide highly accurate feature localization but tend to drift along image edges and can easily lose track of the scene due to camera motion.

In this work we investigate both feature-based and image-based methods and analyze their strengths and weaknesses. We conclude that both types of methods are highly complementary and thus can be successfully combined in order to improve feature detection and feature matching. To that end, we employ Scale Invariant Feature Transform (SIFT) [Low99] for feature detection and Inverse Compositional Algorithm (ICA) [BMO4] to refine the detected features. In particular, we use ICA to register image patches defined around corresponding SIFT features. We propose to search for an optimal image patch size and provide a homography as initial transformation to ICA. Furthermore, we decrease the computational complexity of the SfM pipeline by clustering corresponding 2D and 3D point sets. This improvement also results into a more homogenous features distribution. Finally, we incorporate a registration tree structure designed to order the image sequence prior to image registration, thus reducing the amount of transformations between images.



We evaluate our method on Hannover Table Sequence data set. Experimental results show that the angular reprojection error is reduced by using the features refined by the proposed method yielding a considerable improvement in the 3D scene reconstruction. Our findings also indicate that the proposed method allows to significantly reduce the computational complexity of SfM pipeline leading to a faster convergence.

International Max Planck Research School for Computer Science



Tetiana ZINCHENKO NATIONALITY: Ukraine DEPARTMENT: Databases and Information Systems

THESIS TITLE:

Redescription Mining Over Non-Binary Data Sets Using Decision Trees

ABSTRACT OF MASTER'S THESIS:

Scientific data mining is aimed to extract useful information from huge data sets with the help of computational efforts. Recently, scientists encounter an overload of data which describe domain entities from different sides. Many of them provide alternative means to organize information. And every alternative data set offers a different perspective onto the studied problem.

Redescription mining is tool with a goal of finding various descriptions of the same objects, i.e. giving information on entity from different perspectives. It is a tool for knowledge discovery which helps uniformly reason across data of diverse origin and integrates numerous forms of characterizing data sets.

Redescription mining has important applications. Mainly, redescriptions are useful in biology (e.g. to find bio niches for species), bioinformatics (e.g. dependencies in genes can assist in analysis of diseases) and sociology (e.g. exploration of statistical and political data), etc.

We initiate redescription mining with data set consisting of 2 arrays with Boolean and/or real-valued attributes. In redescription mining we are looking for such queries which would describe nearly the same objects from both given arrays.

Among all redescription mining algorithms there exist approaches which exploits alternating decision tree induction. Only Boolean variables were involved there so far. In this thesis we extend these approaches to non-Boolean data and adopt two methods which allow redescription mining over non-binary data sets.





Dr. Mohamed ABDEL MAKSOUD

NATIONALITY: Egyptian DEPARTMENT: Programming Languages and Compiler Construction

DISSERTATION TITLE:

Processor Pipelines in WCET Analysis

ABSTRACT OF PHD THESIS:

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 microarchitectural 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-computationallydemanding 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%.





Dr. Fidaa ABED NATIONALITY: Palestinian DEPARTMENT: Algorithms & Complexity

DISSERTATION TITLE:

Coordinating Selfish Players in Scheduling Games

ABSTRACT OF PHD THESIS:

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, 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 $R \mid \mid \Sigma \omega_i C_i$ or the jump neighborhood can be found in polynomial time and is within a factor of 2.618 of the optimal solution.



Dr. Foteini ALVANAKI NATIONALITY: Greek DEPARTMENT: Databases and Information Systems

DISSERTATION TITLE:

Mining Interesting Events on Large and Dynamic Data

ABSTRACT OF PHD THESIS:

Nowadays, almost every human interaction produces some form of data. These data are available either to every user, e.g. images uploaded on Flickr or to users with specific privileges, e.g. transactions in a bank. The huge amount of these produced data can easily overwhelm humans that try to make sense out of it. The need for methods that will analyse the content of the produced data, identify emerging topics in it and present the topics to the users has emerged. In this work, we focus on emerging topics identification over large and dynamic data. More specifically, we analyse two types of data: data published in social networks like Twitter, Flickr etc. and structured data stored in relational databases that are updated through continuous insertion queries.

In social networks, users post text, images or videos and annotate each of them with a set of tags describing its content. We define sets of co-occurring tags to represent topics and track the correlations of co-occurring tags over time. We split the tags to multiple nodes and make each node responsible of computing the correlations of its assigned tags. We implemented our approach in Storm, a distributed processing engine, and conducted a user study to estimate the quality of our results.

In structured data stored in relational databases, top-k group-by queries are defined and an emerging topic is considered to be a change in the top-k results. We maintain the top-k result sets in the presence of updates minimising the interaction with the underlying database. We implemented and experimentally tested our approach.





Dr. Yassen ASSENOV NATIONALITY: Bulgarian DEPARTMENT: Computational Biology and Applied Algorithms

DISSERTATION TITLE:

Identification and Prioritization of Genomic Loci with Disease-Specific Methylation

Abstract of PhD Thesis:

Epigenetic systems are an indispensable mechanism in development, they respond to environmental stimuli and are dysregulated in cancer and other diseases. DNA methylation is the best characterized and extensively studied epigenetic mark to date. In the past years, a number of assays have been designed to measure DNA methylation levels genome-wide. This thesis introduces computational techniques for handling DNA methylation data from microarrayand enrichment-based methods. It focuses on disease-oriented studies and addresses the questions of quality control and normalization, inter-and intragroup variability, identification of differentially methylated loci, prioritization of biomarker candidates and prediction of cancer type and other phenotypes.

The presented statistical approaches and heuristics facilitated important discoveries with clinical applications. We showed that neurological and autoimmune disorders can be characterized by their distinct methylation profiles. We observed a strong tissue-specific signal in the methylation profiles of healthy and cancer samples. We were able to accurately predict tumor type of origin ofmetastatic samples. We showed that neither adeno carcinoma, nor squamous cell carcinoma can be separated into two distinct subtypes with a characteristic global methylation profile. In colon cancer, we identified differentially methylated regions with a potential to be used as biomarkers for predicting microsatellite instability.



Dr. Bastian BEGGEL NATIONALITY: German DEPARTMENT: Computational Biology and Applied Algorithms

DISSERTATION TITLE:

Determining and Utilizing the Quasispecies of the Hepatitis B Virus in Clinical Applications

ABSTRACT OF PHD THESIS:

Chronic hepatitis B caused by infection with the hepatitis B virus (HBV) affects about 240 million people worldwide and is one of the major causes of severe liver cirrhosis and liver cancer. Hepatitis B treatment options have improved dramatically in the last decade. Effective direct-acting antiviral drugs, so-called nucleos(t)ide analogs, and one effective immunomodulatory drug (pegylated interferon a-2a) are available presently. Current challenges for treating HBV involve the careful selection of patients who require therapy and the thoughtful choice of the treatment option tailored to each patient individually. Personalized medicine aims to optimize treatment decisions based on the analysis of host factors and virus characteristics.

The population of viruses within a host is called the viral quasispecies. This thesis provides statistical methods to infer relevant information about the viral quasispecies of HBV to support treatment decisions. We introduce a new genotyping methodology to identify dual infections, which can help to quantify the risk of interferon therapy failure. We present a method to infer short-range linkage information from Sanger sequencing chromatograms, a method to support treatment adjustment after the development of resistance to nucleos(t)ide analogs. Additionally, we provide the first full-genome analysis of the G-to-A hypermutation patterns of the HBV genome. Hypermutated viral genomes form a subpopulation of the quasispecies caused by proteins of the human innate immune system editing the genome of exogenous viral agents. We show that hypermutation is associated with the natural progression of hepatitis B, but does not correlate with treatment response to interferon.




Dr. Hagen BLANKENBURG

NATIONALITY: German DEPARTMENT: Computational Biology and Applied Algorithms

DISSERTATION TITLE:

Computational Methods for Integrating and Analyzing Human Systems Biology Data

ABSTRACT OF PHD THESIS:

The combination of heterogeneous biological datasets is a key requirement for modern molecular systems biology. Of particular importance for our understanding of complex biological systems like the human cell are data about the interactions of proteins with other molecules. In this thesis, we develop and apply methods to improve the availability and the quality of such interaction data. We also demonstrate how these data can be used in interdisciplinary studies to discover new biological results.

First, we develop technical systems for the instant integration of interaction data that are stored and maintained in separate online repositories. Second, we implement a computational framework for the application of multiple scoring algorithms to qualitatively assess different aspects of interaction data. Our methods are based on distributed client-server systems, ensuring that the services can be updated continuously. This promotes equal access to interaction data and allows researchers to expand the client-server systems with their own service.

Third, we focus our application studies on integrative network-based analyses of human host factors for viral infections. Our applications provide new biological insights into the life cycle of the hepatitis C virus and identify new potential candidates for antiviral drug therapy.



Dr. Karl BRINGMANN NATIONALITY: German DEPARTMENT: Algorithms and Complexity

DISSERTATION TITLE:

Sampling from Discrete Distributions and Computing Fréchet Distances

ABSTRACT OF PHD THESIS:

In the first part of this dissertation, we study the fundamental problem of sampling from a discrete probability distribution. Specifically, given non-negative numbers $p_1,...,p_n$ the task is to draw *i* with probability proportional to p_i . We extend the classic solution to this problem, Walker's alias method, in various directions:

- 1. We improve upon its space requirements by presenting optimal succinct sampling data structures.
- We present improved trade-offs between preprocessing and query time for sorted inputs, and generalize this from proportional sampling to sampling subsets.
- 3. For Bernoulli, geometric, and binomial random variates we present optimal sampling algorithms on a bounded precision machine.
- 4. As an application, we speed up sampling of internal diffusion limited aggregation.

The second part of this dissertation belongs to the area of computational geometry and deals with algorithms for the Fréchet distance, which is a popular measure of similarity of two curves and can be computed in quadratic time (ignoring logarithmic factors). We provide the first conditional lower bound for this problem: No polynomial factor improvement over the quadratic running time is possible unless the Strong Exponential Time Hypothesis fails. Our various extensions of this main result include conditional lower bounds under realistic input assumptions, which do not match the known algorithms. We close this gap by presenting an improved approximation algorithm for the Fréchet distance.





Dr. Matthias DIETZEN NATIONALITY: German DEPARTMENT: Computational Biology and Applied Algorithms

DISSERTATION TITLE:

Modeling Protein Interactions in Protein Binding Sites and Oligomeric Protein Complexes

ABSTRACT OF PHD THESIS:

Three-dimensional structures of protein-ligand and protein-protein complexes can provide key insights into biochemical processes within living cells, yet, their experimental determination is often expensive, time-consuming, or can fail due to the heterogeneity in the complex composition and thus the binding affinities of different components. A computational prediction of these structures can overcome these problems in certain cases and is thus highly demanded in many areas of research.

In this work, we address two questions: first, can one predict conformational changes of the protein backbone upon ligand binding, using the energetically most favorable motions obtained from normal mode analysis of elastic network models, and second, can one computationally assemble large protein complexes, using the structures and stoichiometries of their monomers and the approximate interaction geometries.

For the first problem, using a diverse set of 433 pairs of bound and unbound protein conformations, we could show that the benefit from such motions is small: modeling ligand-induced conformational changes using normal modes is rather ineffective. To solve the second problem, we have developed a novel scoring function and an efficient algorithm for iterative complex assembly based on pairwise dockings, 3 D-MOSAIC, that, on a diverse benchmark set of 308 complexes, can accurately and efficiently assemble protein complexes of up to 60 monomers and 15 protein types.



Dr. Rayna DIMITROVA NATIONALITY: Bulgarian DEPARTMENT: Software Systems

DISSERTATION TITLE:

Synthesis and Control of Infinite-State Systems with Partial Observability

Abstract of PhD Thesis:

Complex computer systems play an important role in every part of everyday life and their correctness is often vital to human safety. In light of the recent advances in the area of formal methods and the increasing availability and maturity of tools and techniques, the use of verification techniques to show that a system satisfies a specified property is about to become an integral part of the development process. To minimize the development costs, formal methods must be applied as early as possible, before the entire system is fully developed, or even at the stage when only its specification is available. The goal of synthesis is to automatically construct an implementation guaranteed to fulfill the provided specification, and, if no implementation exists, to report that the given requirements cannot be realized. When synthesizing an individual component within a system and its external environment, the synthesis procedure must take into account the component's interface and deliver implementations that comply with it. For example, what a component can observe about its environment may be restricted by imprecise sensors or inaccessible communication channels. In addition, sufficiently precise models of a component's environment are typically infinite-state, for example due to modeling real time or unbounded communication buffers. This thesis presents novel synthesis methods that respect the given interface limitations of the synthesized system components and are applicable to infinite-state models.

The studied computational model is that of infinite-state two-player games under incomplete information. The contributions are structured into three parts,



corresponding to a classification of such games according to the interface between the synthesized component and its environment. In the first part, we obtain decidability results for a class of game structures where the player corresponding to the synthesized component has a given finite set of possible observations and a finite set of possible actions. A prominent type of systems for which the interface of a component naturally defines a finite set of observations are Lossy Channel Systems. We provide symbolic game solving and strategy synthesis algorithms for lossy channel games under incomplete information with safety and reachability winning conditions. Our second contribution is a counterexample-guided abstraction refinement scheme for solving infinite-state under incomplete information in which the actions available to the component are still finitely many, but no finite set of possible observations is given. This situation is common, for example, in the synthesis of mutex protocols or robot controllers. In this setting, the observations correspond to observation predicates, which are logical formulas, and their computation is an integral part of our synthesis procedure. The resulting game solving method is applicable to games that are out of the scope of other available techniques. Last we study systems in which, in addition to the possibly infinite set of observation predicates, the component can choose between infinitely many possible actions. Timed games under incomplete information are a fundamental class of games for which this is the case. We extend the abstraction-refinement procedure to develop the first systematic method for the synthesis of observation predicates for timed control. Automatically refining the set of candidate observations based on counterexamples demonstrates better potential than brute-force enumeration of observation sets, in particular for systems where fine granularity of the observations is necessary.



Dr. Maximilian DYLLA NATIONALITY: German DEPARTMENT: Databases and Information Systems

DISSERTATION TITLE:

Efficient Querying and Learning in Probabilistic and Temporal Databases

ABSTRACT OF PHD THESIS:

Probabilistic databases store, query, and manage large amounts of uncertain information. This thesis advances the state-of-the-art in probabilistic databases in three different ways:

- We present a closed and complete data model for temporal probabilistic databases and analyze its complexity. Queries are posed via temporal deduction rules which induce lineage formulas capturing both time and uncertainty.
- 2. We devise a methodology for computing the top-k most probable query answers. It is based on first-order lineage formulas representing sets of answer candidates. Theoretically derived probability bounds on these formulas enable pruning low-probability answers.
- 3. We introduce the problem of learning tuple probabilities which allows updating and cleaning of probabilistic databases. We study its complexity, characterize its solutions, cast it into an optimization problem, and devise an approximation algorithm based on stochastic gradient descent.

All of the above contributions support consistency constraints and are evaluated experimentally.





Dr. Ahmed ELHAYEK NATIONALITY: Palestinian DEPARTMENT: Computer Graphics

DISSERTATION TITLE:

Marker-less Motion Capture in General Scenes with Sparse Multi-Camera Setups

Abstract of PhD Thesis:

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 motioncapture 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

44

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.



Dr. Lars FEUERBACH

NATIONALITY: German
DEPARTMENT: Computational Biology and Applied Algorithms

DISSERTATION TITLE:

Evolutionary Epigenomics – Identifying Functional Genome Elements by Epigenetic Footprints in the DNA

ABSTRACT OF PHD THESIS:

Over the last decade, advances in genome sequencing have substantially increased the amount of genomic DNA sequences available. While these rich resources have improved our understanding of genome function, research of the epigenome as a transient but heritable memory system of the cell has only profited from this development indirectly. Although epigenetic information in the form of DNA methylation is not directly encoded in the genomic nucleotide sequence, it increases the mutation rate of cytosine-guanine dinucleotides by the CpG decay effect, and thus leaves epigenetic footprints in the DNA. This thesis proposes four approaches to facilitate this information for research. For largely uncharacterized genomes, CgiHunter presents an exhaustive algorithm for an unbiased DNA sequence-based annotation of CpG islands as regions that are protected from CpG decay. For species with well characterized point mutation frequencies, EqiScore identifies regions that evolve under distinct DNA methylation levels. Furthermore, the derived equilibrium distributions for methylated and unmethylated genome regions predict the evolutionary robustness of transcription factor binding site motifs against the CpG decay effect. The AluJudge annotation and underlying L-score provide a method to identify putative active copies of CpG-rich transposable elements within genomes. Additionally, epigenetic footprints in these sequences are applied to predict the germline epigenome of their loci. Moreover, AluJudge provides support for the targeted removal of epigenetically silenced repeat copies from CpG island annotations, which are subjected to a methylation-induced erosion process. Finally, the FFK approach enables the prediction of the germline methylome for homologous genome loci.

In a number of case studies on the human genome, I demonstrate how this evolutionary epigenomics toolkit can be applied to enhance the epigenomic characterization of the large quantity of currently sequenced vertebrate genomes. Furthermore, these studies show how to improve the identification of novel epigenetic functional genome regions in already well characterized species. Finally, the toolkit opens new avenues for computer-based research of the evolution of genome-wide DNA methylation.

Dr. Arnaud Luc FIETZKE

NATIONALITY: German DEPARTMENT: Automation of Logic

DISSERTATION TITLE:

Labelled Superposition

ABSTRACT OF PHD THESIS:

The work presented in this thesis consists of two parts: The first part presents a formalization of the splitting rule for case analysis in superposition and a proof of its correctness, as well as the integration into splitting of a novel approach to lemma learning, which allows the derivation of non-ground lemmas. The second part deals with the application of hierarchic superposition, and in particular superposition modulo linear arithmetic SUP(LA), to the verification of timed and probabilistic timed systems. It contains the following three main contributions: Firstly, a SUP(LA) decision procedure for reachability in timed automata, which is among the first decision procedures for free firstorder logic modulo linear arithmetic; secondly, an automatic induction rule for SUP(LA) based on loop detection, whose application allows a generalization of the decidability result for timed automata to timed automata with unbounded integer variables; and thirdly, a formalism for modelling probabilistic timed systems with first-order background theories, as well as a novel approach for the analysis of such systems based on a reduction to model checking using SUP(LA) proof enumeration.



Dr. Iliyan GEORGIEV

NATIONALITY: Bulgarian DEPARTMENT: Computer Graphics

DISSERTATION TITLE:

Path Sampling Techniques for Efficient Light Transport Simulation

ABSTRACT OF PHD THESIS:

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.

State-of-the-art Monte Carlo methods solve the global illumination problem by sampling random light transport paths in the scene via ray tracing. We analyze the efficiency of these methods, devise new path sampling techniques for rendering surface and volumetric light scattering, and develop novel means of leveraging illumination coherence via path reuse. This results in several practical rendering algorithms that produce images with less noise and remain more resilient to variations in the scene configuration than existing methods. The improved efficiency of these algorithms comes from the use of new and diverse sampling techniques, each specialized for handling a different set of lighting effects. Their robustness is due to the adaptive combination of these techniques in a way that preserves their individual strengths.



Dr. Sebastian GERLING NATIONALITY: German DEPARTMENT: Software Systems

DISSERTATION TITLE:

Plugging in Trust and Privacy: Three Systems to Improve Widely Used Ecosystems

ABSTRACT OF PHD THESIS:

The era of touch-enabled mobile devices has fundamentally changed our communication habits. Their high usability and unlimited data plans provide the means to communicate any place, any time and lead people to publish more and more (sensitive) information. Moreover, the success of mobile devices also led to the introduction of new functionality that crucially relies on sensitive data (e.g., location-based services). With our today's mobile devices, the Internet has become the prime source for information (e.g., news) and people need to rely on the correctness of information provided on the Internet. However, most of the involved systems are neither prepared to provide robust privacy guarantees for the users, nor do they provide users with the means to verify and trust in delivered content.

This dissertation introduces three novel trust and privacy mechanisms that overcome the current situation by improving widely used ecosystems. With WebTrust we introduce a robust authenticity and integrity framework that provides users with the means to verify both the correctness and authorship of data transmitted via HTTP. X-pire! and X-pire 2.0 offer a digital expiration date for images in social networks to enforce post-publication privacy. AppGuard enables the enforcement of fine-grained privacy policies on third-party applications in Android to protect the users privacy.





Dr. Willem HAGEMANN

NATIONALITY: German DEPARTMENT: Automation of Logic

DISSERTATION TITLE:

Symbolic Orthogonal Projections: A New Polyhedral Representation for Reachability Analysis of Hybrid Systems

Abstract of PhD Thesis:

The thesis deals with reachability analysis of linear hybrid systems. Special importance is given to the treatment of the necessary geometrical operations.

In the first part, we introduce a new representation class for convex polyhedra, the symbolic orthogonal projections (sops). A sop encodes a polyhedron as an orthogonal projection of a higher-dimensional polyhedron. This representation is treated purely symbolically, in the sense that the actual computation of the projection is avoided. We show that fundamental geometrical operations, like affine transformations, intersections, Minkowski sums, and convex hulls, can be performed by 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, we investigate the application of sops in reachability analysis of hybrid systems. It turns out that sops are well-suited for the discrete computations. Thereupon, we demonstrate also the applicability of sops for the continuous computation of reachability analysis. By a subtle parallel computation of a precise sop-based representation and a simplified representation, we tackle the problem of monotonic growing sizes of the sops. Additionally, we present experimental results which also show that sops allow an efficient and accurate computation of the reachable states.



Dr. Konstantin HALACHEV NATIONALITY: Bulgarian

DEPARTMENT: Computational Biology and Applied Algorithmics

DISSERTATION TITLE:

Exploratory Visualizations and Statistical Analysis of Large, Heterogeneous Epigenetic Datasets

ABSTRACT OF PHD THESIS:

Epigenetic marks, such as DNA methylation and histone modifications, are important regulatory mechanisms that allow a single genomic sequence to give rise to a complex multicellular organism. When studying mechanisms of epigenetic regulation, the analyses depend on the experimental technologies and the available data. Recent advancements in sequencing technologies allow for the efficient extraction of genome-wide maps of epigenetic marks. A number of large-scale mapping projects, such as ENCODE and IHEC, intensively produce data for different tissues and cell cultures. The increasing quantity of data highlights a major bottleneck in bioinformatic research, namely the lack of bioinformatic tools for analyzing these data. To date, there are bioinformatics tools for detailed (mostly visual) inspection of single genomic loci, allowing biologists to focus research on regions of interest. Also, efficient tools for manipulation and analysis of the data have been published, but often they require computer science abilities. Furthermore, the available tools provide solutions to only already well formulated biological questions. What is missing, in our opinion, are tools (or pipelines of tools) to explore the data interactively, in a process that would facilitate a trained biologist to recognize interesting aspects and pursue them further until concrete hypotheses are formulated. A possible solution stems from the best practices in the fields of information retrieval and exploratory search. In this thesis, I propose EpiExplorer, a paradigm for integration of state-of-the-art information retrieval methods and indexing structures, applied to offer instant interactive exploration of large epigenetic datasets. The algorithms we use are developed for semi-structured text data, but we apply them on bioinformatic data through clever textual map-



ping of biological properties. We demonstrate the power of EpiExplorer in a series of studies that address interesting biological problems. We also present in this manuscript EpiGRAPH, a bioinformatic software that we developed with colleagues. EpiGRAPH helps identify and model significant biological associations among epigenetic and genetic properties for sets of regions. Using EpiExplorer and EpiGRAPH, independently or in a pipeline, provides the bioinformatic community with access to large databases of annotations, allows for exploratory visualizations or statistical analysis and facilitates reproduction and sharing of results.



Dr. Johannes HOFFART

NATIONALITY: German DEPARTMENT: Databases and Information Systems

DISSERTATION TITLE:

Discovering and Disambiguating Named Entities in Text

ABSTRACT OF PHD THESIS:

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 Prof. Dr. Hans Uszkoreit.





Dr. Rashid IBRAGIMOV NATIONALITY: Uzbek DEPARTMENT: Computational Systems Biology

DISSERTATION TITLE:

Exact and Heuristic Algorithms for Network Alignment Using Graph Edit Distance Models

Abstract of PhD Thesis:

In the thesis we aim to 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 Part II we explore some theoretical aspects of the model formulated as three different problems; Part III presents three heuristics for the PPI network alignment problem based on a GED model that counts the number of deleted and inserted edges. In the following we summarize the major contribution of the work.

Covering Tree with Stars (CTS).

In Chapter 3 we study the complexity border of one of the simplest formulations of the Graph Edit Distance problem – Tree Edit Distance with edge Insertions and Deletions (TED-ID). Here, we consider a special case of the problem and reformulate it as CTS. CTS asks if, given a tree and a set of stars, we can connect the input stars by adding edges between them such that the resulting tree is isomorphic to the input tree. The main result of Chapter 3 is the proof that CTS is NP-hard. From the proof it also follows that TED-ID is also NP-hard, even if the diameter of the input trees is bounded by 10. We also show that for CTS with the size of the largest star being bounded by a constant k the problem becomes polynomial-time solvable.

Neighborhood-Preserving Mapping (NPM) on trees.

In Chapter 4 we introduce a new generalization of the graph isomorphism problem, called NPM. The input of NPM consists of two graphs and three integers l, d and k. The problem asks if the first graph without at most k vertices, called isolation vertices, can be mapped to the second graph. Using parameters l and d, for each mapped vertex v in the first graph we specify which neighbors of v (vertices within distance l from v) have to be mapped close enough (within distance d) to the image of v in the second graph. In NPM, the graphs with smaller size of the isolation set are thought to be more similar as less edit operations are required. We study NPM on trees and as the major result of the chapter we present a dichotomy of classical complexity of NPM on trees with respect to different values of l, d and k. Interestingly, NPM on trees is NP-hard even if the mapping forces the first tree to "squeeze" into the second, i.e. for d > l and $l \ge 3$. Despite the fact that all NP-hardness proofs in this thesis are based on reductions from the same problem (3-Partition), the key of every proof is the understanding of the structure of the trees to construct depending on the input parameters.

Compactness-Preserving Mapping (CPM) on trees.

In Chapter 5 we consider another generalization of the graph isomorphism problem, called CPM. The input of CPM consists of two graphs and two integers l and d. The meaning of the parameter l is the same as in NPM. The parameter d corresponds to the total "error" that is allowed when close neighbors of a vertex v in the first graph are mapped to the second graph: Given the sum L v of distances between v and its close neighbors, and the sum L O v of distances between the image of v and the images of the close neighbors of v, CPM asks if there is a one-to-one mapping between the graphs such that for every v, L O v – L v ≤ d. From one side CPM constrains NPM formulation with k = O and relaxes it from another side: In contrast to NPM it is possible that some vertices are mapped at distance greater than d, but most of the vertices have to be closer to each other. Two graphs are thought to be more similar if there is a mapping as solution for CPM with a smaller value of d. In Chapter 5 we study CPM on trees and present a dichotomy of classical complexity of CPM with respect to different values of l and d.

Network Alignment with Bee Colony Optimization.

From a practical side, despite the fact that computing GED is NP-hard, efficient heuristics are required that are able to approximate an optimal solution. In Chapter 8 we describe the first heuristic presented in the thesis, called NABEECO. Approximating GED, NABEECO adopts an artificial bee colony strategy to traverse the search space and refine its solutions. We demonstrate the performance of NABEECO on a set of real protein-protein interaction data. Despite the relatively good performance of the strategy, its current implementation is not efficient enough in escaping local minima. For example it failed to pass the self-alignment test, resulting in about 85% of the optimal solution on most input graphs.

Network Alignment with Evolutionary Algorithm.

In Chapter 9 we present a novel algorithm, called GEDEVO, for the pairwise network alignment problem. GEDEVO aims to optimize GED by exploiting an evolutionary strategy. The main advantages of the approach are its simplicity and flexibility: We perform elementary evolutionary operations on a map-



ping, which are guided by any relevant data that can be incorporated thanks to the GED model. Among its disadvantages are high running times and memory consumption. On a set of real data sets we demonstrate that our approach is comparable to and often outperforms the current tools for Network Alignment using topological information only. We show that, in contrast to other tools, our approach is also able to find a perfect alignment on a self-alignment test, which should be a requirement for any network alignment tool.

Multiple-Network Alignment with Evolutionary Algorithm.

In Chapter 10 we present a heuristic, called GEDEVO-M, for the multiple network alignment problem. Here, we define the total graph edit distance on multiple networks as an optimization criterion and apply the evolutionary strategy presented in Chapter 9. Correspondingly, we extend the evolutionary operators of GEDEVO and evaluate the approach in two settings by aligning two sets of eukaryotic and bacterial PPI networks using topological information only. We also show that GEDEVO-M is able to pass a self-alignment test where it perfectly aligns several copies of a network. However, the running time here depends on the number of input networks. Since with a growing number of input networks the search space grows exponentially, the time required by GEDEVO-M to align multiple copies of the same network also raises exponentially. For the further exploitation of the method on bigger networks, the execution time should be improved by, for example, constraining search space using relevant biological Data.



Dr. Arjun JAIN NATIONALITY: Indian DEPARTMENT: Computer Graphics

DISSERTATION TITLE:

Data-Driven Methods for Interactive Visual Content Creation and Manipulation

ABSTRACT OF PHD THESIS:

Software tools for creating and manipulating visual content — be they for images, video or 3D models — are often difficult to use and involve a lot of manual interaction at several stages of the process. Coupled with long processing and acquisition times, content production is rather costly and poses a potential barrier to many applications. Although cameras now allow anyone to easily capture photos and video, tools for manipulating such media demand both artistic talent and technical expertise. However, at the same time, vast corpuses with existing visual content such as Flickr, YouTube or Google 3D Warehouse are now available and easily accessible.

This thesis proposes a data-driven approach to tackle the above mentioned problems encountered in content generation. To this end, statistical models trained on semantic knowledge harvested from existing visual content corpuses are created. Using these models, we then develop tools which are easy to learn and use, even by novice users, but still produce high-quality content. These tools have intuitive interfaces, and enable the user to have precise and flexible control. Specifically, we apply our models to create tools to simplify the tasks of video manipulation, 3D modeling and material assignment to 3D objects.





Dr. Megha KHOSLA NATIONALITY: Indian DEPARTMENT: Algorithms and Complexity

DISSERTATION TITLE:

Multiple Choice Allocations with Small Maximum Loads

Abstract of PhD Thesis:

The idea of using multiple choices to improve allocation schemes is now well understood and is often illustrated by the following example. Suppose *n* balls are allocated to *n* bins with each ball choosing a bin independently and uniformly at random. The maximum load, or the number of balls in the most loaded bin, will then be approximately $\frac{\log n}{\log \log n}$ with high probability. Suppose now the balls are allocated sequentially by placing a ball in the least loaded bin among the k >2 bins chosen independently and uniformly at random. Azar, Broder, Karlin, and Upfal [1] showed that in this scenario, the maximum load drops to $\frac{\log \log n}{\log k} + \Theta$ (1), with high probability, which is an exponential improvement over the previous case.

In this thesis we investigate multiple choice allocations from a slightly different perspective. Instead of minimizing the maximum load, we fix the bin capacities and focus on maximizing the number of balls that can be allocated without overloading any bin. In the process that we consider we have m = [cn] balls and n bins. Each ball chooses k bins independently and uniformly at random. *Is it* possible to assign each ball to one of its choices such that the no bin receives more than k balls? For all k > 3 and l > 2 we give a critical value, $c_{k,l}^*$, such that when $c < c_{k,l}^*$ an allocation is possible with high probability and when $c > c_{k,l}^*$ this is not the case.

In case such an allocation exists, *how quickly can we find it?* Previous work on total allocation time for case k > 3 and l = 1 has analyzed a breadth first strategy which is shown to be linear only in expectation. We give a simple and efficient

algorithm which we also call local search allocation (LSA) to find an allocation for all k > 3 and l = 1. Provided the number of balls are below (but arbitrarily close to) the theoretical achievable load threshold, we give a linear bound for the total allocation time that holds with high probability. We demonstrate, through simulations, an order of magnitude improvement for total and maximum allocation times when compared to the state of the art method.

Our results find applications in many areas including hashing, load balancing, data management, orientability of random hypergraphs and maximum matchings in a special class of bipartite graphs.





Dr. Sarath Kumar KONDREDDI

NATIONALITY: Indian DEPARTMENT: Databases and Information Systems

DISSERTATION TITLE:

Human Computing and Crowdsourcing Methods for Knowledge Acquisition

Abstract of PhD Thesis:

Ambiguity, complexity, and diversity in natural language textual expressions are major hindrances to automated knowledge extraction. As a result stateof-the-art methods for extracting entities and relationships from unstructured data make incorrect extractions or produce noise. With the advent of human computing, computationally hard tasks have been addressed through human inputs. While text-based knowledge acquisition can benefit from this approach, humans alone cannot bear the burden of extracting knowledge from the vast textual resources that exist today. Even making payments for crowdsourced acquisition can quickly become prohibitively expensive.

In this thesis we present principled methods that effectively garner human computing inputs for improving the extraction of knowledge-base facts from natural language texts. Our methods complement automatic extraction techniques with human computing to reap benefits of both while overcoming each other's limitations. We present the architecture and implementation of HIGGINS, a system that combines an information extraction (IE) engine with a human computing (HC) engine to produce high quality facts. Using automated methods, the IE engine compiles dictionaries of entity names and relational phrases. It further combines statistics derived from large Web corpora with semantic resources like WordNet and ConceptNet to expand the dictionary of relational phrases. It employs specifically designed statistical language models for phrase relatedness to come up with questions and relevant candidate answers that are presented to human workers. Through extensive experiments we establish the superiority of this approach in extracting relation-centric

facts from text. In our experiments we extract facts about fictitious characters in narrative text, where the issues of diversity and complexity in expressing relations are far more pronounced. Finally, we also demonstrate how interesting human computing games can be designed for knowledge acquisition tasks.





Dr. Christian KURZ NATIONALITY: Austrian DEPARTMENT: Computer Graphics

DISSERTATION TITLE:

Constrained Camera Motion Estimation and 3D Reconstruction

Abstract of PhD Thesis:

The creation of virtual content from visual data is a tedious task which requires a high amount of skill and expertise. Although the majority of consumers is in possession of multiple imaging devices that would enable them to perform this task in principle, the processing techniques and tools are still intended for the use by trained experts. As more and more capable hardware becomes available, there is a growing need among consumers and professionals alike for new flexible and reliable tools that reduce the amount of time and effort required to create high-quality content.

This thesis describes advances of the state of the art in three areas of computer vision: camera motion estimation, probabilistic 3D reconstruction, and template fitting. First, a new camera model geared towards stereoscopic input data is introduced, which is subsequently developed into a generalized framework for constrained camera motion estimation. A probabilistic reconstruction method for 3D line segments is then described, which takes global connectivity constraints into account. Finally, a new framework for symmetry-aware template fitting is presented, which allows the creation of high-quality models from low-quality input 3D scans. Evaluations with a broad range of challenging synthetic and real-world data sets demonstrate that the new constrained camera motion estimation methods provide improved accuracy and flexibility, and that the new constrained 3D reconstruction methods improve the current state of the art.



Dr. Manuel LAMOTTE-SCHUBERT

NATIONALITY: German DEPARTMENT: Automation of Logic

DISSERTATION TITLE:

Automatic Authorization Analysis

ABSTRACT OF PHD THESIS:

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.

This 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. This thesis shows that the Hyper-resolution calculus modulo



redundancy elimination terminates on *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* have 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.



Dr. Faraz MAKARI MANSHADI

NATIONALITY: Iranian DEPARTMENT: Databases and Information Systems

DISSERTATION TITLE:

Scalable Optimization Algorithms for Recommender Systems

Abstract of PhD Thesis:

Recommender systems have now gained significant popularity and been widely used in many e-commerce applications. Predicting user preferences is a key step to providing high quality recommendations. In practice, however, suggestions made to users must not only consider user preferences in isolation; a good recommendation engine also needs to account for certain constraints. For instance, an online video rental that suggests multimedia items (e.g., DVDs) to its customers should consider the availability of DVDs in stock to reduce customer waiting times for accepted recommendations. Moreover, every user should receive a small but sufficient number of suggestions that the user is likely to be interested in.

This thesis aims to develop and implement scalable optimization algorithms that can be used (but are not restricted) to generate recommendations satisfying certain objectives and constraints like the ones above. State-of-the-art approaches lack efficiency and/or scalability in coping with large real-world instances, which may involve millions of users and items. First, we study large-scale matrix completion in the context of collaborative filtering in recommender systems. For such problems, we propose a set of novel shared-nothing algorithms which are designed to run on a small cluster of commodity nodes and outperform alternative approaches in terms of efficiency, scalability, and memory footprint. Next, we view our recommendation task as a generalized matching problem, and propose the first distributed solution for solving such problems at scale. Our algorithm is designed to run on a small cluster of commodity nodes (or in a MapReduce environment) and has strong approxima-



tion guarantees. Our matching algorithm relies on linear programming. To this end, we present an efficient distributed approximation algorithm for mixed packing-covering linear programs, a simple but expressive subclass of linear programs. Our approximation algorithm requires a poly-logarithmic number of passes over the input, is simple, and well-suited for parallel processing on GPUs, in shared-memory architectures, as well as on a small cluster of commodity nodes.

Dr. Steffen METZGER

NATIONALITY: German DEPARTMENT: Databases and Information Systems

DISSERTATION TITLE:

User-Centric Knowledge Extraction and Maintenance

ABSTRACT OF PHD THESIS:

An ontology is a machine readable knowledge collection. There is an abundance of information available for human consumption. Thus, large general knowledge ontologies are typically generated tapping into this information source using imperfect automatic extraction approaches that translate human readable text into machine readable semantic knowledge. This thesis provides methods for user-driven ontology generation and maintenance. In particular, this work consists of three main contributions:

- 1. An interactive human-supported extraction tool: LUKe. The system extends an automatic extraction framework to integrate human feedback on extraction decisions and extracted information on multiple levels.
- 2. A document retrieval approach based on semantic statements: S3K. While one application is the retrieval of documents that support extracted information to verify the correctness of the piece of information, another application in combination with an extraction system is a fact based indexing of a document corpus allowing statement based document retrieval.
- 3. A method for similarity based ontology navigation: QBEES. The approach enables search by example. That is, given a set of semantic entities, it provides the most similar entities with respect to their semantic properties considering different aspects.

All three components are integrated into a modular architecture that also provides an interface for third-party components.





Dr. Chuong Hoang NGUYEN

NATIONALITY: Vietnamese DEPARTMENT: Computer Graphics

DISSERTATION TITLE:

Data-Driven Approaches for Interactive Appearance Editing

Abstract of PhD Thesis:

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 examplebased 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.

Dr. Bojan PEPIK

NATIONALITY: Macedonian
DEPARTMENT: Computer Vision and Multimodal Computing

DISSERTATION TITLE:

Richer Object Representations for Object Class Detection in Challenging Real World Images

ABSTRACT OF PHD THESIS:

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, finegrained 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.





Dr. Ali POURMIRI

NATIONALITY: Iranian DEPARTMENT: Algorithms and Complexity

DISSERTATION TITLE:

Random Walk-Based Algorithms on Networks

ABSTRACT OF PHD THESIS:

The present thesis studies some important random walk-based algorithms, which are randomized rumor spreading and balanced allocation protocols on networks. In the first part of the thesis, we study the Push and the Push-Pull protocols introduced by [DGG⁺87], which are basic randomized protocols for information dissemination on networks. In Chapter 2, we propose a new model where the number of calls of each node in every round is chosen independently according to a probability distribution *R* with bounded mean determined at the beginning of the process. In addition to the model being a natural extension of the standard protocols, it also serves as a more realistic model for rumor spreading in a network whose entities are not completely uniform and may have different levels of power. We provide both lower and upper bounds on the rumor spreading time depending on statistical properties of R such as the mean or the variance. While it is well-known that the standard protocols need θ (log n) rounds to spread a rumor on a complete network with n nodes, we show that, if *R* follows a power law distribution with exponent $\beta \in (2,3)$, then the Push-Pull protocol spreads a rumor in θ (log log n) rounds. Moreover, when β = 3, we show a runtime of θ ($\frac{\log n}{\log \log n}$). In Chapter 3, we analyze the behavior of the standard Push-Pull protocol on a class of random graphs, called random k-trees for every integer $k \ge 2$, that are suitable to model poorly connected, small-world and scale free networks. Here, we show that the Push-Pull protocol propagates a rumor from a randomly chosen informed node to almost all nodes of a random *k*-tree with n nodes in O ((log n) 1+ ck) rounds with high probability, where $O < c_k \le 1$ is a decreasing function in k. We also derive a lower bound of $n^{\Omega(1)}$ for the runtime of the protocol to inform all nodes of the graph. Our technique for proving the upper bound is successfully carried over to a closely related class of random graphs called random *k*-Apollonian networks.

We devote the rest of the thesis to the study of random walks on graphs, covering both practical and theoretical aspects. In Chapter 4, we show the existence of a cutoff phenomenon for simple random walks on Kneser graphs. A cutoff phenomenon for a given sequence of ergodic Markov chains describes a sharp transition in the convergence of the chains to its stationary distribution over a negligible period of time, known as the *cutoff window*. In order to establish the cutoff phenomenon, we combine the spectral information of the transition matrix and a probabilistic technique, known as Wilson's method [WilO4]. And finally in Chapter 5, by using *non-backtracking* random walks introduced by Alon et al. [ABLSO7], we propose a new algorithm for sequentially allocating *n* balls into *n* bins that are organized as a *d*-regular graph with n nodes, say *G*, where *d* > 3 can be any integer. Let *l* be a given positive integer. In each round $t, 1 \le t \le n$, ball *t* picks a node of *G* uniformly at random and performs a non-Backtracking random walk of length *l* from the chosen node. Then it deterministically selects a subset of the visited nodes as the potential choices and allocates itself on one of the choices with minimum load (ties are broken uniformly at random). 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. We also show that the upper bound is tight up to a O (log log n) factor. In particular, we show that if we set $l=[(log n) \frac{1+\epsilon}{2}]$, for any constant $\epsilon \in (0,1]$, and *G* has girth at least $\omega(l)$, then the maximum load is bounded by $O(1/ \in)$ with high probability.





Dr. Fahimeh RAMEZANI

NATIONALITY: Iranian DEPARTMENT: Algorithms and Complexity

DISSERTATION TITLE:

Application of Multiplicative Weights Update Method in Algorithmic Game Theory

Abstract of PhD Thesis:

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 [LSO5, LS11] 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 (e.g., see [LSO5, LS11, CEF10, HKV11] for applications), but unlikely to be efficient in practice, because it uses the Ellipsoid method. In Chapter 2, 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.

In Chapter 3, we consider the problem of computing an approximate saddle point, or equivalently equilibrium, for a convex-concave functions $F: X \times Y \rightarrow \mathbb{R}$, where X and Y are convex sets of arbitrary dimensions. Our main contribution is the design of a randomized algorithm for computing an ε -approximation saddle point for *F*. Our algorithm is based on combining a technique developed

by Grigoriadis and Khachiyan [GK95], which is a randomized variant of Brown's fictitious play [Bro51], with the recent results on random sampling from convex sets (see, e.g., [LV06, Vem05]). The algorithm finds an ε -approximation saddle point in an expected number of $O\left(\frac{p^2(n+m)}{\varepsilon^2}\log\frac{R}{\varepsilon}\right)$ iterations, where in each iteration two points are sampled from log-concave distributions over strategy sets. It is assumed that X and Y have inscribed balls of radius 1/R and circumscribing balls of radius R and $p = \max_{x \in X, y \in Y} |F(x,y)|$. In particular, the algorithm requires $O^*\left(\frac{p^2(n+m)^6}{\varepsilon^2}\log R\right)$ calls to a membership oracle, where $O^*(\cdot)$ suppresses polylogarithmic factors that depend on n,m, and ε .




Dr. Ilya RESHETOUSKI

NATIONALITY: Belarusian DEPARTMENT: Generalized Image Acquisition and Analysis

DISSERTATION TITLE:

Kaleidoscopic Imaging

ABSTRACT OF PHD THESIS:

Kaleidoscopes have a great potential in computational photography as a tool for redistributing light rays. In time-of-flight imaging the concept of the kaleidoscope is also useful when dealing with the reconstruction of the geometry that causes multiple reflections. This work is a step towards opening new possibilities for the use of mirror systems as well as towards making their use more practical. The focus of this work is the analysis of planar kaleidoscope systems to enable their practical applicability in 3D imaging tasks.

We analyse important practical properties of mirror systems and develop a theoretical toolbox for dealing with planar kaleidoscopes. Based on this theoretical toolbox we explore the use of planar kaleidoscopes for multi-view imaging and for the acquisition of 3D objects. The knowledge of the mirrors positions is crucial for these multi-view applications. On the other hand, the reconstruction of the geometry of a mirror room from time-of-flight measurements is also an important problem. We therefore employ the developed tools for solving this problem using multiple observations of a single scene point.



Dr. Christine RIZKALLAH

NATIONALITY: Egyptian DEPARTMENT: Algorithms and Complexity

DISSERTATION TITLE:

Verification of Program Computations

ABSTRACT OF PHD THESIS:

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. 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, Australia's Information and Communications Technology Research Centre of Excellence. This work contributes to building a feasible framework for verifying efficient file systems code. As opposed to the algorithmic problems we address in this thesis, file systems code is mostly straightforward and hence a good candidate for automation.





Dr. Marcus ROHRBACH

NATIONALITY: German
DEPARTMENT: Computer Vision and Multimodal Computing

DISSERTATION TITLE:

Combining Visual Recognition and Computational Linguistics: Linguistic Knowledge for Visual Recognition and Natural Language Descriptions of Visual Content

ABSTRACT OF PHD THESIS:

Extensive efforts are being made to improve visual recognition and semantic understanding of language. However, surprisingly little has been done to exploit the mutual benefits of combining both fields. In this thesis we show how the different fields of research can profit from each other.

First, we scale recognition to 200 unseen object classes and show how to extract robust semantic relatedness from linguistic resources. Our novel approach extends zero-shot to few shot recognition and exploits unlabeled data by adopting label propagation for transfer learning.

Second, we capture the high variability but low availability of composite activity videos by extracting the essential information from text descriptions. For this we recorded and annotated a corpus for fine-grained activity recognition. We show improvements in a supervised case but we are also able to recognize unseen composite activities.

Third, we present a corpus of videos and aligned descriptions. We use it for grounding activity descriptions and for learning how to automatically generate natural language descriptions for a video. We show that our proposed approach is also applicable to image description and that it outperforms baselines and related work.

In summary, this thesis presents a novel approach for automatic video description and shows the benefits of extracting linguistic knowledge for object and activity recognition as well as the advantage of visual recognition for understanding activity descriptions.



Dr. Richard RÖTTGER NATIONALITY: German DEPARTMENT: Computational Biology and Applied Algorithmics

DISSERTATION TITLE:

Active Transitivity Clustering of Large-Scale Biomedical Datasets

ABSTRACT OF PHD THESIS:

Clustering is a popular computational approach for partitioning data sets into groups of objects that share common traits. Due to recent advances in wetlab technology, the amount of available biological data grows exponentially and increasingly poses problems in terms of computational complexity for current clustering approaches. In this thesis, we introduce two novel approaches, TransClustMV and ActiveTransClust, that enable the handling of large scale datasets by reducing the amount of required information drastically by means of exploiting missing values.

Furthermore, there exists a plethora of different clustering tools and standards making it very difficult for researchers to choose the correct methods for a given problem. In order to clarify this multifarious field, we developed ClustEval which streamlines the clustering process and enables practitioners conducting large-scale cluster analyses in a standardized and bias-free manner.

We conclude the thesis by demonstrating the power of clustering tools and the need for the previously developed methods by conducting real-world analyses. We transferred the regulatory network of *E. coli K-12* to pathogenic EHEC organisms based on evolutionary conservation therefore avoiding tedious and potentially dangerous wet-lab experiments. In another example, we identify pathogenicity specific core genomes of actinobacteria in order to identify potential drug targets.





Dr. Sven Eric SCHELHORN

NATIONALITY: German DEPARTMENT: Computational Biology and Applied Algorithmics

DISSERTATION TITLE:

Going Viral: an Integrated View on Virological Data Analysis from Basic Research to Clinical Applications

ABSTRACT OF PHD THESIS:

Viruses are of considerable interest for several fields of life science research. The genomic richness of these entities, their environmental abundance, as well as their high adaptability and, potentially, pathogenicity make treatment of viral diseases challenging. This thesis proposes three novel contributions to antiviral research that each concern analysis procedures of high-throughput experimental genomics data. First, a sensitive approach for detecting viral genomes and transcripts in sequencing data of human cancers is presented that improves upon prior approaches by allowing detection of viral nucleotide sequences that consist of human-viral homologs or are diverged from known reference sequences. Second, a computational method for inferring physical protein contacts from experimental protein complex purification assays is put forward that allows statistically meaningful integration of multiple data sets and is able to infer protein contacts of transiently binding protein classes such as kinases and molecular chaperones. Third, an investigation of minute changes in viral genomic populations upon treatment of patients with the mutagen ribavirin is presented that first characterizes the mutagenic effect of this drug on the hepatitis C virus based on deep sequencing data.



Dr. Stephan SEUFERT NATIONALITY: German DEPARTMENT: Databases and Information Systems

DISSERTATION TITLE:

Algorithmic Building Blocks for Relationship Analysis over Large Graphs

ABSTRACT OF PHD THESIS:

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.





Dr. Martin SUDA NATIONALITY: Czech DEPARTMENT: Automation of Logic

DISSERTATION TITLE:

Resolution-Based Methods for Linear Temporal Reasoning

Abstract of PhD Thesis:

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. It is expected that other implementations would benefit from it in an analogous way. 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.



Dr. Tomasz TYLENDA NATIONALITY: Polish DEPARTMENT: Databases and Information Systems

DISSERTATION TITLE:

Methods and Tools for Summarization of Entities and Facts in Knowledge Bases

Abstract of PhD Thesis:

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 document.





Dr. Zhongjie WANG NATIONALITY: Chinese DEPARTMENT: Computer Graphics

DISSERTATION TITLE:

Pattern Search for the Visualization of Scalar, Vector, and Line Fields

Abstract of PhD Thesis:

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 userdefined 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.





Dr. Chenglei WU NATIONALITY: Chinese DEPARTMENT: Computer Graphics

DISSERTATION TITLE:

Inverse Rendering for Scene Reconstruction in General Environments

ABSTRACT OF PHD THESIS:

Demand for high-quality 3D content has been exploding recently, owing to the advances in 3D displays and 3D printing. However, due to insufficient 3D content, the potential of 3D display and printing technology has not been realized to its full extent. Techniques for capturing the real world, which are able to generate 3D models from captured images or videos, are a hot research topic in computer graphics and computer vision. Despite significant progress, many methods are still highly constrained and require lots of prerequisites to succeed. Marker-less performance capture is one such dynamic scene reconstruction technique that is still confined to studio environments. The requirements involved, such as the need for a multi-view camera setup, specially engineered lighting or green-screen backgrounds, prevent these methods from being widely used by the film industry or even by ordinary consumers.

In the area of scene reconstruction from images or videos, this thesis proposes new techniques that succeed in general environments, even using as few as two cameras. Contributions are made in terms of reducing the constraints of marker-less performance capture on lighting, background and the required number of cameras. The primary theoretical contribution lies in the investigation of light transport mechanisms for high-quality 3D reconstruction in general environments. Several steps are taken to approach the goal of scene reconstruction in general environments. At first, the concept of employing inverse rendering for scene reconstruction is demonstrated on static scenes, where a high-quality multi-view 3D reconstruction method under general unknown illumination is developed. Then, this concept is extended to dynamic scene

reconstruction from multi-view video, where detailed 3D models of dynamic scenes can be captured under general and even varying lighting, and in front of a general scene background without a green screen. Finally, efforts are made to reduce the number of cameras employed. New performance capture methods using as few as two cameras are proposed to capture high-quality 3D geometry in general environments, even outdoors.





Dr. Mohamed Amir YOSEF

NATIONALITY: Egyptian DEPARTMENT: Databases and Information Systems

DISSERTATION TITLE:

U-AIDA: a Customizable System for Named Entity Recognition, Classification, and Disambiguation

Abstract of PhD Thesis:

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 finegrained type classification of names in natural language text.

Current Students

ΝΑΜΕ	NATIONALITY
AGARWAL, Prabal	Indian
AMANOVA, Dilafruz	Uzbek
BOZCA, Sinan	Turkish
CHU, Cuong Xuan	Vietnamese
DARWISH, Omar	Egyptian
DIVO, Michaël	French
FIERARU, Mihai	Romanian
GROVER, Priyanka	Indian
HANKA, Adam	Czech
HEINEN, Tobias	German
MUKKAMALA, Mahesh Chandra	Indian
RAKHMANBERDIEVA, Nurzat	Kyrgyztan
RAMIREZ CHAVEZ, Luis Enrique	Mexican
SETHI, Riccha	Indian
SHARIFULLAEVA, Firuza	Uzbekistan
TEKLEHAIMANOT, Birhan Tadele	Ethiopian
ZHENG, Menglin	Chinese

* Master's students are assigned a scientific supervisor and become members of a research group when they start work on their thesis. Until then, therefore, they are supervised by the IMPRS-CS coordinator.

C	00	
C	00	

NAME	SUPERVISOR	DEPARTMENT	NATIONALITY
ALAGI, Gábor	Christoph Weidenbach	Automation of Logic	Hungarian
BASTYS, Tomas	Olga Kalinina	Computational Biology and Applied Algorithmics	Lithuanian
BECKER, Ruben	Andreas Karrenbauer	Algorithms and Complexity	German
BICHHAWAT, Abhishek	Christian Hammer	Algorithms and Complexity	Indian
BIEGA, Joanna	Gerhard Weikum	Databases and Information Systems	Polish
BYELOZYOROV, Sergiy	Philipp Slusallek, Thorsten Thormählen	Computer Graphics	Ukrainian
CHIU, Wei-Chen	Mario Fritz	Computer Vision and Multimodal Computing	Chinese
CHOUDHARY, Aruni	Michael Kerber	Algorithms and Complexity	Indian
CROITORU, Cosmina	Kurt Melhorn	Algorithms and Complexity	Romanian
DANILEWSKI, Piotr	Philipp Slusallek, Christian Theobalt	Computer Graphics	Polish
DEHGHANI AMIRABAD, Azim	Marcel Schulz	Computational Biology and Applied Algorithmics	Iranian
DEL CORRO, Luciano	Gerhard Weikum	Databases and Information Systems	Spanish
DIRNBERGER, Michael	Kurt Mehlhorn	Algorithms and Complexity	Austrian
DOERFERT, Johannes	Christoph Weidenbach	Automation of Logic	German
DÖRING, Matthias	Nico Pfeifer	Computational Biology and Applied Algorithmics	German
DONCHEVA, Nadezhda	Mario Albrecht, Thomas Lengauer	Computational Biology and Applied Algorithmics	Bulgarian
DUDZIAK, Tomasz	Reinhard Wilhelm, Sebastian Hack	Programming Languages and Compiler Construction	Polish
DUTTA, Sourav	Gerhard Weikum	Databases and Information Systems	Indian
EBERT, Peter	Thomas Lengauer	Computational Biology and Applied Algorithmics	German
ELDAROV, Nasimi	Thorsten Herfet, Gerhard Weikum	Telecommunications	Azerbaijanian
ELEK, Oskar	Tobias Ritschel	Computer Graphics	Czech
ERNST, Patrick	Gerhard Weikum	Databases and Information Systems	German

NAME	SUPERVISOR	DEPARTMENT	NATIONALITY
FELDMANN, Anna	Nico Pfeifer	Computational Biology and Applied Algorithmics	German
FERNANDES ALBRECHT, Felipe	Thomas Lengauer	Computational Biology and Applied Algorithmics	German
GAD-ELRAB, Mohamed	Gerhard Weikum	Databases and Information Systems	Egyptian
GHAFFAARI, Ali	Tobias Marschall	Computational Biology and Applied Algorithmics	Iranian
GROSS, Florian	Andreas Zeller, Gerhard Weikum	Software Engineering	German
GRYADITSKAYA, Iulia	Karol Myszkowski	Computer Graphics	Russian
GRYCNER, Adam	Gerhard Weikum	Databases and Information Systems	Polish
GUPTA, DHRUV	Klaus Berberich	Databases and Information Systems	Indian
GURAJADA, Venkata Sairam	Gerhard Weikum	Databases and Information Systems	Indian
HASHEMI, Vahid	Holger Herrmanns	Dependable Systems and Software	Iranian
HATEFIARDAKANI, Hassan	Holger Hermanns, Christoph Weidenbach	Dependable Systems and Software	Iranian
HAUSCHILD, Anne-Christin	Jan Baumbach, Volkhard Helms	Computational Systems Biology	German
HOPPE, Sabrina	Andreas Bulling	Computer Vision and Multimodal Computing	German
HUI, Kai	Klaus Berberich, Gerhard Weikum	Databases and Information Systems	Chinese
HUSSEIN AZMY, Noran	Christoph Weidenbach	Automation of Logic	Egyptian
ISSAC, Davis	Andreas Karrenbauer	Algorithms and Complexity	Indian
JOHN, Maximilian	Andreas Karrenbauer	Algorithms and Complexity	German
KALAGHATGI, Prabhav	Glenn Lawyer	Computational Biology and Applied Algorithmics	Indian
KALOJANOV, Javor	Philipp Slusallek	Computer Graphics	Bulgarian
KARAEV, Sanjar	Pauli Miettinen	Databases and Information Systems	Uzbek
KELLNHOFER, Petr	Karol Myszkowski	Computer Graphics	Czech

89

International Max Planck Research School for Computer Science

NAME	SUPERVISOR	DEPARTMENT	NATIONALITY
KHOREVA, Anna	Bernt Schiele	Computer Vision and Multimodal Computing	Russian
KLEIN, Felix	Bernd Finkbeiner, Christoph Weidenbach	Reactive Systems	German
KOBEL, Alexander	Michael Sagraloff	Algorithms and Complexity	German
KODRIC, Bojana	Martin Hoefer	Algorithms and Complexity	Croatian
KOLEV, Pavel	He Sun	Algorithms and Complexity	Bulgarian
KOSTA, Marek	Christoph Weidenbach	Automation of Logic	Slovakian
KUZEY, Erdal	Gerhard Weikum	Databases and Information Systems	Turkish
LI, Wenbin	Mario Fritz	Computer Vision and Multimodal Computing	Chinese
MALINOWSKI, Mateusz	Mario Fritz	Computer Vision and Multimodal Computing	Polish
MANDROS, Panagiotis	Jilles Vreeken	Exploratory Data Analysis	Greek
MARSALEK, Lukas	Philipp Slusallek	Computer Graphics	Czech
MEKA, Abhimitra	Christian Theobalt	Computer Graphics	Indian
METZLER, Saskia	Pauli Miettinen	Databases and Information Systems	German
MORAN, Shay	Kurt Mehlhorn	Algorithms and Complexity	Israeli
MÜLLER, Fabian	Thomas Lengauer	Computational Biology and Applied Algorithmics	German
MUKHERJEE, Subhabrata	Gerhard Weikum	Databases and Information Systems	Indian
NAG CHOWDHURY, Sreyasi	Gerhard Weikum	Databases and Information Systems	Indian
NGUYEN, Dat Ba	Gerhard Weikum	Databases and Information Systems	Vietnamese
NIKUMBH, Sarvesh Prakash	Nico Pfeifer	Computational Biology and Applied Algorithmics	Indian
OMRAN, Mohamed	Bernt Schiele	Computer Vision and Multimodal Computing	Egyptian
OTT, Sebastian	Rob van Stee	Algorithms and Complexity	German
PISHCHULIN, Leonid	Bernt Schiele	Computer Vision and Multimodal Computing	Russian

ΝΑΜΕ	SUPERVISOR	DEPARTMENT	NATIONALITY
PRYTKOVA, Natalia	Gerhard Weikum	Databases and Information Systems	Russian
RANGAPURAM, Syama Sundar Yadav	Matthias Hein, Kurt Mehlhorn	Machine Learning	Indian
RHODIN, Helge	Christian Theobalt	Computer Graphics	German
ROBERTINI, Nadia	Christian Theobalt	Computer Graphics	Italian
ROHRBACH, Anna	Bernt Schiele	Computer Vision and Multimodal Computing	Ukrainian
SATTAR, Hosnieh	Andreas Bulling, Mario Fritz	Computer Vision and Multimodal Computing	Iranian
SIMEONOVSKI, Milivoj	Michael Backes, Gerhard Weikum	Information Security and Cryptography	Macedonian
SIU, Yin Amy	Gerhard Weikum	Databases and Information Systems	Australian
SPEICHER, Nora	Nico Pfeifer	Computational Biology and Applied Algorithmics	German
SRIDHAR, Srinath	Christian Theobalt, Antti Oulasvirta	Computer Graphics	Indian
STEIL, JULIAN	Andreas Bulling	Computer Vision and Multimodal Computing	German
SUN, Peng	Jan Baumbach, Jiong Guo	Computational Systems Biology	Chinese
TANDON, Niket	Gerhard Weikum	Databases and Information Systems	Indian
TANG, Ching Hoo	Christoph Weidenbach	Automation of Logic	British
TEFLIOUDI, Christina	Gerhard Weikum	Databases and Information Systems	Greek
VAZ, Daniel	Parinya Chalermsook	Algorithms and Complexity	Portuguese
VOIGT, Marco	Christoph Weidenbach	Automation of Logic	German
WU, Xiaokun	Thorsten Thormählen	Computer Graphics	Chinese
YAHYA, Mohamed	Gerhard Weikum	Databases and Information Systems	Palestinian
ZHANG, Xucong	Andreas Bulling	Computer Vision and Multimodal Computing	Chinese

Staff



Aaron ALSANCAK

NATIONALITY:	German
POSITION:	Student Assistant
E-MAIL:	aalsancak@mpi-inf.mpg.de
EDUCATION:	Comparative Languages and Literatures and Translation
UNIVERSITY:	Saarland University



Meike BRODÉ

NATIONALITY:	German
Position:	Student Assistant
E-MAIL:	mbrode@mpi-inf.mpg.de
EDUCATION:	Intercultural Communication (French, Spanish) and Business Administration
UNIVERSITY:	Saarland University



Mark DUANCE

NATIONALITY:	British
Position:	English Teacher
E-MAIL:	mduance@mpi-inf.mpg.de
EDUCATION:	Psychology
UNIVERSITY:	Goldsmiths University of London



Jennifer GERLING

NATIONALITY:	German
POSITION:	IMPRS-CS Coordinator
E-MAIL:	jgerling@mpi-inf.mpg.de
EDUCATION: UNIVERSITY:	Applied Linguistics and Translating (Spanish, English, and German) Saarland University



Isthar GREINER

NATIONALITY:	German
POSITION:	German Teacher
E-MAIL:	igreiner@mpi-inf.mpg.de
EDUCATION:	Teacher for Secondary Education (Academic High School: German and English), German as a Foreign Language
UNIVERSITY:	Saarland University



Olga GRIGORIEVA

NATIONALITY:	Moldavian
POSITION:	German Teacher
E-MAIL:	ogrigori@mpi-inf.mpg.de
EDUCATION:	German, German as a Foreign Language
UNIVERSITY:	Alecu Russo State University of Bălți, Saarland University



Sarah HELLICAR

NATIONALITY:	British
Position:	English Teacher
E-MAIL:	hellicar@mpi-inf.mpg.de
EDUCATION:	BA in Comparative Literature
UNIVERSITY:	King's College London



Jeanine HORN

NATIONALITY:	German / French
POSITION:	German Teacher
E-MAIL:	jhorn@mpi-inf.mpg.de
EDUCATION:	Teacher for Secondary Education (Academic High School: German and History). German as a Foreign Language
UNIVERSITY:	Saarland University



Stephanie JÖRG

NATIONALITY:	German
POSITION:	Assistant to the Coordinator
E-MAIL:	sjoerg@mpi-inf.mpg.de
EDUCATION:	Translation (English and Spanish)
UNIVERSITY:	Saarland University



Dorothy MACDONALD

NATIONALITY:	British
POSITION:	English Teacher
E-MAIL:	dorothy@mpi-inf.mpg.de
EDUCATION:	Health Sciences
UNIVERSITY:	University College of Ripon and York



Andrea RUFFING

NATIONALITY:	German
POSITION:	IMPRS-CS Coordinator
E-MAIL:	aruffing@mpi-inf.mpg.de
EDUCATION:	Applied Linguistics and Translation (French, English, German); Literature and Linguistics
UNIVERSITY:	Saarland University



Elena TREGUBOVA

NATIONALITY:	Russian
POSITION:	German Teacher
E-MAIL:	etregubo@mpi-inf.mpg.de
EDUCATION:	Philology, German, Translating/Interpreting, Education
UNIVERSITY:	Voronezh State University; Voronezh State Technical University



Janine VIEWEG

NATIONALITY:	German
POSITION:	Student Assistant
E-MAIL:	jvieweg@mpi-inf.mpg.de
EDUCATION:	Sports Sciences
UNIVERSITY:	Saarland University



Prof. Dr.-Ing. Gerhard WEIKUM

NATIONALITY:	German
POSITION:	Spokesperson
E-MAIL:	weikum@mpi-inf.mpg.de
EDUCATION:	Computer Science
UNIVERSITY:	University of Darmstadt

Awards

AWARDS

THE FOLLOWING IMPRS STUDENTS RECEIVED AWARDS FOR THEIR WORK IN 2014:

NAME OF STUDENT	Award
Jindal Gorav	Günter Hotz Medal
Sourav Dutta	Google European Doctoral Fellowship 2014
Ndapandula Nakashole	Otto Hahn Medal 2014
<i>Maximilian</i> Dylla	Dissertation Award of the German Computer Science Society (Gesellschaft für Informatik, GI)

THE FOLLOWING IMPRS STUDENTS RECEIVED AWARDS FOR THEIR WORK IN 2015:

NAME OF STUDENT	Award
Mohamed Gad-Elrab	Günter Hotz Medal
Anne-Christin Hauschild	Metabolites Best Paper Award, Third prize



International Max Planck Research School for Computer Science

Publisher

International Max Planck Research School

for Computer Science



International Max Planck Research School for Computer Science Max Planck Institute for Informatics Campus E1 4 | 66123 Saarbrücken | Germany Telefon +49 681 9325 1800 Fax +49 681 9325 1899 E-mail imprs@mpi-inf.mpg.de Internet www.imprs-cs.de

www.imprs-cs.de





Max Planck Institute for Software Systems



COMPUTER SCIENCE

