Yearbook 2011/2012




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 2011/2012

CONTENTS

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 that 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.

IMPRS-CS was established in 2001. Since then, 150 IMPRS-CS students have obtained a doctoral degree, including 23 women and 44 non-Germans. Among these students, 11 have been awarded the Otto Hahn Medal of the Max Planck Society, and two have received the Otto Hahn Award, which includes a 5-year scholarship for an independent research group.

Eight IMPRS-CS students have received the Eduard Martin Dissertation Award at Saarland University. 26 of the institute's graduates since 2001 have become professors or have comparable positions in research labs (11 in Germany). Several of the institute's former students have won prestigious awards later in their careers. Currently, IMPRS-CS has 122 doctoral students, 24 of them women and 78 non-Germans. In addition, IMPRS-CS currently supports 20 students with fellowships towards master's degrees. We have successfully attracted industrial sponsors to provide a total of 14 three-year fellowships for doctoral students.

In 2012, the Max Planck Society (MPG) evaluated IMPRS-CS with the support of several internationally renowned computer science professors. One of the reviewers stated in his assessment: "I am very impressed with the school – it is certainly one of the best computer science research schools in the world and probably the best in Europe. The output of the school, both in terms of research results and PhD candidate production, is impressive. One key indicator of the quality of the school is the large number of PhD candidates that are now faculty at major universities or research scientists at major research labs." As a result of the positive evaluation in 2012, funding for the school was prolonged for another six years.

Alumni: M.Sc.

Nicolás ALCARAZ MILMAN

NATIONALITY: Mexican
DEPARTMENT: Computational Systems Biology

THESIS TITLE:

KeyPathwayMiner – Detecting Case-specific Biological Pathways by Using Expression Data

ABSTRACT OF MASTER'S THESIS:

Advances in the field of systems biology have provided the biological community with massive amounts of pathway data that describe the interplay of genes and their products. The resulting biological networks usually consist of thousands of entities and interactions that can be modeled mathematically as graphs. Since these networks only provide a static picture of the accumulated knowledge, pathways that are affected during development of complex diseases cannot be extracted easily. This gap can be filled by means of OMICS technologies such as DNA microarrays, which measure the activity of genes and proteins under different conditions. Integration of both interaction and expression datasets can increase the quality and accuracy of analysis when compared to independant inspection of each. However, sophisticated computational methods are needed to deal with the size of the datasets while also accounting for the presence of biological and technological noise inherent in the data generating process.

In this dissertation the KeyPathwayMiner is presented, a method that enables the extraction and visualization of affected pathways given the results of a series of gene expression studies. Specifically, given network and gene expression data, KeyPathwayMiner identifies those maximal subgraphs where all but k nodes of the subnetwork are differentially expressed in all but at most l cases in the gene expression data. This new formulation allows users to control the number of outliers with two parameters that provide good interpretability of the solutions. Since identifying these subgraphs is computationally intensive, an heuristic algorithm based on Ant Colony Optimization was designed and adapted to this problem, where solutions are reported in the order of seconds on a standard personal computer. The Key-PathwayMiner was tested on real Huntington's Disease and Breast

Cancer datasets, where it is able to extract pathways containing a large percentage of known relevant genes when compared to other similar approaches.

KeyPathwayMiner has been implemented as a plugin for Cytoscape, one of the most widely used open source biological network analysis and visualization platforms. The Key-PathwayMiner is available online at http://keypathwayminer.mpi-inf.mpg.de or through the plugin manager of Cytoscape.

Nikolaos ARVANITOPOULOS DARGINIS

NATIONALITY: Greek
DEPARTMENT: Machine Learning

THESIS TITLE:

Aggregation of Multiple Clusterings and Active Learning in a Transductive Setting

ABSTRACT OF MASTER'S THESIS:

In this work we proposed a novel transductive method to solve the problem of learning from partially labeled data. Our main idea was to aggregate information obtained from several clusterings to infer the labels of the unlabeled data. While our method is not restricted to a specific clustering method, we chose to use in our experiments the normalized variant of 1-spectral clustering, which was demonstrated to produce in most cases better clusterings than the standard spectral clustering method. Our approach yielded results which were at least comparable to, and in some cases even significantly better than the best results obtained by state-of-the-art methods reported in the literature.

Furthermore, we proposed a novel active learning framework that is able to query the labels of the most informative points which help in the classification of the unlabeled points. For the majority vote scheme we provided some guarantees on the number of points that should be drawn from each cluster in order to infer the correct label of the cluster with high probability. Moreover, in the ridge regression scheme we proposed an algorithm that in each step selects the most uncertain point in terms of the prediction function of the classifier (the point that lies near the decision boundary of the classifier). In both cases, experimental results show the strength of our methods and confirm our theoretical guarantees.

The results look very promising and open several interesting directions of future research. For the SSL scheme, it is interesting to test the performance of several other clustering approaches, such as k-means, standard spectral

clustering, hierarchical clustering, e.t.c. and combine them together in one general method. Our intuition is that the algorithm should be able to select only the good clusterings that provide discriminative information for each specific problem. Apart from ridge regression, it would be beneficial to experiment with other fitting approaches that produce sparse representations in our constructed basis. For the active learning framework, one interesting direction is to further generalize it into more general clusterings that take into account the hierarchical structure of data. In that way, we will take advantage of the underlying hierarchy and by adaptively selecting the pruning of the cluster tree we can (potentially) further improve our sampling strategy. Additionally, we believe that in the multi-clustering scenario extensive improvements of our algorithm can be proposed in order to better take advantage of the variation in the multiple clustering representations of the data. Finally, as our methods scale to large-scale problems and partially labeled data occurs in many different areas ranging from web documents to protein data, there is room for many interesting applications of the proposed methods.

Noran AZMY

NATIONALITY: Egyptian
DEPARTMENT: Atomation of Logics

THESIS TITLE:

Formula Renaming with Generalizations

ABSTRACT OF MASTER'S THESIS:

Most automated reasoning systems operate on formulas in conjunctive normal form (CNF). However, standard CNF translation of a given formula may result in an exponential blow-up of the formula size. In 1968, Tseitin introduced a workaround which came to be known as Tseitin's encoding or renaming: all subformulas are replaced by fresh atoms to escape repeated applications of the distributivity law and consequently obtain a shorter result, while maintaining satisfiability. Instead of renaming every subformula, Nonnengart and Weidenbach present criteria that determine the subformulas whose renaming yields positive benefit, i.e. a shorter CNF result. For first-order logic, we may employ one more significant enhancement: we may use one symbol to rename a set of compatible subformulas, i.e. subformulas that are instances of a common "general" formula. We denote this by *generalized renaming*. This work studies the generalized renaming problem and gives the first formal algorithm for generalized renaming, as well as an implementation of the algorithm as part of the SPASS theorem prover. Our tests show that generalized renaming significantly reduces the number of clauses in the general case.

Ervina CERGANI

NATIONALITY: **Albanian**

DEPARTMENT: **Databases and Information Systems**

THESIS TITLE:

Relation Extraction using Matrix Factorization Methods

ABSTRACT OF MASTER'S THESIS:

Relation extraction has an important role within the information extraction domain. Given an initial ontology specifying noun categories, instances from these categories and text corpora, the relation extraction task consists of extracting the relations that connect instances from these categories. Current research works related with this topic, are mostly based on the clustering methods. One well known method for this purpose is k-means clustering, which is used to extend the ontology of computer systems such as NELL, enabling them to increase over time their learning and extraction capabilities. In this work, we propose the use of matrix factorization methods for extracting relations between pairs of noun categories. Our approach uses two kinds of matrix factorization: Non-negative matrix factorization and Boolean matrix factorization. Using these methods, we intend to overcome the drawbacks that follows from k-means clustering, which is a hard clustering method, in the meaning that it forces the context patterns to be disjointly assigned to one cluster. In the real data, it might happen that not every context pattern belong semantically to one of the clusters. In the same time, it might happen as well that one context pattern can semantically belong to multiple clusters. These issues can be easily solved using matrix factorization methods. We tested our method on 2 different data sets: YAGO data set, and ClueWeb09 data set. For YAGO data set we have the disambiguated relations and instance pairs, on which we could compare our results. The ClueWeb09 data set, is the same data set used with the k-means clustering method presented in, giving us the opportunity to compare the results from both methods. Based on the experimental results, we conclude that matrix factorization is a useful approach for extracting relations between noun categories.

Ines CIOLACU

NATIONALITY: **Romanian**
DEPARTMENT: **Cryptographic Protocols**

THESIS TITLE:

Universally Composable Relativistic Commitments

ABSTRACT OF MASTER'S THESIS:

Designing communications protocols specifically adapted to relativistic situations (i.e. constrained by special relativity theory) is taking advantage of uniquely relativistic features to accomplish otherwise impossible tasks. Kent [Ken99] has demonstrated, for example, that secure bit commitment is possible using a protocol exploiting relativistic causality constraints, even though it is known to be impossible otherwise. Therefore, Kent's protocol gives a theoretical solution to the problem of finding commitment schemes secure over arbitrarily long time intervals. The functionality only requires from the committer a sequence of communications, including a post-revelation validation, each of which is guaranteed to be independent of its predecessor.

We propose to verify the security of the relativistic commitment not as a standalone protocol, but as an entity which is part of an unpredictable environment. To achieve this task we use the universal composability paradigm defined by Canetti [Can01]. The relevant property of the paradigm is the guarantee of security even when a secure protocol is composed with an arbitrary set of protocols, or, more generally, when the protocol is used as an element of a possibly complex system. Unfortunately, Kent's relativistic bit commitment satisfies universal composability only with certain restrictions on the adversarial model. However, we construct a two-party universal composable commitment protocol, also based on general relativistic assumptions.

Cosmina CROITORU

NATIONALITY: Romanian
DEPARTMENT: Algorithms and Complexity

THESIS TITLE:

Algorithmic Aspects of Abstract Argumentation Frameworks

ABSTRACT OF MASTER'S THESIS:

Arguments differ from proofs by the fact that they are defeasible, that is, the validity of their conclusions can be disputed by other arguments. The various Artificial Intelligence approaches to model this kind of non-monotonic reasoning in the presence of conflicting and incomplete information can be studied using the so called abstract argumentation frameworks. These are directed graphs, in which different families of sets of vertices are considered as possible outputs, representing sets of collectively accepted arguments. Finding these outputs is in general a computational intractable task.

In this thesis I use graph-theoretic and algorithmic tools in order to identify practically efficient argumentation systems.

I introduced a new type of argument acceptability, which can be used when the classical approaches fail, I proved that the corresponding decision problem is Π_2^p-complete, but I identified interesting polynomial time restrictions of it. Since argumentation frameworks model social interactions, I introduced a discipline policy for their construction without changing a specific outcome. Finally, I started a polyhedral approach for describing classical extensions, which can pave the way of using linear programming techniques for selecting solutions in the case of multiple status semantics.

The results obtained show that a study of practical problems raised by argumentation frameworks using advanced graph theoretic and algorithmic tools is important and necessary.

Mahdi EBRAHIMI

NATIONALITY: Iranian

DEPARTMENT: Databases and Information Systems

THESIS TITLE:

Solving Linear Programs in MapReduce

ABSTRACT OF MASTER'S THESIS:

Most interesting discrete optimization problems are NP-hard, thus no efficient algorithm to find optimal solution to such problems is likely to exist. Linear programming plays a central role in design and analysis of many approximation algorithms. However, linear program instances in real-world applications grow enormously. In this thesis, we study the Awerbuch-Khandekar parallel algorithm for approximating linear programs, provide strategies for efficient realization of the algorithm in MapReduce, and discuss methods to improve its performance in practice. Further, we characterize numerical properties of the algorithm by comparing it with partially-distributed optimization methods. Finally, we evaluate the algorithm on a weighted maximum satisfiability problem generated by SOFIE knowledge extraction framework on the complete Academic Corpus.

Ilya GONCHAROV

NATIONALITY: Russian

DEPARTMENT: Mathematical Image Analysis

THESIS TITLE:

Local Constancy Assumption Selection for Variational Optical Flow

ABSTRACT OF MASTER'S THESIS:

Variational methods are among the most successful approaches for computing high-quality optical flow. However, there are still many ways to improve. In this thesis we first provide a general overview of the main ideas of existing approaches by the example of the complementary optical flow method of Zimmer et al. This serves us as a starting point for introducing the concept of automatic local selection of the most suitable constancy assumption on image features, which allows to further improve the quality of optical flow estimation. As a main contribution, we provide the variational formulation, that directly leads to the proposed behaviour. The derived model is then analysed and evaluated in the series of experiments.

Javeria IQBAL

NATIONALITY: Pakistani

DEPARTMENT: Databases and Information Systems

THESIS TITLE:

Lineage Enabled Query Answering in Uncertain Knowledge Bases

ABSTRACT OF MASTER'S THESIS:

We present a unified framework for query answering over uncertain RDF knowledge bases. Specifically, our proposed design combines *correlated base facts* with a query driven, top down deductive grounding phase of first-order logic formulas (i.e., Horn rules) followed by a probabilistic inference phase. In addition to static *input correlations* among base facts, we employ the *lineage structure* obtained from processing the rules during grounding phase, in order to trace the logical dependencies of query answers (i.e., derived facts) back to the base facts. Thus, correlations (or more precisely: dependencies) among facts in a knowledge base may arise from two sources: 1) static input dependencies obtained from real-world observations; and 2) dynamic dependencies induced at query time by the rule based lineage structure of the query answer.

Our implementation employs state-of the-art inference techniques: We apply exact inference whenever tractable, the detection of *shared factors, shrink-age of Boolean formula* when feasible, and *Gibbs sampling* in the general case. Our experiments are conducted on real data sets with synthetic expansion of correlated base facts. The experimental evaluation demonstrates the practical viability and scalability of our approach, achieving interactive query response times over a very large knowledge base. The experimental results provide the success guarantee of our presented framework.

Violeta IVANOVA

NATIONALITY: Bulgarian
DEPARTMENT: Computational Biology and Applied Algorithmics

THESIS TITLE:

Comparison of Methods for the Discovery of Copy Number Aberrations Relevant to Cancer

ABSTRACT OF MASTER'S THESIS:

Recurrent genomic amplifications and deletions characterize cancer genomes and contribute to disease evolution. Array Comparative Genomic Hybridization (aCGH) technology allows detection of chromosomal copy number aberrations in the genomic DNA of tumors with high resolution. The association of consistent copy number aberrations with particular types of cancer facilitates the understanding of the pathogenesis of the disease, and contributes towards the improvement of diagnosis, prognosis and the development of drugs. However, distinguishing aberrations that are relevant to cancer from random background aberrations is a difficult task, due to the high dimensionality of the aCGH data. Different statistical methods have been developed to identify non-random gains and losses across multiple samples. Their approaches vary in several aspects: requirements necessary for an aberration to be recurrent, preprocessing of the input data, statistical approaches used for assessing significance of a recurrent aberration and other biological considerations they use. So far, multiple-sample analysis methods have only been evaluated qualitatively and their relative merits remain unknown. In this work we propose an approach for quantitative evaluation of the performance of four selected methods. We use simulated data with known aberrations to validate each method and we interpret the different outcomes. We also compare the performance of the methods on a collection of neuroblastoma tumors by quantifying the agreement between methods. We select appropriate techniques to combine the outputs of the methods into a meaningful aggregation in order to obtain a high confidence lists of significant copy number aberrations.

Hossein KHOSHNEVIS

NATIONALITY: Iranian

DEPARTMENT: Telecommunications

THESIS TITLE:

Discriminating 4G and Broadcast Signals via Cyclostationary Feature Detection

ABSTRACT OF MASTER'S THESIS:

According to the FCC, spectrum allocation will be one of the problems of future telecommunication systems. Indeed, the available parts of the spectrum have been assigned statically to some applications such as mobile networks and broadcasting systems; hence finding a proper operating band for new systems is difficult. These telecommunication systems are called *primary users*. However, primary users do not always use their entire bandwidth, and therefore a lot of spectrum holes can be detected. These spectrum holes can be utilized for undefined systems called *secondary users*. Federal communication commission (FCC) introduced *cognitive radio* which detects these holes and assigns them to secondary users.

There are several techniques for detention of signals such as energy based detection, matched filter detection and cyclostationary based detection. Cyclostationary based detection as one of the most sensitive methods, can be used for detection and classification of different systems. However, traditional multi-cycle and single-cycle detectors suffer from high complexity. Fortunately, using some prior knowledge about the signal, this shortcoming can be solved.

In this thesis, signals of DVB-T2 as a broadcasting system and 3GPP LTE and IEEE 802.16 (WiMAX) as mobile networks have been evaluated and two cyclostationary based algorithms for detection and classification of these signals in SISO and MIMO antenna configurations are proposed.

Yagiz KARGIN

NATIONALITY: Turkish

DEPARTMENT: Databases and Information Systems

THESIS TITLE:

Distributed Analytics over Web Archives

ABSTRACT OF MASTER'S THESIS:

Evolving content of the Web is being accumulated over time into Web archival collections. This creates the need for time travel search to explore the dynamics of the content. Text analytics has also a key role in exploring interesting information in text collections. Moreover, frequent phrase mining, a special case of text analytics, is an important analytical task that is motivated by the need of knowledge on frequent phrases in various areas of computer science, such as information retrieval and machine translation etc. However, time travel search and frequent phrase mining have to be conducted on increasingly large-scale data. Distributed approaches such as MapReduce, which is mainly designed to work on vast amount of text, can be utilized in this case. We address two separate problems in this thesis. The first problem is that time travel inverted index, which enables searching on the time dimension, is proposed in centralized setting. In our work, we parallelize the construction of time travel inverted index using MapReduce, having a distributed index as an end product. The second problem is that finding frequent phrases through naïve counting, even in MapReduce, is a time consuming task, because data to be processed gets much larger in size, when phrases are considered. As our work, we present *partitioned approximate phrase counting*, a very fast way to retrieve most of the frequent phrases together with their counts out of a collection to enable interactive analysis of the content. Included in this, we propose and a novel technique, *partitioned in-mapper combining*, which enables us to aggregate data in memory correctly, even though the data to be aggregated is larger than the available memory. Evaluation of experiments on New York Times Annotated Corpus, which contains roughly 2 million documents, show that our approach works at least 2 times faster as compared to naïve approach. It finds more than 90 % of frequent phrases with high precision. Moreover, it is able to find all highly frequent phrases exactly, along with their accurate counts. Furthermore, by a quick second pass on the data, we precisely provide most of the frequent phrases with their corresponding true counts, still being significantly faster than naïve approach.

Erdal KUZEY

NATIONALITY: Turkish

DEPARTMENT: Databases and Information System

THESIS TITLE:

Extraction of Temporal Facts and Events from Wikipedia

ABSTRACT OF MASTER'S THESIS:

In recent years, the great success of Wikipedia and the progress in information extraction techniques led to automatic construction of large scale knowledge bases which have Subject-Predicate-Object style facts extracted from both semi structured and natural language text of Wikipedia articles. Those knowledge bases consist of millions of entities, relations about them and their semantic types. Unfortunately, most of the current knowledge bases focus on static facts and ignore their temporal dimension, although, the vast majority of facts are evolving with time and are valid during a particular time period.

In this thesis, we introduce a complete information extraction framework which harvests temporal facts and events from semi-structured and free text of Wikipedia articles to enrich a temporal ontology (T-YAGO). Furthermore, this thesis discusses methods for introducing a temporal dimension to time-agnostic knowledge bases. In addition, several experiments and evaluations are presented to show the effectiveness of the methods proposed.

Dina MAHMOUD

NATIONALITY: Egyptian
DEPARTMENT: Mathematical Image Analysis

THESIS TITLE:

Multiple-Frame Image Super Resolution Based on Optic Flow

ABSTRACT OF MASTER'S THESIS:

Super resolution is the task of reconstructing one or several high resolution images, from one or several low resolution images. A variety of super resolution methods have been proposed over the past three decades, some following a single-frame based methodology while the others utilizing a multiple-frame based one. These methods are usually very sensitive to their underlying model of data and noise, which limits their performance. In this thesis, we propose and compare two multiple-frame based approaches that address such shortcomings. In the first proposal we investigate a fast, local approach which combines the low resolution frames via warping and then performs diffusion-based inpainting. The second proposal models the image formation process in a variational framework with regularization that is robust to errors in motion and blur estimation. In addition, we introduce a brightness adaptation step which results in images with sharper edges. An accurate estimation of optical flow among the low resolution measurements is a fundamental step towards high quality super resolution for both methods. Experiments confirm the effectiveness of our method on a variety of super resolution benchmark sequences, as well as its superiority in performance to other closely-related methods.

Mateusz MALINOWSKI

NATIONALITY: **Polish**
DEPARTMENT: **Probabilistic Machine Learning and Medical Image Processing**

THESIS TITLE:

Optimization Algorithms in the Reconstruction of MR Images: A Comparative Study

ABSTRACT OF MASTER'S THESIS:

Time that an imaging device needs to produce results is one of the most crucial factors in medical imaging. Shorter scanning duration causes fewer artifacts such as those created by the patient motion. In addition, it increases patient comfort and in the case of some imaging modalities also decreases exposure to radiation.

There are some possibilities, hardware-based or software-based, to improve the imaging speed. One way is to speed up the scanning process by acquiring fewer measurements. A recently developed mathematical framework called compressed sensing shows that it is possible to accurately recover undersampled images provided a suitable measurement matrix is used and the image itself has a sparse representation.

Nevertheless, not only measurements are important but also good reconstruction models are required. Such models are usually expressed as optimization problems.

In this thesis, we concentrated on the reconstruction of the undersampled Magnetic Resonance (MR) images. For this purpose a complex valued reconstruction model was provided. Since the reconstruction should be as quick as possible, fast methods to find the solution for the reconstruction problem are required. To meet this objective, three popular algorithms *FISTA, Augmented Lagrangian* and *Non-linear Conjugate Gradient* were adopted to work with our model.

By changing the complex-valued reconstruction model slightly and dualizing the problem, we obtained an instance of the quadratically constrained quadratic program where both the objective function and the constraints are twice differentiable. Hence new model opened doors to two other methods, the first order method which resembles *FISTA* and is called in this thesis *Normed Constrained Quadratic FGP*, and the second order method called *Truncated Newton Primal Dual Interior Point*.

Next, in order to compare performance of the methods, we set up the experiments and evaluated all presented methods against the problem of reconstructing undersampled MR images. In the experiments we used a number of invocations of the Fourier transform to measure the performance of all algorithms.

As a result of the experiments we found that in the context of the original model the performance of Augmented *Lagrangian* is better than the other two methods. Performance of *Non-linear Conjugate Gradient* and *FISTA* are about the same. In the context of the extended model Normed Constrained Quadratic FGP beats the *Truncated Newton Primal Dual Interior Point* method.

Shay MORAN

NATIONALITY: Israeli
DEPARTMENT: Algorithms and Complexity

THESIS TITLE:

Shattering Extremal Systems

ABSTRACT OF MASTER'S THESIS:

The Shatters relation and the VC dimension have been investigated since the early seventies. These concepts have found numerous applications in statistics, combinatorics, learning theory and computational geometry. Shattering extremal systems are set-systems with a very rich structure and many different characterizations. The goal of this thesis is to elaborate on the structure of these systems.

Dat Ba NGUYEN

NATIONALITY: **Vietnamese**
DEPARTMENT: **Databases and Information Systems**

THESIS TITLE:

Efficient Entity Disambiguation via Similarity Hashing

ABSTRACT OF MASTER'S THESIS:

The task of Named Entity Disambiguation (NED), which maps mentions of ambiguous names in natural language onto a set of known entities, has been an important issue in many areas including machine translation and information extraction. Working with a huge amount of data (e.g. more than three million entities in YAGO), some parts in an NED system which estimate the probability of a mention matching an entity, the similarity between a mention and an entity and the coherence among entity candidates for all mentions together might become bottlenecks. Thus, it is challenging for an interactive NED system to reach not only high accuracy but also efficiency.

This thesis presents an efficient way of disambiguating named entities by similarity hashing. Our framework is integrated with AIDA which is an on-line tool for entity detection and disambiguation developed at Max-Planck Institute for Informatics. We apply various state-of-the-art approaches, for example Locality Sensitive Hashing (LSH) and Spectral Hashing, to some forms of similarity search problem such as near-duplicate search for mention-entity matching, and especially related pair detection for entity-entity mapping which is not the default application of using hashing techniques due to the usually low similarities between entities.

Justin PARKS

NATIONALITY: USA
DEPARTMENT: Computer Graphics

THESIS TITLE:

Detecting Structural Regularity in Perspective Images

ABSTRACT OF MASTER'S THESIS:

The presence of structural regularity in a scene is a powerful source of information that reveals advantageous geometric and visual properties of an image, yet is notably difficult to accurately and fully recover. This work focuses on lattices, for which, despite many years of interest, current state of the art solutions continue to achieve less than optimal accuracy.

Much fruitful effort has gone into the contrivance of an automated system for detecting lattices, focusing on the system level at the expense of inquiry into the full recovery of the lattice. In fact, many systems use the same mechanism, identifying a single median motif for the lattice and evaluating its correspondence with the rest of the image by normalized cross-correlation. We recognize the tempered attention in this aspect of the process as an opportunity for improvement and identify two critical aspects in want of consideration: the method of choosing the motif and the technique for evaluating correspondence with the motif.

We develop a fully automated system for robustly detecting planar lattices in a perspective image as a framework for experimentation with motif selection and cell-correspondence evaluation. This system exemplifies a successful method for detecting lattices, recovering both the cell motif and the group operations that define the geometric displacement of the cells without any prior knowledge about the image or even the type of lattice that might be present.

Our primary contributions are the development of several novel lattice completion techniques and a structured evaluation of these and extant methods

resulting in revelations on this aspect of lattice detection. We devise five motif methods and several variants thereof, four of which we categorize as globally optimizing methods, and the last we term locally adaptive. In terms of cell-to-cell correspondence, we evaluate the commonly used SSD and normalized cross-correlation methods. We additionally test a mutual information technique of our own design and a composition of this method with normalized cross-correlation.

The striking result of our evaluation is twofold. Firstly, if a system utilizes a global motif method, it is overwhelmingly critical to choose the correct cell correspondence method, and that method is likely to be *mutual information*. Secondly, the advantage of using any variant of the local methods over any other local method is nearly imperceptible in contrast with the advantage of simply picking a local method versus a global method. We conclude that optimum results in any situation will be best and most easily obtained with a combination of mutual information and a motif model that adapts to the changing local environment of candidate cells.

Ebad PIR MORADIAN

NATIONALITY: Iranian
DEPARTMENT: Computer Graphics

THESIS TITLE:

Monocular Performance Capture via ToF Camera

ABSTRACT OF MASTER'S THESIS:

Performance capture is getting more and more importance among researchers. The main reason is its wide application range. it has applications from the wealthy entertainment industry to the robotics and from sports to surgery. It enables the movie makers to transfer every complex movement or body gesture to the digital models. Robots can use it to detect objects and navigate. Sport equipments could be designed simpler using this technology. Though the marker-based performance capture methods with their advantages and disadvantages are the dominant methods in the movie industry, markerless methods are also under great research aiming at solving the problems and limitations of marker-based methods. Parallel to advances in performance capture methods, in recent years depth cameras based on ToF principle has improved a lot. These cameras measures depth using infra-red signal enabling us having real-time 2.5D depth data with a tolerable noise level. This thesis is aimed to combine the Marker-less performance capture with the ToF technology and to verify its suitability for this application.

Natalia PRYTKOVA

NATIONALITY: Russian

DEPARTMENT: Databases and Information Systems

THESIS TITLE:

Modelling and Evaluation of Co-Evolution in Collective Web Memories

ABSTRACT OF MASTER'S THESIS:

The constantly evolving Web reflects the evolution of society in the cyberspace. Projects like the Open Directory Project (dmoz.org) can be understood as a collective memory of society on the Web. The main assumption is that such collective Web memories evolve when a certain cognition level about a concept has been exceeded. In the scope of our work we analyse the New York Times archive for concept detection. There are several approaches to the concept modelling. We introduce an alternative model for concepts, which does not make any additional assumptions about types of contained entities or the number of entities in the corpus. Moreover, the proposed distributed concept computation algorithm enables the large scale archive analysis. We also introduce a model of cognition level and explain how it can be employed to predict changes in the category system of DMOZ.

Bahjat SALIBA

NATIONALITY: Syrian
DEPARTMENT: Artificial Intelligence

THESIS TITLE:

An Evaluation Method For Indoor Positioning Systems On The Example Of LORIOT

ABSTRACT OF MASTER'S THESIS:

In this thesis an evaluation method on indoor positioning system called LORIOT is presented. This positioning system combines two technologies (RFID and IR) for positioning depending on Geo-referenced dynamic bayesian networks. LORIOT allows the users to calculate their position on their own device without sending any data to a server responsible for calculating the position. This property provides less complexity and fast calculation. This positioning method is developed by placing the tags in the environment and letting the user carry the sensors that are used to read data from these tags. The user is then able to choose either to pass the positioning data to any third party application or not. The main focus here is to check the actual accuracy and performance of indoor positioning systems using the proposed evaluation method which is tested on LORIOT. Most of the evaluation methods that have been used to test the level of accuracy of indoor positioning systems are biased and not good enough. For instance, the system is tested under optimal conditions of the environment. To achieve this goal, the evaluation method will be used to test LORIOT in a natural environment and by using data of natural traces of people walking in the environment without giving them any task to do. This type of evaluation criteria improves the results because the system would be installed in an environment which has the same properties that the environment has in this study, (where the evaluation tests are done). In addition, the system will position people while walking naturally (unlike most evaluation methods which test indoor positioning systems not while walking).

Martin SIMONOVSKY

NATIONALITY: Czech
DEPARTMENT: Computer Graphics

THESIS TITLE:

Hand Shape Recognition Using a ToF Camera: An Application to Sign Language

ABSTRACT OF MASTER'S THESIS:

This master's thesis investigates the benefit of utilizing depth information acquired by a time-of-flight (ToF) camera for hand shape recognition from unrestricted viewpoints. Specifically, we assess the hypothesis that classical 3D content descriptors might be inappropriate for ToF depth images due to the 2.5D nature and noisiness of the data and possible expensive computations in 3D space. Instead, we extend 2D descriptors to make use of the additional semantics of depth images. Our system is based on the appearance based retrieval paradigm, using a synthetic 3D hand model to generate its database. The system is able to run at interactive frame rates. For increased robustness, no color, intensity, or time coherence information is used. A novel, domain-specific algorithm for segmenting the forearm from the upper body based on reprojecting the acquired geometry into the lateral view is introduced. Moreover, three kinds of descriptors exploiting depth data are proposed and the made design choices are experimentally supported. The whole system is then evaluated on an American sign language fingerspelling dataset. However, the retrieval performance still leaves room for improvements. Several insights and possible reasons are discussed.

Niket TANDON

NATIONALITY: Indian

DEPARTMENT: Databases and Information Systems

THESIS TITLE:

Deriving a Web-Scale Common Sense Fact Knowledge Base

ABSTRACT OF MASTER'S THESIS:

The fact that birds have feathers and ice is cold seems trivially true. Yet, most machine-readablesources of knowledge either lack such common sense facts entirely or have only limited coverage. Prior work on automated knowledge base construction has largely focused on relations between named entities and on taxonomic knowledge, while disregarding common sense properties. Extracting such structured data from text is challenging, especially due to the scarcity of explicitly expressed knowledge. Even when relying on large document collections, patternbased information extraction approaches typically discover insufficient amounts of information.

This thesis investigates harvesting massive amounts of common sense knowledge using the textual knowledge of the entire Web, yet staying away from the massive engineering efforts in procuring such a large corpus as a Web. Despite the advancements in knowledge harvesting, we observed that the state of the art methods were limited in terms of accuracy and discovered insufficient amounts of information under our desired setting.

This thesis shows how to gather large amounts of common sense facts from Web N-gram data, using seeds from the existing knowledge bases like ConceptNet. Our novel contributions include scalable methods for tapping onto Web-scale data and a new scoring model to determine which patterns and facts are most reliable.

An extensive experimental evaluation is provided for three different binary relations, comparing different sources of n-gram data as well as different algorithms. The experimental results show that this approach extends ConceptNet by many orders of magnitude (more than 200-fold) at comparable levels of precision.

Christina TEFLIOUDI

NATIONALITY: **Greek**

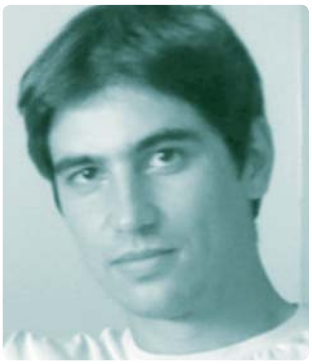
DEPARTMENT: **Databases and Information Systems**

THESIS TITLE:

Learning Soft Inference Rules in Large and Uncertain Knowledge Bases

ABSTRACT OF MASTER'S THESIS:

Recent progress in information extraction has enabled us to create large semantic knowledge bases with millions of RDF facts extracted from the Web. Nevertheless, the resulting knowledge bases are still incomplete or might contain inconsistencies, either because of the heuristic nature of the extraction process, or due to the varying reliability of the Web sources from which they were collected. One possible way of resolving both issues is to reinforce the knowledge base with deductive power by appending first-order logical inference rules, which help to describe and to further constrain the domain with which the ontology deals. In our work, we investigate learning these rules directly from the data using Inductive Logic Programming (ILP), a well-known technique, which lies in the intersection of machine learning and logic. Although powerful, ILP is inherently expensive as there is a combinatorial growth of the search space when constructing these rules, as the size of the background knowledge grows. In addition, the evaluation of each rule becomes more expensive, as the number of the training examples is rising. Apart from that, it is not always obvious how to automatically select positive and negative training examples needed for learning new rules over an incomplete knowledge base. This Master thesis explores the issues involved when applying ILP in an incomplete and large RDF-knowledge base.



Liviu TERIS

NATIONALITY: **Romanian**
DEPARTMENT: **Information Security and Cryptography**

THESIS TITLE:

Securing User-data in Android A Conceptual Approach for Consumer and Enterprise Usage

ABSTRACT OF MASTER'S THESIS:

Nowadays, smartphones and tablets are replacing the personal computer for the average user. As more activities move to these gadgets, so does the sensitive data with which they operate. However, there are few data protection mechanisms for the mobile world at the moment, especially for scenarios where the attacker has full access to the device (e.g. when the device is lost or stolen). In this thesis, we tackle this problem and propose a novel encryption system for Android, the top-selling mobile operating system.

Our investigation of the Android platform leads to a set of observations that motivate our effort. Firstly, the existing defense mechanisms are too weak or too rigid in terms of access control and granularity of the secured data unit. Secondly, Android can be corrupted such that the default encryption solution will reveal sensitive content via the debug interface. In response, we design and (partially) implement an encryption system that addresses these shortcomings and operates in a manner that is transparent to the user. Also, by leveraging hardware security mechanisms, our system offers security guarantees even when running on a corrupted OS. Moreover, the system is conceptually designed to operate in an enterprise environment where mobile devices are administered by a central authority. Finally, we provide a prototypical implementation and evaluate our system to show the practicality of our approach.

Anh TUAN TRAN

NATIONALITY: **Vietnamese**
DEPARTMENT: **Databases and Information Systems**

THESIS TITLE:

Context-Aware Timeline for Entity Exploration

ABSTRACT OF MASTER'S THESIS:

With millions of articles in multiple languages, Wikipedia has become the de-facto source of reference on the Internet today. Each article on Wikipedia contains encyclopedic information about various topics (people, events, inventions, etc.) and implicitly represents an entity. Extracting the most important facts about such entity will help users to find desired information more quickly and effectively. However, this task is challenging due to the incomplete and noisy nature of Wikipedia articles. This calls for a mechanism to detect and summarize the most important information about an entity on Wikipedia.

This thesis proposes and implements CATE (Context-Aware Timeline for Entity Exploration), a framework that utilizes Wikipedia to summarize and visualize the important aspects of entities in a timeline fashion. Such a system will help users to draw quickly an informative picture of an entity (e.g. life of a person, or evolution of a research topic, etc.). The novelty of CATE lies in seeing the entity in different contexts, synchronous with contemporaneous events. In addition, CATE puts the entity in a relationship with other entities, and thus offers a broader portrait about it. In order to efficiently query and visualize the events related to the entity, a number of techniques have been developed, combining information extraction and information retrieval with a novel ranking model. The thesis also discusses several experiments and evaluation results to show the effectiveness of the methods proposed.

Mahendiran VENKATACHALAPATHY

NATIONALITY: Indian

DEPARTMENT: Information Systems

THESIS TITLE:

Scheduling Strategies in a Main-Memory MapReduce Framework Approach for Countering Reduce Side Skew

ABSTRACT OF MASTER'S THESIS:

Over the past few decades, there is a multifold increase in the amount of digital data that is being generated. Various attempts are being made to process this vast amount of data in a fast and efficient manner. Hadoop-MapReduce is one such software framework that has gained popularity in the last few years. It provides a reliable and easier way to process huge amount of data in-parallel on large computing cluster. However, Hadoop always persists intermediate results to the local disk. As a result, Hadoop usually suffers from long execution runtimes as it typically pays a high I/O cost for running jobs.

The state-of-the-art computing clusters have enough main memory capacity to hold terabytes of data in main memory. We have built M3R (Main Memory MapReduce) framework, a prototype for generic main memory-based data processing. M3R can execute MapReduce jobs and also in addition it can execute general data processing jobs.

This master thesis in particular, focuses on countering the data-skewness problem for MapReduce jobs on M3R. Intermediate data following skewed distribution could lead to computational imbalance amongst the reduce tasks, resulting in longer MapReduce job execution times. This provides a scope for rebalancing the intermediate data and thereby reducing the total job runtimes.

We propose a novel dynamic approach of data rebalancing, to counter the reducer side data skewness. Our proposed on-the-fly skew countering approach,

attempts to detect the level of skewness in the intermediate data and rebalances the intermediate data amongst the reduce tasks. The proposed mechanism performs all the skew-countering related activities during the execution of actual MapReduce job. We have implemented this reduce side skew countering mechanism as a part of the M3R framework. The experiments conducted to study the behavior of this M3R data-rebalancing approach shows there is a significant reduction in the map-reduce job runtimes. In case of the data-skewed input, our proposed skew-control approach for M3R has reduced the total map-reduce job runtime (up to 31 %) when compared to M3R without skew-control.

Qinqing ZHENG

NATIONALITY: Chinese
DEPARTMENT: Machine Learning

THESIS TITLE:

Sparse Dictionary Learning with Simplex Constraints and Application to Topic Modeling

ABSTRACT OF MASTER'S THESIS:

Probabilistic mixture model is a powerful tool to provide a low-dimensional representation of count data. In the context of topic modeling, this amounts to representing the distribution of one document as a mixture of multiple distributions known as topics. The mixing proportions are called coefficients. A common attempt is to introduce sparsity into both the topics and the coefficients for better interpretability. We first discuss the problem of recovering sparse coefficients of given documents when the topics are known. This is formulated as a penalized least squares problem on the probability simplex, where the sparsity is achieved through regularization. However, the typical ℓ_1 regularizer becomes toothless in this case since it is constant over the simplex. To overcome this issue, we propose a group of concave penalties for inducing sparsity. An alternative approach is to post-process the solution of non-negative lasso to produce result that conform to the simplex constraint. Our experiments show that both kinds of approaches can effectively recover the sparsity pattern of coefficients. We then elaborately compare their robustness for different characteristics of input data. The second problem we discuss is to model both the topics and the coefficients of a collection of documents via matrix factorization. We propose the LpT approach, in which all the topics and coefficients are constrained on the simplex, and the ℓ_p penalty is imposed on each topic to promote sparsity. We also consider procedures that post-process the solutions of other methods. For example, the L1 approach first solves the problem where the simplex constraints imposed on the topics are relaxed into the non-negativity constraints, and the ℓ_p penalty is the replaced by the ℓ_1 penalty. Afterwards, L1 normalize the estimated topics to generate results satisfying the simplex

constraints. As detecting the number of mixture components inherent in the data is of central importance for the probabilistic mixture model, we analyze how the regularization techniques can help us to automatically find out this number. We compare the capabilities of these approaches to recover the low-rank structure underlying the data, when the number of topics are correctly specified and over-specified, respectively. The empirical results demonstrate that LpT and L1 can discover the sparsity pattern of the ground truth. In addition, when the number of topics is over-specified, they adapt to the true number of topics.

Alumni: PhD

Dr. Rawia AWADALLAH

NATIONALITY: **Palestinian**
DEPARTMENT: **Databases and Information Systems**

DISSERTATION TITLE:

Methods for Constructing an Opinion Network for Politically Controversial Topics

ABSTRACT OF PHD THESIS:

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

Dr. Andreas BAAK

NATIONALITY: German
DEPARTMENT: Computer Graphics

DISSERTATION TITLE:

Retrieval-based Approaches for Tracking and Reconstructing Human Motions

ABSTRACT OF PHD THESIS: :

Tracking, reconstructing, and analyzing human motions constitute central topics in computer vision and computer graphics. Although marker-less motion tracking has been an active research field for more than two decades, there are still major challenges, in particular when dealing with only few cameras, noise in the image data, occlusions, or fast motions. In this thesis, we introduce novel approaches for increasing the stability, accuracy, and efficiency of marker-less human motion tracking and 3D human pose reconstruction. As one common underlying concept, the presented approaches contain a retrieval component making use of database knowledge in the form of previously recorded marker-based motion capture (mocap) data. In particular, we contribute to three different areas dealing with various types of sensors including video cameras, optical mocap systems, inertial sensors, and depth cameras. Firstly, we introduce content-based retrieval techniques for automatically segmenting and annotating mocap data that is originally provided in form of unstructured data collections. Secondly, we show how such robust annotation procedures can be used to support and stabilize marker-less motion tracking. Thirdly, we develop algorithms for reconstructing human motions from noisy depth sensor data in real-time. In all these contributions, a particular focus is put on efficiency issues in order to keep the run time as low as possible.

Dr. Jasmina BOGOJESKA

NATIONALITY: Macedonian
DEPARTMENT: Computational Biology and Applied Algorithmics

DISSERTATION TITLE:

Statistical Learning Methods for Bias-aware HIV Therapy Screening

ABSTRACT OF PHD THESIS:

The human immunodeficiency virus (HIV) is the causative agent of the acquired immunodeficiency syndrome (AIDS) which claimed nearly 30 million lives and is arguably among the worst plagues in human history. With no cure or vaccine in sight, HIV patients are treated by administration of combinations of antiretroviral drugs. The very large number of such combinations makes the manual search for an effective therapy practically impossible, especially in advanced stages of the disease. Therapy selection can be supported by statistical methods that predict the outcomes of candidate therapies. However, these methods are based on clinical data sets that are biased in many ways. The main sources of bias are the evolving trends of treating HIV patients, the sparse, uneven therapy representation, the different treatment backgrounds of the clinical samples and the differing abundances of the various therapy-experience levels.

In this thesis we focus on the problem of devising bias-aware statistical learning methods for HIV therapy screening – predicting the effectiveness of HIV combination therapies. For this purpose we develop five novel approaches that when predicting outcomes of HIV therapies address the aforementioned biases in the clinical data sets. Three of the approaches aim for good prediction performance for every drug combination independent of its abundance in the HIV clinical data set. To achieve this, they balance the sparse and uneven therapy representation by using different routes of sharing common knowledge among related therapies. The remaining two approaches additionally account for the bias originating from the differing treatment histories of the samples making up the HIV clinical data sets. For this purpose, both methods predict

the response of an HIV combination therapy by taking not only the most recent (target) therapy but also available information from preceding therapies into account. In this way they provide good predictions for advanced patients in mid to late stages of HIV treatment, and for rare drug combinations.

All our methods use the time-oriented evaluation scenario, where models are trained on data from the less recent past while their performance is evaluated on data from the more recent past. This is the approach we adopt to account for the evolving treatment trends in the HIV clinical practice and thus offer a realistic model assessment.

Dr. Martin BOKELOH

NATIONALITY: German
DEPARTMENT: Computer Graphics

DISSERTATION TITLE:

Symmetry in 3D Shapes - Analysis and Applications to Model Synthesis

ABSTRACT OF PHD THESIS:

Symmetry is an essential property of a shapes' appearance and presents a source of information for structure-aware deformation and model synthesis. This thesis proposes feature-based methods to detect symmetry and regularity in 3D shapes and demonstrates the utilization of symmetry information for content generation. First, we will introduce two novel feature detection techniques that extract salient keypoints and feature lines for a 3D shape respectively. Further, we will propose a randomized, feature-based approach to detect symmetries and decompose the shape into recurring building blocks. Then, we will present the concept of docking sites that allows us to derive a set of shape operations from an exemplar and will produce similar shapes. This is a key insight of this thesis and opens up a new perspective on inverse procedural modeling. Finally, we will present an interactive, structure-aware deformation technique based entirely on regular patterns.

Dr. Katarzyna BOZEK

NATIONALITY: Polish
DEPARTMENT: Computational Biology and Applied Algorithmics

DISSERTATION TITLE:

Analysis of HIV-host Interaction on Different Scales

ABSTRACT OF PHD THESIS:

The human immunodeficiency virus depends on molecular pathways of the host for efficient replication and spread. The intricate network of host-virus interactions shapes the virus' evolution by driving the pathogen to evade immune recognition and constraining it to maintain its capacity to replicate. Study of the HIV-host interactions provides important insights into viral evolution, pathogenicity and potential treatment strategies. This thesis presents an analysis of HIV-host interactions on several scales, ranging from individual protein interactions to whole genomes.

On the scale of individual interaction we analyze structural and physical determinants of the interaction between host TRIM5α and virus capsid – an interaction of potential therapeutic interest due to the capacity of TRIM5α to block retroviral infections. On the scale of viral population we present two studies of a highly variable region of the virus genome involved in the interaction with host cell coreceptors upon virus cell entry. The studies provide insights into the virus evolution and the physicochemical and structural properties related to its interaction with cellular coreceptors. On the scale of the single cell we develop models of HIV cell entry involving virus, host and environmental factors. The models represent a comprehensive picture of the virus phenotype and allow one to view the variability of virus phenotypes on 2D phenotype maps. On the genomic scale we perform a large-scale analysis of all HIV-host interactions. This study reveals insights into general patterns of the host-pathogen evolution and suggests candidate host proteins involved in interactions potentially important for the infection and interesting for further study on other scales.

Interactions and processes crucial for the HIV infection reemerge across the scales pointing to the importance of integrative, multi-scale studies of host-pathogen biology.

Dr. Andreas BROSCHART

NATIONALITY: German

DEPARTMENT: Efficient Search in Semistructured Data Spaces

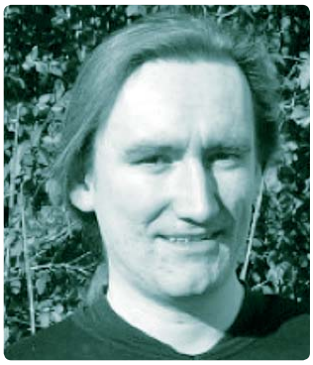
DISSERTATION TITLE:

Efficient Query Processing and Index Tuning using Proximity Scores

ABSTRACT OF PHD THESIS:

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

- We present a comprehensive comparative analysis of proximity score models and a rigorous analysis of the potential of phrases and adapt a leading proximity score model for XML data.
- We discuss the feasibility of all presented proximity score models for top-k query processing and present a novel index combining a content and proximity score that helps to accelerate top-k query processing and improves result quality.
- We present a novel, distributed index tuning framework for term and term pair index lists that optimizes pruning parameters by means of well-defined optimization criteria under disk space constraints. Indexes can be tuned with emphasis on efficiency or effectiveness: the resulting indexes yield fast processing at high result quality.
- We show that pruned index lists processed with a merge join outperform top-k query processing with unpruned lists at a high result quality.
- Moreover, we present a hybrid index structure for improved cold cache run times.



Dr. Tom CRECELIUS

NATIONALITY: German
DEPARTMENT: Databases and Information Systems

DISSERTATION TITLE:

Socially Enhanced Search and Exploration in Social Tagging Networks

ABSTRACT OF PHD THESIS:

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

We address three research areas in our work utilising these intrinsic features of social tagging networks.

- We investigate new approaches for exploiting the social knowledge of and the relationships between users for searching and recommending relevant contents, and integrate them in a comprehensive framework, coined SENSE, for search in social tagging networks.
- To dynamically update precomputed lists of transitive friends in descending order of their distance in user graphs of social tagging networks, we provide an algorithm for incrementally solving the all pairs shortest distance problem in large, disk-resident graphs and formally prove its correctness.
- Since users are content providers in social tagging networks, users may keep their own data at independent, local peers that collaborate in a distributed P2P network. We provide an algorithm for such systems to counter cheating of peers in authority computations over social networks.

The viability of each solution is demonstrated by extensive experiments regarding effectiveness and efficiency.

Dr. Dimitar DENEV

NATIONALITY: Romanian
DEPARTMENT: Databases and Information Systems

DISSERTATION TITLE:

Models and Methods for Web Archive Crawling

ABSTRACT OF PHD THESIS:

Web archives offer a rich and plentiful source of information to researchers, analysts, and legal experts. For this purpose, they gather Web sites as the sites change over time. In order to keep up to high standards of data quality, Web archives have to collect all versions of the Web sites. Due to limited resources and technical constraints this is not possible. Therefore, Web archives consist of versions archived at various time points without guarantee for mutual consistency.

This thesis presents a model for assessing the data quality in Web archives as well as a family of crawling strategies yielding high-quality captures. We distinguish between single-visit crawling strategies for exploratory and visit-revisit crawling strategies for evidentiary purposes. Single-visit strategies download every page exactly once aiming for an “undistorted” capture of the ever-changing Web. We express the quality of such the resulting capture with the “blur” quality measure. In contrast, visit-revisit strategies download every page twice. The initial downloads of all pages form the visit phase of the crawling strategy. The second downloads are grouped together in the revisit phase. These two phases enable us to check which pages changed during the crawling process. Thus, we can identify the pages that are consistent with each other. The quality of the visit-revisit captures is expressed by the “coherence” measure.

Quality-conscious strategies are based on predictions of the change behaviour of individual pages. We model the Web site dynamics by Poisson processes with page specific change rates. Furthermore, we show that these rates can be statistically predicted. Finally, we propose visualization techniques for exploring the quality of the resulting Web archives.

A fully functional prototype demonstrates the practical viability of our approach.

Dr. Piotr DIDYK

NATIONALITY: **Polish**
DEPARTMENT: **Computer Graphics**

DISSERTATION TITLE:

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

ABSTRACT OF PHD THESIS:

Existing displays have a number of limitations, which make it difficult to realistically reproduce real-world appearance; discrete pixels are used to represent images, which are refreshed only a limited number of times per second, the output luminance range is much smaller than in the real world, and only two dimensions are available to reproduce a three-dimensional scene.

While in some cases technology advanced and higher frame rates, higher resolution, higher luminance, and even disparity-based stereo is possible, these solutions are often costly and, further, it is challenging to produce adequate content.

On the other hand, the human visual system has certain limitations itself, such as the density of photoreceptors, imperfections in the eye optics, or the limited ability to discern high-frequency information. The methods presented in this dissertation show that taking these properties into account can improve the efficiency and perceived quality of displayed imagery. More precisely, those techniques make use of perceptual effects, which are not measurable physically, that will allow us to overcome the physical limitations of display devices in order to enhance apparent image qualities.

Dr. Laura DIETZ

NATIONALITY: German
DEPARTMENT: Databases and Information Systems

DISSERTATION TITLE:

Exploiting Graph-Structured Data in Generative Probabilistic Models

ABSTRACT OF PHD THESIS:

Unsupervised machine learning aims to make predictions when labeled data is absent, and thus, supervised machine learning cannot be applied. These algorithms build on assumptions about how data and predictions relate to each other. One technique for unsupervised problem settings are generative models, which specify the set of assumptions as a probabilistic process that generates the data.

The subject of this thesis is how to most effectively exploit input data that has an underlying graph structure in unsupervised learning for three important use cases. The first use case deals with localizing defective code regions in software, given the execution graph of code lines and transitions. Citation networks are exploited in the next use case to quantify the influence of citations on the content of the citing publication. In the final use case, shared tastes of friends in a social network are identified, enabling the prediction of items from a user a particular friend of his would be interested in.

For each use case, prediction performance is evaluated via held-out test data that is only scarcely available in the domain. This comparison quantifies under which circumstances each generative model best exploits the given graph structure.

Dr. Zhao DONG

NATIONALITY: Chinese
DEPARTMENT: Computer Graphics

DISSERTATION TITLE:

Visually Pleasing Real-time Global Illumination Rendering for Fully-dynamic Scenes

ABSTRACT OF PHD THESIS:

Global illumination (GI) rendering plays a crucial role in the photo-realistic rendering of virtual scenes. With the rapid development of graphics hardware, GI has become increasingly attractive even for real-time applications nowadays. However, the computation of physically-correct global illumination is time consuming and cannot achieve real-time, or even interactive performance. Although the realtime GI is possible using a solution based on precomputation, such a solution cannot deal with fully-dynamic scenes.

This dissertation focuses on solving these problems by introducing visually pleasing real-time global illumination rendering for fully-dynamic scenes. To this end, we develop a set of novel algorithms and techniques for rendering global illumination effects using the graphics hardware. All these algorithms not only result in real-time or interactive performance, but also generate comparable quality to the previous works in off-line rendering. First, we present a novel implicit visibility technique to circumvent expensive visibility queries in hierarchical radiosity by evaluating the visibility implicitly. Thereafter, we focus on rendering visually plausible soft shadows, which is the most important GI effect caused by the visibility determination. Based on the pre-filtering shadow mapping theory, we successively propose two real-time soft shadow mapping methods: “convolution soft shadow mapping” (CSSM) and “variance soft shadow mapping” (VSSM). Furthermore, we successfully apply our CSSM method in computing the shadow effects for indirect lighting. Finally, to explore the GI rendering in participating media, we investigate a novel technique to interactively render volume caustics in the single-scattering participating media.

Dr. Sandra EBERT

NATIONALITY: German
DEPARTMENT: Computer Vision and Multimodal Computing

DISSERTATION TITLE:

Semi-Supervised Learning for Image Classification

ABSTRACT OF PHD THESIS:

Object class recognition is an active topic in computer vision still presenting many challenges. In most approaches, this task is addressed by supervised learning algorithms that need a large quantity of labels to perform well. This leads either to small datasets ($< 10,000$ images) that capture only a subset of the real-world class distribution (but with a controlled and verified labeling procedure), or to large datasets that are more representative but also add more label noise. Therefore, semi-supervised learning is a promising direction. It requires only few labels while simultaneously making use of the vast amount of images available today. We address object class recognition with semi-supervised learning. These algorithms depend on the underlying structure given by the data, the image description, and the similarity measure, and the quality of the labels. This insight leads to the main research questions of this thesis: “Is the structure given by labeled and unlabeled data more important than the algorithm itself?”, “Can we improve this neighborhood structure by a better similarity metric or with more representative unlabeled data?”, and “Is there a connection between the quality of labels and the overall performance and how can we get more representative labels?”. We answer all these questions, i.e., we provide an extensive evaluation, we propose several graph improvements, and we introduce a novel active learning framework to get more representative labels.

Dr. Shady ELBASSUONI

NATIONALITY: Egyptian
DEPARTMENT: Databases and Information Systems

DISSERTATION TITLE:

Effective Searching of RDF Knowledge Bases

ABSTRACT OF PHD THESIS:

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

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

Dr. Pavel EMELIYANENKO

NATIONALITY: Russian
DEPARTMENT: Algorithms and Complexity

DISSERTATION TITLE:

Harnessing the Power of GPUs for Problems in Real Algebraic Geometry

ABSTRACT OF PHD THESIS:

This thesis presents novel parallel algorithms to leverage the power of GPUs (Graphics Processing Units) for exact computations with polynomials having large integer coefficients. The significance of such computations, especially in real algebraic geometry, is hard to undermine. On massively-parallel architectures such as GPU, the degree of datalevel parallelism exposed by an algorithm is the main performance factor. We attain high effiency through the use of structured matrix theory to assist the realization of relevant operations on polynomials on the graphics hardware. A detailed complexity analysis, assuming the PRAM model, also confirms that our approach achieves a substantially better parallel complexity in comparison to classical algorithms used for symbolic computations.

Aside from the theoretical considerations, a large portion of this work is dedicated to the actual algorithm development and optimization techniques where we pay close attention to the specifics of the graphics hardware. As a byproduct of this work, we have developed high-throughput modular arithmetic which we expect to be useful for other GPU applications, in particular, open-key cryptography. We further discuss the algorithms for the solution of a system of polynomial equations, topology computation of algebraic curves and curve visualization which can profit to the full extent from the GPU acceleration. Extensive benchmarking on a real data demonstrates the superiority of our algorithms over several state-of-the-art approaches available to date.

Dr. Dorothea EMIG

NATIONALITY: German
DEPARTMENT: Computational Biology and Applied Algorithmics

DISSERTATION TITLE:

Novel Analysis Approaches to Context-Dependent Molecular Networks

ABSTRACT OF PHD THESIS:

Proteins are key players in all kinds of biological processes and accurate knowledge of their presence and their interactions is fundamental for understanding the functioning of the cells. Over the last years, many large-scale studies have been performed in order to unravel the complete human interactome. However, the results of these studies usually depend on the cellular conditions, in which the protein interactions were detected. Furthermore, additional biological mechanisms or temporal and spatial constraints contribute to the context-dependent formation of protein interactions.

In this thesis, we focus on different biological aspects that are important for the formation of protein-protein interactions. We first analyze protein interactions in a structural context and demonstrate that interacting proteins may collide in three-dimensional space, rendering the interaction impossible. Second, we investigate the tissue-specific formation of protein interactions. We analyze the ability of different technologies such as microarray platforms and next-generation RNA-sequencing to reliably detect tissue-specific gene expression. We further use gene expression data to identify tissue-specific protein interactions and their functional implications. Finally, we concentrate on protein variants that arise by alternative splicing events. We describe our software DomainGraph that allows for visually exploring protein variants and their interactions in different biological conditions.

Dr. Peter M. GROSCHE

NATIONALITY: German
DEPARTMENT: Computer Graphics

DISSERTATION TITLE:

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

ABSTRACT OF PHD THESIS:

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

Dr. David GÜNTHER

NATIONALITY: German
DEPARTMENT: Computer Graphics

DISSERTATION TITLE:

Topological Analysis of Discrete Scalar Data

ABSTRACT OF PHD THESIS:

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

Dr. Catalin HRITCU

NATIONALITY: Romanian
DEPARTMENT: Information Security and Cryptography

DISSERTATION TITLE:

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

ABSTRACT OF PHD THESIS:

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

Dr. Verena KONZ

NATIONALITY: German
DEPARTMENT: Computer Graphics

DISSERTATION TITLE:

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

ABSTRACT OF PHD THESIS:

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

Dr. Zhao LI

NATIONALITY: Chinese
DEPARTMENT: Telecommunications

DISSERTATION TITLE:

Multicast MAC Extensions for High Rate Real-Time Traffic in Wireless LANs

ABSTRACT OF PHD THESIS:

Nowadays we are rapidly moving from a mainly textual-based to a multimedia-based Internet, for which the widely deployed IEEE 802.11 wireless LANs can be one of the promising candidates to make them available to users anywhere, anytime, on any device. However, it is still a challenge to support group-oriented real-time multimedia services, such as video-on-demand, video conferencing, distance educations, mobile entertainment services, interactive games, etc., in wireless LANs, as the current protocols do not support multicast, in particular they just send multicast packets in open-loop as broadcast packets, i.e., without any possible acknowledgements or retransmissions. In this thesis, we focus on MAC layer reliable multicast approaches which outperform upper layer ones with both shorter delays and higher efficiencies. Different from polling based approaches, which suffer from long delays, low scalabilities and low efficiencies, we explore a feedback jamming mechanism where negative acknowledgement (NACK) frames are allowed from the non-leader receivers to destroy the acknowledgement (ACK) frame from the single leader receiver and prompts retransmissions from the sender. Based on the feedback jamming scheme, we propose two MAC layer multicast error correction protocols, SEQ driven Leader Based Protocol (SEQ-LBP) and Hybrid Leader Based Protocol (HLBP), the former is an Automatic Repeat reQuest (ARQ) scheme while the later combines both ARQ and the packet level Forward Error Correction (FEC). We evaluate the feedback jamming probabilities and the performances of SEQ-LBP and HLBP based on theoretical analyses, NS-2 simulations and experiments on a real test-bed built with consumer wireless LAN cards. Test results confirm the feasibility of the feedback jamming scheme and the outstanding performances of the proposed protocols SEQ-LBP and HLBP, in particular SEQ-LBP is good for small multicast groups due to its short delay, effectiveness and simplicity while HLBP is better for large multicast groups because of its high efficiency and high scalability with respect to the number of receivers per group.

Dr. Madhusudan MANJUNATH

NATIONALITY: Indian

DEPARTMENT: Algorithms and Complexity

DISSERTATION TITLE:

A Riemann-Roch Theory for Sublattices of the Root Lattice A_n, Graph Automorphisms and Counting Cycles in Graphs

ABSTRACT OF PHD THESIS:

This thesis consists of two independent parts. In the first part of the thesis, we develop a Riemann-Roch theory for sublattices of the root lattice A_n extending the work of Baker and Norine (Advances in Mathematics, 215(**2**): 766-788, 2007) and study questions that arise from this theory. Our theory is based on the study of critical points of a certain simplicial distance function on a lattice and establishes connections between the Riemann-Roch theory and the Voronoi diagrams of lattices under certain simplicial distance functions. In particular, we provide a new geometric approach for the study of the Laplacian of graphs. As a consequence, we obtain a geometric proof of the Riemann-Roch theorem for graphs and generalize the result to other sub-lattices of A_n. Furthermore, we use the geometric approach to study the problem of computing the rank of a divisor on a finite multigraph G to obtain an algorithm that runs in polynomial time for a fixed number of vertices, in particular with running time $2^{O(n \log n)}$ *poly (size (G))* where n is the number of vertices of G. Motivated by this theory, we study a dimensionality reduction approach to the graph automorphism problem and we also obtain an algorithm for the related problem of counting automorphisms of graphs that is based on exponential sums.

In the second part of the thesis, we develop an approach, based on complex-valued hash functions, to count cycles in graphs in the data streaming model. Our algorithm is based on the idea of computing instances of complex-valued random variables over the given stream and improves drastically upon the naïve sampling algorithm.

Dr. Yana M. MILEVA

NATIONALITY: **Bulgarian**
DEPARTMENT: **Software Engineering**

DISSERTATION TITLE:

Mining the Evolution of Software Component Usage

ABSTRACT OF PHD THESIS:

The topic of this thesis is the analysis of the evolution of software components. In order to track the evolution of software components, one needs to collect the evolution information of each component. This information is stored in the version control system (VCS) of the project – the repository of the history of events happening throughout the project's lifetime. By using software archive mining techniques one can extract and leverage this information.

The main contribution of this thesis is the introduction of *evolution usage trends and evolution change patterns*. The raw information about the occurrences of each component is stored in the VCS of the project. By organizing it in evolution trends and patterns, we are able to draw conclusions and issue recommendations concerning each individual component and the project as a whole.

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

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

Dr. Ndapandula T. NAKASHOLE

NATIONALITY: **Namibian**
DEPARTMENT: **Databases and Information Systems**

DISSERTATION TITLE:

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

ABSTRACT OF PHD THESIS:

Equipping machines with knowledge, through the construction of machine readable knowledge bases, presents a key asset for semantic search, machine translation, question answering, and other formidable challenges in artificial intelligence. However, human knowledge predominantly resides in books and other natural language text forms. This means that knowledge bases must be extracted and synthesized from natural language text. When the source of text is the Web, extraction methods must cope with ambiguity, noise, scale, and updates.

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

All methods have been evaluated through experiments involving Web-scale text collections.

Dr. Hans-Jörg PETER

NATIONALITY: German
DEPARTMENT: Reactive Systems

DISSERTATION TITLE:

A Uniform Approach to the Complexity and Analysis of Succinct Systems

ABSTRACT OF PHD THESIS:

This thesis provides a unifying view on the succinctness of systems: the capability of a modeling formalism to describe the behavior of a system of exponential size using a polynomial syntax.

The key theoretical contribution is the introduction of sequential circuit machines as a new universal computation model that focuses on succinctness as the central aspect. The thesis demonstrates that many well-known modeling formalisms such as communicating state machines, linear-time temporal logic, or timed automata exhibit an immediate connection to this machine model. Once a (syntactic) connection is established, many complexity bounds for structurally restricted sequential circuit machines can be transferred to a certain formalism in a uniform manner. As a consequence, besides a far-reaching unification of independent lines of research, we are also able to provide matching complexity bounds for various analysis problems, whose complexities were not known so far. For example, we establish matching lower and upper bounds of the small witness problem and several variants of the bounded synthesis problem for timed automata, a particularly important succinct modeling formalism.

Also for timed automata, our complexity-theoretic analysis leads to the identification of tractable fragments of the timed synthesis problem under partial observability. Specifically, we identify timed controller synthesis based on discrete or template-based controllers to be equivalent to model checking. Based on this discovery, we develop a new model checking-based algorithm to efficiently find feasible template instantiations.

From a more practical perspective, this thesis also studies the preservation of succinctness in analysis algorithms using symbolic data structures. While efficient techniques exist for specific forms of succinctness considered in isolation, we present a general approach based on abstraction refinement to combine of-the-shelf symbolic data structures. In particular, for handling the combination of concurrency and quantitative timing behavior in networks of timed automata, we report on the tool SYNTHIA which combines binary decision diagrams with difference bound matrices. In a comparison with the timed model checker UPPAAL and the timed game solver UPPAAL-TIGA running on standard benchmarks from the timed model checking and synthesis domain, respectively, the experimental results clearly demonstrate the effectiveness of our new approach.

Dr. Stefan POPOV

NATIONALITY: **Bulgarian**
DEPARTMENT: **Computer Graphics**

DISSERTATION TITLE:

Algorithms and Data Structures for Interactive Ray Tracing on Commodity Hardware

ABSTRACT OF PHD THESIS:

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

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

For the static portions of the scene, we also address the topic of creating the “perfect” acceleration structure. We develop a polynomial time non-greedy BVH construction algorithm. We then modify it to produce a new type of acceleration structure that inherits both the high performance of KD-trees and the small size of BVHs.

Finally, we focus on bringing real-time ray tracing to commodity desktop computers. We develop several new KD-tree and BVH traversal algorithms specifically tailored for the GPU. With them, we show for the first time that GPU ray tracing is indeed feasible, and it can outperform CPU ray tracing by almost an order of magnitude, even on large CAD models.

Dr. Imran RAUF

NATIONALITY: **Pakistani**
DEPARTMENT: **Algorithms and Complexity**

DISSERTATION TITLE:

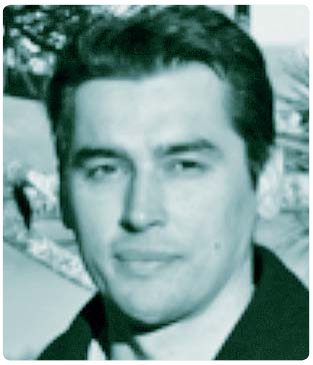
Polynomially Solvable Cases of Hypergraph Transversal and Related Problems

ABSTRACT OF PHD THESIS:

This thesis is mainly concerned with the *hypergraph transversal problem*, which asks to generate all minimal transversals of a given hypergraph. While the current best upper bound on the complexity of the problem is quasi-polynomial in the combined input and output sizes, it is shown to be solvable in output polynomial time for a number of hypergraph classes. We extend this polynomial frontier to the hypergraphs induced by hyperplanes and constant-sided polytopes in fixed dimension $\mathbb{R}^d$ and hypergraphs for which every minimal transversal and hyperedge intersection is bounded. We also show the problem to be fixed parameter tractable with respect to the minimum integer k such that the input hypergraph is k-degenerate, and also with respect to its maximum complementary degree. Whereas we improve the known bounds when the parameter is the maximum degree of a hypergraph.

We also study the *readability* of a monotone Boolean function which is defined as the minimum integer r such that it can be represented by a $\vee-\wedge-$formula with every variable occurrence is bounded by r. We prove that it is NP-hard to approximate the readability of even a depth three Boolean formula. We also give tight sublinear upper bounds on the readability of a monotone Boolean function given in CNF (or DNF) form, parameterized by the number of terms in the CNF and the maximum number of variables in the intersection of any constant number of terms. For interval DNF's we give much tighter logarithmic bounds on the readability.

Finally, we discuss an implementation of a quasi-polynomial algorithm for the hypergraph transversal problem that runs in polynomial space. We found our implementation to be competitive with all but one previous implementation on various datasets.



Dr. Art TEVS

NATIONALITY: German
DEPARTMENT: Computer Graphics

DISSERTATION TITLE:

Deformable Shape Matching

ABSTRACT OF PHD THESIS:

Deformable shape matching has become an important building block in academia as well as in industry. Given two three dimensional shapes A and B the deformation function f aligning A with B has to be found. The function is discretized by a set of corresponding point pairs. Unfortunately, the computation cost of a brute-force search of correspondences is exponential. Additionally, to be of any practical use the algorithm has to be able to deal with data coming directly from 3D scanner devices which suffers from acquisition problems like noise, holes as well as missing any information about topology.

This dissertation presents novel solutions for solving shape matching: First, an algorithm estimating correspondences using a randomized search strategy is shown. Additionally, a planning step dramatically reducing the matching costs is incorporated. Using ideas of these both contributions, a method for matching multiple shapes at once is shown. The method facilitates the reconstruction of shape and motion from noisy data acquired with dynamic 3D scanners. Considering shape matching from another perspective a solution is shown using Markov Random Fields (MRF). Formulated as MRF, partial as well as full matches of a shape can be found. Here, belief propagation is utilized for inference computation in the MRF. Finally, an approach significantly reducing the space-time complexity of belief propagation for a wide spectrum of computer vision tasks is presented.

Dr. Alexander THIELEN

NATIONALITY: German
DEPARTMENT: Computational Biology and Applied Algorithmics

DISSERTATION TITLE:

Genotypic Analysis of HIV-1 Coreceptor Usage

ABSTRACT OF PHD THESIS:

The acquired immunodeficiency syndrome (AIDS) is one of the biggest medical challenges in the world today. Its causative pathogen, the human immunodeficiency virus (HIV), is responsible for millions of deaths per year. Although about two dozen antiviral drugs are currently available, progression of the disease can only be delayed but patients cannot be cured.

In recent years, the new class of coreceptor antagonists has been added to the arsenal of antiretroviral drugs. These drugs block viral cell-entry by binding to one of the receptors the virus requires for infection of a cell. However, some HIV variants can also use another coreceptor so that coreceptor usage has to be tested before administration of the drug.

This thesis analyzes the use of statistical learning methods to infer HIV coreceptor usage from viral genotype. Improvements over existing methods are achieved by using sequence information of so far not used genomic regions, next generation sequencing technologies, and by combining different existing prediction systems.

In addition, HIV coreceptor usage prediction is analyzed with respect to clinical outcome in patients treated with coreceptor antagonists. The results demonstrate that inferring HIV coreceptor usage from viral genotype can be reliably used in daily routine.

Dr. Laura TOLOSI

NATIONALITY: Romanian
DEPARTMENT: Computational Biology and Applied Algorithmics

DISSERTATION TITLE:

Finding Regions of Aberrant DNA Copy Number Associated with Tumor Phenotype

ABSTRACT OF PHD THESIS:

DNA copy number alterations are a hallmark of cancer. Understanding their role in tumor progression can help improve diagnosis, prognosis and therapy selection for cancer patients and can contribute to the development of personalised therapies. High-resolution, genome-wide measurements of DNA copy number changes for large cohorts of tumors are currently available, owing to the rapid development of technologies like microarray-based array comparative hibridization (arrayCGH). In this manuscript, we introduce a computational pipeline for statistical analysis of tumor cohorts, which can help extract relevant patterns of copy number aberrations and infer their association with various phenotypical indicators. The pipeline makes use of machine learning techniques for classification and feature selection, with emphasis on interpretable models (linear models with penalties, tree-based models).

The main challenges that our methods meet are the high dimensionality of the arrays compared to the small number of tumor samples available, as well as the large correlations between copy number estimates measured at neighboring genomic locations. Consequently, feature selection is unstable, depending strongly on the set of training samples, leading to un-reproducible signatures across different clinical studies. We also show that the feature ranking given by several widely-used methods for feature selection is biased due to the large correlations between features. In order to correct for the bias and instability of the feature ranking, we introduce a dimension reduction step in our pipeline, consisting of multivariate segmentation of the set of arrays. We present three algorithms for multivariate segmentation, which are based on indentifying recurrent DNA breakpoints or DNA regions of constant copy number profile. The

multivariate segmentation constitutes the basis for computing a smaller set of super-features, by summarizing the DNA copy number within the segmentation regions. Using the super-features for supervised classification, we improve the interpretability and stability of the models, where the baseline for comparison consists of classification models trained on probe data.

We validated the methods by training models for prediction of the phenotype of breast cancers and neuroblastoma tumors. We show that the multivariate segmentation step affords higher model stability and it does not decrease the accuracy of the prediction. We obtain substantial dimension reduction (up to 200-fold less predictors), which recommends the multivariate segmentation procedures not only for the purpose of phenotype prediction, but also as pre-processing step for downstream integration with other data types.

The interpretability of the models is also improved, revealing important associations between copy number aberrations and phenotype. For example, we show that a very informative predictor that distinguishes between inflammatory and non-inflammatory breast cancers with ERBB2 amplification is the co-amplification of the genomic region located in the immediate vicinity of the ERBB2 gene locus. Therefore, we conclude that the size of the amplicon is associated with the cancer subtype, a hypothesis present elsewhere in the literature. In the case of neuroblastoma tumors, we show that patients belonging to different age subgroups are characterized by distinct copy number patterns, especially when the subgroups are defined as older or younger than 16-18 months. Indeed, considering a large set of age cutoffs, our prediction models are most accurate if the cutoff is around 16 – 18 months. We thereby confirm the recommendation for a higher age cutoff than 12 months (current clinical practice) for differential diagnosis of neuroblastoma.

Dr. Carola WINZEN

NATIONALITY: German
DEPARTMENT: Algorithms and Complexity

DISSERTATION TITLE:

Toward a Complexity Theory for Randomized Search Heuristics: Black-Box Models

ABSTRACT OF PHD THESIS:

Randomized search heuristics are a broadly used class of general-purpose algorithms. Analyzing them via classical methods of theoretical computer science is a growing field. While several strong runtime bounds exist, a powerful complexity theory for such algorithms is yet to be developed.

We contribute to this goal in several aspects. In a first step, we analyze existing black-box complexity models. Our results indicate that these models are not restrictive enough. This remains true if we restrict the memory of the algorithms under consideration.

These results motivate us to enrich the existing notions of black-box complexity by the additional restriction that not actual objective values, but only the relative quality of the previously evaluated solutions may be taken into account by the algorithms. Many heuristics belong to this class of algorithms. We show that our ranking-based model gives more realistic complexity estimates for some problems, while for others the low complexities of the previous models still hold.

Surprisingly, our results have an interesting game-theoretic aspect as well. We show that analyzing the black-box complexity of the ONEMAXn function class – a class often regarded to analyze how heuristics progress in easy parts of the search space – is the same as analyzing optimal winning strategies for the generalized Mastermind game with 2 colors and length-n codewords. This connection was seemingly overlooked so far in the search heuristics community.

Dr. Patrick WISCHNEWSKI

NATIONALITY: German
DEPARTMENT: Automation of Logics

DISSERTATION TITLE:

Efficient Reasoning Procedures for Complex First-Order Theories

ABSTRACT OF PHD THESIS:

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

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

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

Dr. Gernot ZIEGLER

NATIONALITY: **Austrian**
DEPARTMENT: **Computer Graphics**

DISSERTATION TITLE:

GPU Data Structures for Graphics and Vision

ABSTRACT OF PHD THESIS:

Graphics hardware has in recent years become increasingly programmable, and its programming APIs use the stream processor model to expose massive parallelization to the programmer. Unfortunately, the inherent restrictions of the stream processor model, used by the GPU in order to maintain high performance, often pose a problem in porting CPU algorithms for both video and volume processing to graphics hardware. Serial data dependencies which accelerate CPU processing are counterproductive for the data-parallel GPU.

This thesis demonstrates new ways for tackling well-known problems of large scale video/volume analysis. In some instances, we enable processing on the restricted hardware model by reintroducing algorithms from early computer graphics research. On other occasions, we use newly discovered, hierarchical data structures to circumvent the random-access read/fixed write restriction that had previously kept sophisticated analysis algorithms from running solely on graphics hardware. For 3D processing, we apply known game graphics concepts such as mip-maps, projective texturing, and dependent texture lookups to show how video/volume processing can benefit algorithmically from being implemented in a graphics API.

The novel GPU data structures provide drastically increased processing speed, and lift processing heavy operations to real-time performance levels, paving the way for new and interactive vision/graphics applications.

Dr. Henning Lars ZIMMER

NATIONALITY: German
DEPARTMENT: Mathematical Image Analysis

DISSERTATION TITLE:

Correspondence Problems in Computer Vision Novel Models, Numerics, and Applications

ABSTRACT OF PHD THESIS:

Correspondence problems like optic flow belong to the fundamental problems in computer vision. Here, one aims at finding correspondences between the pixels in two (or more) images. The correspondences are described by a displacement vector field that is often found by minimising an energy (cost) function. In this thesis, we present several contributions to the energy-based solution of correspondence problems: (i) We start by developing a robust data term with a high degree of invariance under illumination changes. Then, we design an anisotropic smoothness term that works complementary to the data term, thereby avoiding undesirable interference. Additionally, we propose a simple method for determining the optimal balance between the two terms. (ii) When discretising image derivatives that occur in our continuous models, we show that adapting one-sided upwind discretisations from the field of hyperbolic differential equations can be beneficial. To ensure a fast solution of the nonlinear system of equations that arises when minimising the energy, we use the recent fast explicit diffusion (FED) solver in an explicit gradient descent scheme. (iii) Finally, we present a novel application of modern optic flow methods where we align exposure series used in high dynamic range (HDR) imaging. Furthermore, we show how the alignment information can be used in a joint super-resolution and HDR method.

Current Students

Name	Supervisor	Department	Nationality
AFSARIYEGANEH, Elham	Jennifer Gerling	IMPRS-CS	*Iranian*
ARVANITIDIS, Georgios	Jennifer Gerling	IMPRS-CS	*Greek*
ASLAM, Shahzad	Jennifer Gerling	IMPRS-CS	*Pakistani*
BELET, Razvan	Jennifer Gerling	IMPRS-CS	*Romanian*
BOLDYREV, Artem	Jennifer Gerling	IMPRS-CS	*Russian*
BOZCA, Sinan	Jennifer Gerling	IMPRS-CS	*Turkish*
DEDIK, Darya	Jennifer Gerling	IMPRS-CS	*Ukrainian*
DINU, Lavinia	Jennifer Gerling	IMPRS-CS	*Romanian*
EGHBALI, Mandana	Jennifer Gerling	IMPRS-CS	*Iranian*
GRISHCHENKO, Ilya	Jennifer Gerling	IMPRS-CS	*Russian*
HARIMAN, Charles Darwis	Jennifer Gerling	IMPRS-CS	*Indonesian*
ILIEVA, Evica	Jennifer Gerling	IMPRS-CS	*Macedonian*
KAPCARI, Edite	Jennifer Gerling	IMPRS-CS	*Albanian*
KHOREVA, Anna	Jennifer Gerling	IMPRS-CS	*Russian*
KOLEV, Pavel	Jennifer Gerling	IMPRS-CS	*Polish*
LEVINKOV, Evgeny	Jennifer Gerling	IMPRS-CS	*Russian*
MERCADO LOPEZ, Pedro Eduardo	Jennifer Gerling	IMPRS-CS	*Mexican*
MUKHA, Viktor	Jennifer Gerling	IMPRS-CS	*Russian*
NGYUEN, Quynh Ngoc	Jennifer Gerling	IMPRS-CS	*Vietnamese*
OMRAN, Mohamed	Jennifer Gerling	IMPRS-CS	*Egyptian*
PODOSINNIKOVA, Anastasia	Jennifer Gerling	IMPRS-CS	*Russian*
REZNITSKII, Maxim	Jennifer Gerling	IMPRS-CS	*Canadian*
TIAB, John	Jennifer Gerling	IMPRS-CS	*Egyptian*
VASUDEVAN, Aravind	Jennifer Gerling	IMPRS-CS	*Indian*
VIDRIALES ESCOBAR, Mónica Paola	Jennifer Gerling	IMPRS-CS	*Mexican*
ZINCHENKO, Tania	Jennifer Gerling	IMPRS-CS	*Russian*

Name	Supervisor	Department	Nationality
ABDEL MAKSOUD, Mohamed	Reinhard Wilhelm, Gerhard Weikum	Compiler Construction	*Egyptian*
ABED, Fidaa	Kurt Mehlhorn	Algorithms and Complexity	*Palestinian*
ALVANAKI, Foteini	Sebastian Michel	Querying, Indexing, and Discovery in Dynamic Data	*Greek*
ANAND, Avishek	Gerhard Weikum	Databases and Information Systems	*Indian*
ASSENOV, Yassen	Thomas Lengauer	Computational Biology and Applied Algorithmics	*Bulgarian*
AZMY, Noran	Christoph Weidenbach	Automation of Logic	*Egyptian*
BEEDKAR, Kaustubh	Rainer Gemulla	Databases and Information Systems	*Indian*
BEGGEL, Bastian	Thomas Lengauer	Computational Biology and Applied Algorithmics	*German*
BLANKENBURG, Hagen	Thomas Lengauer	Computational Biology and Applied Algorithmics	*German*
BRINGMANN, Karl	Tobias Friedrich	Algorithms and Complexity	*German*
BYELOZYOROV, Sergiy	Philipp Slusallek, Thorsten Thormählen	Computer Graphics	*Ukrainian*
CHIU, Walon Wei-Chen	Mario Fritz	Computer Vision and Multimodal Computing	*Chinese*
CIOBOTARU, Oana	Michael Backes	Information Security and Cryptography	*Romanian*
CROITORU, Cosmina	Kurt Mehlhorn	Algorithms and Complexity	*Romanian*
DANILEWSKI, Piotr	Philipp Slusallek, Christian Theobalt	Computer Graphics	*Polish*
DEL CORRO GONZALEZ GASS, Luciano	Rainer Gemulla	Databases and Information Systems	*Spanish*
DIETZEN, Matthias	Thomas Lengauer	Computational Biology and Applied Algorithmics	*German*
DIMITROVA, Rayna	Bernd Finkbeiner, Gerhard Weikum	Reactive Systems	*Bulgarian*

Name	Supervisor	Department	Nationality
DONCHEVA, Nadezhda	Mario Albrecht, Thomas Lengauer	Computational Biology and Applied Algorithmics	*Bulgarian*
DYLLA, Maximilian	Martin Theobald	Databases and Information Systems	*German*
EBERT, Peter	Thomas Lengauer	Computational Biology and Applied Algorithmics	*German*
ELEK, Oskar	Tobias Ritschel	Rendering and GPUs	*Czech*
ELHAYEK, Ahmed	Christian Theobalt	Computer Graphics	*Palestinian*
ERNST, Patrick	Gerhard Weikum	Databases and Information Systems	*German*
FERNANDES ALBRECHT, Felipe	Thomas Lengauer	Computational Biology and Applied Algorithmics	*German*
FEUERBACH, Lars	Thomas Lengauer	Computational Biology and Applied Algorithmics	*German*
FIETZKE, Arnaud	Christoph Weidenbach	Automation of Logic	*German*
FLINDERS, Mathieu	Glenn Lawyer	Computational Biology and Applied Algorithmics	*British*
GALÁRRAGA DEL PRADO, Luis	Fabian Suchanek	Ontologies	*Ecuadorian*
GEORGIEV, Iliyan	Philipp Slusallek, Hans-Peter Seidel	Computer Graphics	*Bulgarian*
GERLING, Sebastian	Michael Backes	Information Security and Cryptography	*German*
GRANADOS VELASQUEZ, Miguel Andres	Christian Theobalt	Computer Graphics	*Colombian*
GROCHULLA, Martin Peter	Thorsten Thormählen	Computer Graphics	*German*
GROSS, Florian	Andreas Zeller, Gerhard Weikum	Software Engineering	*German*
HAGEMANN, Willem	Uwe Waldmann	Automation of Logic	*German*
HALACHEV, Konstantin	Thomas Lengauer	Computational Biology and Applied Algorithmics	*Bulgarian*

Name	Supervisor	Department	Nationality
HATEFIARDAKANI, Hassan	Holger Hermanns, Christoph Weidenbach	Dependable Systems and Software	*Iranian*
HAUSCHILD, Anne-Christin	Jan Baumbach, Volkhard Helms	Computational Systems Biology	*German*
HELTEN, Thomas	Meinard Müller	Multimedia Information Retrieval and Music Processing	*German*
HOFFART, Johannes	Gerhard Weikum	Databases and Information Systems	*German*
IBRAGIMOV, Rashid	Jan Baumbach	Computational Systems Biology	*Uzbek*
JAIN, Arjun	Thorsten Thormählen	Computer Graphics	*Indian*
JANSEN, Silke	Michael Wand	Computer Graphics	*German*
JIANG, Nanzhu	Meinard Müller	Computer Graphics	*Chinese*
JURKIEWICZ, Tomasz	Kurt Mehlhorn	Algorithms and Complexity	*Polish*
KALAGHATGI, Prabhav	Glenn Lawyer	Computational Biology and Applied Algorithmics	*Indian*
KALOJANOV, Javor	Philipp Slusallek, Michael Wand	Computer Graphics	*Bulgarian*
KELLNHOFER, Petr	Karol Myszkowsky	Computer Graphics	*Czech*
KERBER, Jens	Michael Wand	Computer Graphics	*German*
KHOSLA, Megha	Kurt Mehlhorn	Algorithms and Complexity	*Indian*
KOBEL, Alexander	Michael Sagraloff	Algorithms and Complexity	*German*
KONDREDDI, Sarath Kumar	Gerhard Weikum	Databases and Information Systems	*Indian*
KOSTA, Marek	Christoph Weidenbach	Automation of Logic	*Slovakian*
KURZ, Christian	Thorsten Thormählen	Computer Graphics	*Austrian*
KUZEY, Erdal	Gerhard Weikum	Databases and Information Systems	*Turkish*
LAMOTTE-SCHUBERT, Manuel	Christoph Weidenbach	Automation of Logic	*German*

Name	Supervisor	Department	Nationality
LI, Bi-Qing	Thomas Lengauer	Computational Biology and Applied Algorithmics	*Chinese*
LU, Tianxiang	Christoph Weidenbach	Automation of Logic	*Chinese*
MAKARI MANSHADI, Faraz	Rainer Gemulla	Databases and Information Systems	*German*
MALINOWSKI, Mateusz	Mario Fritz	Computer Vision and Multimodal Computing	*Polish*
MANAKOV, Alkhazur	Ivo Ihrke	Generalized Image Acquisition and Analysis	*Russian*
MARSALEK, Lukas	Philipp Slusallek, Thorsten Thormählen	Computer Graphics	*Czech*
METZGER, Steffen	Ralf Schenkel	Efficient Search in Semistructured Data Spaces	*German*
MORAN, Shay	Kurt Mehlhorn	Algorithms and Complexity	*Israeli*
NGUYEN HOANG, Chuong	Karol Myszkowski	Computer Graphics	*Vietnamese*
OTT, Sebastian	Rob van Stee	Algorithms and Complexity	*German*
PEPIKJ, Bojan	Bernt Schiele	Computer Vision and Multimodal Computing	*Macedonian*
PISHCHULIN, Leonid	Bernt Schiele	Computer Vision and Multimodal Computing	*Russian*
POURIMI, Ali	Kurt Mehlhorn	Algorithms and Complexity	*Iranian*
PRYTKOVA, Natalia	Gerhard Weikum, Marc Spaniol	Databases and Information Systems	*Russian*
QU, Lizhen	Gerhard Weikum	Databases and Information Systems	*Chinese*
RAMEZANI, Fahimeh	Kurt Mehlhorn	Algorithms and Complexity	*Iranian*
RANGAPURAM, Syama Sundar Yadav	Matthias Hein, Kurt Mehlhorn	Machine Learning	*Indian*
RESHETOUSKI, Ilya	Ivo Ihrke	Generalized Image Acquisition and Analysis	*Belarusian*

Name	Supervisor	Department	Nationality
RIZKALLAH, Christine Sherif	Rob van Stee	Algorithms and Complexity	*Canadian*
ROHRBACH, Marcus	Bernt Schiele	Computer Vision and Multimodal Computing	*German*
RÖTTGER, Richard Philipp	Jan Baumbach	Computational Systems Biology	*German*
SCHELHORN, Sven Eric	Mario Albrecht	Computational Biology and Applied Algorithmics	*German*
SCHERBAUM, Kristina	Thorsten Thormählen	Computer Graphics	*German*
SENINA, Anna	Mykhaylo Andriluka	Computer Vision and Multimodal Computing	*Ukrainian*
SEUFERT, Stephan	Gerhard Weikum	Databases and Information Systems	*German*
SHAHEEN, Mohammed	Hans-Peter Seidel	Computer Graphics	*Palestinian*
SIMEONOVSKI, Milivioj	Michael Backes, Gerhard Weikum	Information Security and Cryptography	*Macedonian*
SIU, Yin Amy	Gerhard Weikum	Databases and Information Systems	*Australian*
SRIDHAR, Srinath	Christian Theobalt, Antti Oulasvirta	Computer Graphics	*Indian*
STÖTER, Torsten	Tino Weinkauf	Computer Graphics	*German*
STUPAR, Aleksandar	Sebastian Michel, Ralf Schenkel	Querying, Indexing, and Discovery in Dynamic Data	*Serbian*
SUDA, Martin	Christoph Weidenbach	Automation of Logic	*Czech*
TANDON, Niket	Gerhard Weikum	Databases and Information Systems	*Indian*
TANEVA, Bilyana	Gerhard Weikum	Databases and Information Systems	*Bulgarian*
TANG, Ching Hoo	Christoph Weidenbach	Automation of Logic	*German*
TEFLIOUDI, Christina	Rainer Gemulla	Databases and Information Systems	*Greek*

Name	Supervisor	Department	Nationality
TEMPLIN, Krzysztof	Karol Myszkowski	Computer Graphics	*Polish*
TYLENDA, Tomasz	Gerhard Weikum	Databases and Information Systems	*Polish*
WANG, Yafang	Gerhard Weikum	Databases and Information Systems	*Chinese*
WANG, Zhongjie	Thorsten Thormählen	Computer Graphics	*Chinese*
WU, Chenglei	Christian Theobalt	Computer Graphics	*Chinese*
WU, Xiaokun	Thorsten Thormählen	Computer Graphics	*Chinese*
YAHYA, Mohamed	Gerhard Weikum	Databases and Information Systems	*Palestinian*
YOSEF, Mohamed Amir	Gerhard Weikum	Databases and Information Systems	*Egyptian*

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

Krista AMES

NATIONALITY: US-American
POSITION: German and English Teacher
E-MAIL: ames@mpi-inf.mpg.de

EDUCATION: German
UNIVERSITY: Willamette University, Salem, OR; LMU München

Patrick DEMPSEY

NATIONALITY: Irish
POSITION: English Teacher
E-MAIL: dempsey@mpi-inf.mpg.de

EDUCATION: Comparative Literature & Cinema studies
UNIVERSITY: San Francisco State University

Irena DOTCHEVA

NATIONALITY: Bulgarian
POSITION: English Teacher
E-MAIL: irena@mpi-inf.mpg.de

EDUCATION: English Linguistics & Computational Linguistics
UNIVERSITY: Saarland University

Aikaterini GEORGIOU

NATIONALITY: Greek
POSITION: German Teacher
E-MAIL: georgiou@mpi-inf.mpg.de

EDUCATION: Translating, German as a Foreign Language
UNIVERSITY: Ionian University, Saarland University

Jennifer GERLING

NATIONALITY: German
POSITION: IMPRS-CS Coordinator
E-MAIL: jgerling@mpi-inf.mpg.de

EDUCATION: Applied Linguistics and Translating (Spanish, English, and German)
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 Bălţi, Saarland University

Stephanie JÖRG

NATIONALITY: German
POSITION: Assistant to the Coordinator
E-MAIL: sjoerg@mpi-inf.mpg.de

EDUCATION: Translator for English and Spanish
UNIVERSITY: Saarland University

Heike PRZYBYL

NATIONALITY: German
POSITION: Student Assistant
E-MAIL: hprzybyl@mpi-inf.mpg.de

EDUCATION: Translating / Interpreting for English and French
UNIVERSITY: Saarland University

Marc SCHMITT

NATIONALITY: German
POSITION: Student Assistant
E-MAIL: mschmitt@mpi-inf.mpg.de

EDUCATION: Historically Oriented Cultural Sciences
UNIVERSITY: Saarland University

Katja SCHÖNBORN

NATIONALITY: German
POSITION: Webmaster
E-MAIL: schoenb@mpi-inf.mpg.de

EDUCATION: Sociology, Psychology, Political Science, English
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

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-CS STUDENTS RECEIVED AWARDS FOR THEIR WORK:

2012

NAME OF STUDENT	AWARD
Alejandro Pironti	Best Poster Presentation of 2012, 12th European Meeting on HIV & Hepatitis, Barcelona
Aleksandar Stupar	Google Travel Grant. SIGMOD Conference, 2012
Gerard de Melo	Dr. Eduard Martin-Preis (FOR HIS DISSERTATION)
Tunc Aydin	Eurographics Best PhD Thesis Award
Matthias Hullin	Otto Hahn Medal (FOR HIS DISSERTATION)
Carola Winzen, *Madhusudan* Manjunath	Feodor Lynen Fellowship of the Humboldt Foundation

2011

NAME OF STUDENT	AWARD
Johannes Hoffart	Best Demo Paper Award at the World Wide Web Conference 2011
Adrian Neumann	Günther Hotz Medal (FOR HIS MASTER'S THESIS)
Natalia Prytkova	Student Grant of the 8th European Summer School on Information Retrieval
Karolina Soltys	Recipient of the American Association of University Women Fellowship
Elham Afsari Yeganeh, *Oana* Ciobotaru	Google Anita Borg Scholarship Finalist
Nanzhu Jiang, *Peter M.* Grosche	Best Paper Award, ISMIR 2011
Tobias Ritschel	Eurographics Best PhD Thesis Award

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 institut informatik

Max Planck Institute for Software Systems

SAARLAND UNIVERSITY
COMPUTER SCIENCE