SCHLOSS DAGSTUHL
Leibniz-Zentrum für Informatik

Dagstuhl News

January - December 2008

Volume 11
2009

Dagstuhl Publications
Dagstuhl News January-December 2008
http://drops.dagstuhl.de/opus/volltexte/2009/2057

ISSN 1438-7581

Period: January - December 2008

Frequency: One per year

Schloss Dagstuhl, the Leibniz Center for Informatics is operated by a non-profit organization. Its objective is to promote world-class research in computer science and to host research seminars which enable new ideas to be showcased, problems to be discussed and the course to be set for future development in this field.

Online version: http://www.dagstuhl.de/files/Reports/News/

Associates: Gesellschaft für Informatik e.V. (GI), Bonn
Technical University of Darmstadt
Frankfurt University
Technical University of Kaiserslautern
University of Karlsruhe
University of Stuttgart
University of Trier
Saarland University
Max Planck Society e.V. (MPG)
French National Institute for Research in Informatics and Automatic Control (INRIA)
Dutch National Research Institute for Mathematics and Informatics (CWI)

Membership: The Center is a member of the Leibniz Association.

Funded by: Federal funding and state funding of Saarland and Rhineland Palatinate

Information: Schloss Dagstuhl Office
Saarland University
P.O. Box 15 11 50
66041 Saarbrücken, Germany
Phone: +49-681-302-4396
Fax: +49-681-302-4397
E-mail: service@dagstuhl.de
http://www.dagstuhl.de/

Welcome

Here are the Dagstuhl News for 2008, the 11th edition of the "Dagstuhl News", a publication for the members of the Foundation "Informatikzentrum Schloss Dagstuhl", the *Dagstuhl Foundation* for short.

The main part of this volume consists of collected resumees from the Dagstuhl Seminar Reports 2008 and Manifestos of Perspectives Workshops. We hope that you will find this information valuable for your own work or informative as to what colleagues in other research areas of Computer Science are doing. The full reports for 2008 are on the Web under URL: `http://drops.dagstuhl.de/opus/institut.php?fakultaet=01&year=08/`

Our online-publication service, started to publish online proceedings of our Dagstuhl Seminars, is catching on as a service to the Computer Science community. The Dagstuhl Research Online Publication Server (DROPS) (`http://www.dagstuhl.de/publikationen/publikationsserver-drops/`) hosts the proceedings of a few external workshop and conference series. A group of scientists around the STACS and FSTTCS conference series have negotiated with us the establishment of a high-quality proceedings series, the Leibniz International Proceedings in Informatics, LIPIcs.
`http://www.dagstuhl.de/publikationen/lipics/`
`http://drops.dagstuhl.de/opus/institut_lipics.php?fakultaet=04`
Please read the announcement to learn more about LIPIcs.

It is hard to believe, but Dagstuhl gets even more popular than it already is. We have received a record number of 93 proposals for Dagstuhl Seminars and Perspectives Workshops in 2009. We try to accommodate more workshops by adding an extension building with 7 more rooms. It will allow us to run two Seminars in parallel once the building is finished. We have selected an architecture for the building and hope to finish it next year.

In July, we will be evaluated by the Leibniz Senate for Evaluation. A successful evaluation will grant Dagstuhl an extension of the current financial scheme by another 7 years.

The State and the Activities of the *Dagstuhl Foundation*

The foundation currently has 45 personal members and 7 institutional members.
We are experiencing a growing number of requests for travel support or a reduction of the seminar fees. In 2008, we have supported a number of guests in either of these ways.

Thanks

I would like to thank you for supporting Dagstuhl through your membership in the *Dagstuhl Foundation*. Thanks go to Fritz Müller for editing the resumees collected in this volume.

Reinhard Wilhelm (Scientific Director)

Saarbrücken, May 2009

Contents

Chapter 1

Data Structures, Algorithms, Complexity

1.1 Theory of Evolutionary Algorithms

Seminar No. **08051** Date **27.01.–01.02.2008**
Organizers: Dirk V. Arnold, Anne Auger, Jonathan E. Rowe, Carsten Witt

Motivation and Goals

Evolutionary algorithms (EAs) are randomised heuristics for search and optimisation that are based on principles derived from natural evolution. Mutation, recombination, and selection are iterated with the goal of driving a population of candidate solutions toward better and better regions of the search space. Since the underlying idea is easy to grasp and almost no information about the problem to be optimised is necessary in order to apply it, EAs are widely used in many practical disciplines, mainly in computer science and engineering.

It is a central goal of theoretical investigations of EAs to assist practitioners with the tasks of selecting and designing good strategy variants and operators. Due to the rapid pace at which new strategy variants and operators are being proposed, theoretical foundations of EAs still lag behind practice. However, EA theory has gained much momentum over the last few years and has made numerous valuable contributions to the field of evolutionary computation.

EA theory today consists of a wide range of different approaches. Run-time analysis, facet-wise analysis that concentrates on the one-step behaviour of EAs (e.g., the well-known schema theory), analysis of the dynamics of EAs, and systematic empirical analysis all consider different aspects of EA behaviour. Moreover, they employ different methods and tools for attaining their goals, such as Markov chains, infinite population models, or ideas based on statistical mechanics or population dynamics.

The focus of the 2006 Dagstuhl Seminar on the "Theory of Evolutionary Algorithms" was to find common ground among the multitude of theoretical approaches and identify open

questions central to the field as a whole. In the ensuing discussions during the seminar it became clear that the existence of a wide variety of approaches can be considered evidence of the richness of the field, and that each stream has its own raison d'être. For example, rigorous analysis yields exact results, but can only deal with relatively simple problems. Purely experimental analysis can deal with much more complicated problems, but does not generalise well.

The goals of the 2008 seminar were twofold. The first goal was to gain a systematic understanding of what the capabilities and limitations of the different methodological approaches are, and how they can assist and complement each other. The second goal was to address some of the central open questions identified during the previous seminar, and to establish what the various theoretical approaches can contribute to answering them with a focus on issues related to design and analysis of EAs. For situations where a complete theoretical analysis is out of reach, the goal was to establish a rigorous framework for empirical studies.

Participants and Results

The seminar brought together 47 researchers from nine countries, and from across the spectrum of EA theory. Talks were arranged into eight sessions, grouped loosely around common themes, such as hardness and complexity, multi-objective optimisation, coevolution, and continuous optimisation.

Thomas Jansen (Technische Universität Dortmund)
Bit flip mutations vs. local search: An open problem presentation

Frank Neumann (MPI für Informatik, Saarbrücken)
Making problems easier by multi-objective optimization

Holger Hoos (University of British Columbia, Vancouver)
Automated tuning, configuration and synthesis of complex stochastic local search algorithms

Alden Wright (University of Montana)
Does a temporally or spatially varying environment speed up evolution?

Dirk V. Arnold (Dalhousie University)
Step length adaptation on ridge functions

Steffen Finck (Fachhochschule Vorarlberg)
Performance of evolution strategies on PDQFs

Olivier Teytaud (Université Paris Sud)
Complexity lower bounds for evolution strategies

Christian Igel (Ruhr-Universität Bochum)
Efficient covariance matrix update for evolution strategies

Dirk Sudholt (Technische Universität Dortmund)
Runtime analysis of binary PSO

Benjamin Doerr (MPI für Informatik, Saarbrücken)
A non-artificial problem where crossover provably helps

Ingo Wegener (Technische Universität Dortmund)
Tight bounds for blind search on the integers

Yossi Borenstein (University of Essex)
Kolmogorov complexity and hardness

Darrell Whitley (Colorado State University)
Relating experiments and theory

Adam Prügel-Bennett (University of Southampton)
Solving problems with critical variables

Jonathan E. Rowe (University of Birmingham)
Representation-invariant crossover and mutation operators

Anthony Bucci (Icosystem)
Solution concepts and the subdomain representation

Elena Popovici (Icosystem)
Monotonic convergence and memory requirements of algorithms for "interactive" search problems

R. Paul Wiegand (University of Central Florida, Orlando)
Conditions for the robustness of compositional coevolution

Jonathan L. Shapiro (Manchester University)
Convergence in co-adapting agents with opponents modelling

Eckart Zitzler (ETH Zürich)
Approximating the pareto set using set preference relations: A new perspective on evolutionary multiobjective optimization

Jürgen Branke (Universität Karlsruhe)
Evolutionary multi-objective worst case optimization

Nicholas Freitag McPhee (University of Minnesota, Morris)
N-gram GP: Early result and questions

Anton Eremeev (Sobolev Institute of Mathematics, Omsk)
NP-hard cases of optimal recombination

Nikolaus Hansen (INRIA Futurs, Orsay)
Toward a convergence proof for CMA-ES — and beyond

Alexander Melkozerov (Fachhochschule Vorarlberg)
Weighted multirecombination evolution strategy with mutation strength self-adaptation on the quadratic sphere

Olivier Francois (TIMC Laboratory)
Evolution strategies with random numbers of offspring

Jun He (University of Birmingham)
A theoretical analysis to an experimental comparison of GAs using penalty function methods and repairing methods

---

Boris S. Mitavskiy (University of Sheffield)
Estimating the stationary distributions of Markov chains modeling evolutionary algorithms using quotient construction method

Lothar M. Schmitt (The University of Aizu)
Banach space techniques for the analysis of evolutionary algorithms

Riccardo Poli (University of Essex)
Bye, bye, bloat

Kenneth A. De Jong (George Mason University)
Closing discussion

Many of the presentations and discussions were dedicated to identifying the limitations of the various approaches, shedding light on their complementarity, and arriving at a wider consent with regard to advantages and disadvantages of the different techniques. A new theme that had been largely absent from previous Dagstuhl seminars on the "Theory of Evolutionary Algorithms" and that recurred again and again during the present seminar was a discussion of how experimental work can complement and inspire theoretical analysis. Further themes that were discussed and that are expected to have an impact on future research are the adaptation of mutation distributions in continuous search spaces, and how tools for investigating the stability of Markov chains can be used for the analysis of adaptive algorithms. Finally, important steps were made toward the goal of applying the tools and techniques that have been developed for the runtime analysis of evolutionary algorithms to other modern search heuristics, such as ant colony optimisation and swarm intelligence approaches.

Conclusion

Besides the presentations, and as at past Dagstuhl seminars on "Theory of Evolutionary Algorithms", fruitful and stimulating discussions among varying groups of participants occurred throughout the week. The Dagstuhl seminars are firmly established in the community as a biannual event, and we hope to be able to build on this success and continue to promote discussions between researchers in different areas of EA theory at further workshops in the future.

1.2 Data Structures

Seminar No. **08081** Date **17.02.–22.02.2008**
Organizers: Lars Arge, Robert Sedgewick, Raimund Seidel

The purpose of this workshop was to discuss recent developments in various aspects of data structure research and also to familiarize the community with some of the problems that arise in the context of using so-called flash memory as a storage medium. 49 researchers attended the workshop producing a congenial and productive atmosphere.

A number of contributions reported on new advances on old problems in data structure research: E.g. John Iacono summarized recent advances regarding the dynamic optimality conjecture of splay trees, Bob Tarjan showed new results on dynamically maintaining an acyclic graph, Bob Sedgewick presented a particularly simple version of red-black trees, and Mihai Patrascu offered a new, plausible approach for proving super-polylogarithmic lower bounds to dynamic data structure problems.

The design of data structures and algorithms that can deal well with the memory hierarchy used in today's computers was another core topic of the workshop. E.g. Nobert Zeh discussed this problem in the context of red-blue line segment intersection, Rico Jacob considered the multiplication of dense vectors by a sparse matrix, and Michael Bender presented so-called streaming B-trees that can implement dictionaries well in this memory hierarchy context.

Reducing the actual space requirement of data structures to a minimum was the topic of several contributions. For instance Martin Dietzfelbinger and also Arash Farzan considered this problem from the theoretical point of view, and Peter Sanders and also Holger Bast from a very practical point of view.

A highlight of the workshop were the presentations by San Lyul Min and by Ethan Miller on so-called flash memory, which is physical memory that has found its way into many storage devices but has technological properties that lead to the following peculiarities: (i) Before a bit is written it has to be cleared, however this can only be achieved by clearing an entire block of bits. (ii) Every block of bits can only be cleared a (large) constant number of times before the storage capabilities of this block become too unreliable.

There were many more interesting contributions than listed here and, of course, countless discussions and collaborations. The Dagstuhl atmosphere provided just the right environment for all this.

1.3 Graph Drawing with Applications to Bioinformatics and Social Sciences

Seminar No. **08191** Date **04.05.–09.05.2008**
Organizers: Stephen P. Borgatti, Stephen Kobourov, Oliver Kohlbacher, Petra Mutzel

Graph drawing deals with the problem of communicating the structure of relational data through diagrams, or drawings. Graphs with vertices and edges are typically used to model relational data. The vertices represent the objects (or data points) and the edges represent the relationships between the objects. The main problem in relational visualization is to display the data in a meaningful fashion that may heavily depend on the application domain. Some of the application areas where graph drawing tools are needed include computer science (data base design, data mining, software engineering), bioinformatics (metabolic networks, protein-protein interaction), business informatics (business process models), and the social sciences and criminalistics (social networks, phone-call graphs).

The ability to represent relational information in a graphical form is a powerful tool which allows us to perform analysis through visual exploration. With the aid of graph visualization we can find important patterns, trends, and correlations. Real-world applications such as bioinformatics and sociology pose additional challenges, e.g., semantic information carried by the diagram has to be used for obtaining meaningful layouts and application-specific drawing conventions need to be fulfilled. Moreover, the underlying data often stems from huge data bases, but only a small fraction shall be displayed at a time; the user interactively selects the data to be displayed and explores the graph by expanding interesting and collapsing irrelevant parts. This requires powerful graph exploration tools with navigation capabilities that allow dynamic adaption of the graph layout in real time.

Topics of the Seminar

In this seminar we focused on the application of graph drawing in two important application domains: bioinformatics (metabolic pathways, regulatory networks, protein-protein interaction)and social sciences and criminalistics (case information diagrams, phone-call graphs). In both application domains, the underlying information is usually stored in large data bases constituting a huge and complex graph, but only a suitable fraction of this graph is visualized; the selection of that subgraph is guided by the user and even more user interaction occurs in order to further explore the underlying graph. Thus, the user becomes a central actor that triggers dynamic updates of the displayed graph and its layout. The support of application-specific update functionality in conjunction with high quality graph layout is essential for achieving user acceptance in the targeted application areas.

The interactive navigation through the graph poses new challenges to graph drawing algorithms. Whereas traditional graph drawing deals with the visualization of static graphs, we are now concerned with graphs that change over time, and the layout has to be adjusted

in real time. The new layout has to observe aesthetic and application specific drawing criteria, as well as the preservation of the user's mental map; in particular only few changes in the layout are desired.

A similar dynamic component occurs in the visualization of graphs that evolve over time like minute-by-minute phone-call graphs which have application in police investigations. Here, we have a graph at each time point and an edge corresponds to a phone call between two telephones. In contrast to interactive navigation, we know in advance all of the changes the graph will undergo, and we can exploit this fact for producing a smoother animation sequence.

In summary, it is our impression that the participants enjoyed the great scientific atmosphere offered by Schloss Dagstuhl, and profited from the scientific program. We are grateful for having had the opportunity to organize this seminar. Special thanks are due to Carsten Gutwenger and Karsten Klein for their assistance in the organization and the running of the seminar.

1.4 Design and Analysis of Randomized and Approximation Algorithms

Seminar No. **08201** Date **11.05.–16.05.2008**
Organizers: Martin E. Dyer, Mark Jerrum, Marek Karpinski

The workshop was concerned with the newest developments in the design and analysis of randomized and approximation algorithms. The main focus of the workshop was on three specific topics: approximation algorithms for optimization problems, approximation algorithms for measurement problems, and decentralized networks as well as various interactions between them. This included all sorts of completely new algorithmic questions that lie on the interface of several different areas. Here, some new broadly applicable techniques have emerged recently for designing efficient approximation algorithms for various optimization and measurement problems. This workshop has addressed the above topics and also some new fundamental paradigms and insights into the algorithm design techniques.

The 30 lectures delivered at this workshop covered a wide body of research in the above areas. The meeting was held in a very pleasant and stimulating atmosphere. Thanks to everyone who made it a very interesting and enjoyable event.

1.5 Structure-Based Compression of Complex Massive Data

Seminar No. **08261** Date **22.06.–27.06.2008**
Organizers: Stefan Böttcher, Markus Lohrey, Sebastian Maneth, Wojciech Rytter

Compression of massive complex data is a promising technique to reduce stored and transferred data volumes. In industry, semi-structured data and XML in particular, have become the de facto data exchange format, and e.g. grammar-based compression offers an efficient memory representation that can be used within XML databases. Therefore, we expect improvements on compression techniques for structured data and XML to have significant impact on a variety of industry applications where data exchange or limited data storage are bottlenecks.

The big advantage of structure and grammar-based compression over other compression techniques is that many algorithms can be performed directly on the compressed structure, without prior decompression. This idea of "computing on compressed structures" has given rise to efficient algorithms in several areas; e.g., term graph rewriting, model-checking using BDDs, and querying XML; all three areas profit from the use of dags (directed acyclic graphs) which allow to share common subtrees and which can be seen as a particular instance of grammar-based compression. In these areas and others, we expect efficiency improvements through the use of more sophisticated grammar-based compression techniques.

The goal of the seminar was to bring together researchers working on various aspects of structure-based compression and to discuss new ideas and directions for doing research on compression and on computation over compressed structures. In particular, the sub-goals were to achieve a deeper understanding of the whole field of compressing structured data, to discover new interconnections between different research directions in compression, to interchange ideas between theory and application oriented research, to distinguish between solved and open questions in the field, to identify key problems for further research, and to initiate new collaborations.

Within the seminar, many different forms of cooperations took place. The seminar started with a short introduction of each participant. The introduction was followed by a first series of overview talks, namely:

1. "Grammar-Based Compression (Survey)" by Wojciech Rytter and
2. "Pattern Matching on Compressed Strings", jointly given by Shunsuke Inenaga and Ayumi Shinohara.
3. "Practical Search on Compressed Text" by Gonzalo Navarro and
4. "XML Compression Techniques" by Gregory Leighton.
5. "Algorithmics on Compressed Strings" by Markus Lohrey.

The seminar brought together researchers from different fields of data compression. Many fruitful discussions provided a unique opportunity to discover interconnections between different research directions in compression, to identify open key problems for further research, and to get a much broader understanding of the field. Furthermore, a creative interchange of ideas between theory and application oriented research led to many deep insights for further research in the field. Finally, Dagstuhl initiated new collaborations

which have been fruitful already (one open problem has been resolved, as mentioned in the section on Working Group II), and a large software project software project has been initiated (see the section on Working Group I). As always, Dagstuhl provided a highly enjoyable atmosphere to work in.

1.6 Sublinear Algorithms

Seminar No. **08341** Date **17.08.–22.08.2008**
Organizers: Artur Czumaj, S. Muthu Muthukrishnan, Ronitt Rubinfeld, Christian Sohler

With the increasing role of information technologies we are often confronted with a huge amount of information that is generated without pace by distributed sources or by large scale complex information systems.

In many scenarios, it is not possible to entirely store this information on standard storage devices.

Examples include the World Wide Web, data accumulated in network traffic monitoring, or sensor network data. One of the key challenges in this context is to efficiently process these massive data sets and to extract knowledge by summarizing and aggregating their major features. Most of the time, it is impossible to use traditional algorithms for this purpose. Even linear time algorithms are typically too slow because they require random access to the input data. We require algorithms that either look only at a small random sample of the input or process the data as it arrives extracting a small summary.

Algorithms of this type are called *sublinear algorithms*.

The purpose of this workshop was to bring together leading researchers in the are of sublinear algorithms to discuss recent advances in the area, identify new research directions, and discuss open problems.

The area of sublinear algorithms can be split into three subareas: *property testing*, *sublinear time approximation*, and *data streaming algorithms*.

These areas are not only connected by the fact that they require algorithms with sublinear resources but also that they heavily rely on randomization and random sampling. Researchers from all three areas came to attend this workshop and we believe that it helped to exchange ideas between these different areas.

During the seminar one could obtain a good overview of the current state of sublinear algorithms.

In many interesting talks new algorithms and models as well as solutions to well-known open problems were presented. To give an idea about the topics of the seminar we present a few examples of topics that were discussed in a number of talks at the seminar. These examples are not meant to be exhaustive. Sublinear algorithms, property testing, data streaming, graph algorithms, approximation algorithms.

1.7 Computational Complexity of Discrete Problems

Seminar No. **08381** Date **14.09.–19.09.2008**
Organizers: Peter Bro Miltersen, Rüdiger Reischuk, Georg Schnitger, Dieter van Melkebeek

Estimating the computational complexity of discrete problems constitutes one of the central and classical topics in the theory of computation. Mathematicians and computer scientists have long tried to classify natural families of Boolean relations according to fundamental complexity measures like time and space, both in the uniform and in the nonuniform setting. Several models of computation have been developed in order to capture the various capabilities of digital computing devices, including parallelism, randomness, and quantum interference.

In addition, complexity theorists have studied other computational processes arising in diverse areas of computer science, each with their own relevant complexity measures. Several of those investigations have evolved into substantial research areas in their own right, including:

- proof complexity, interactive proofs, and probabilistically checkable proofs (motivated by verification),
- approximability (motivated by optimization),
- pseudo-randomness and zero-knowledge (motivated by cryptography and security),
- computational learning theory (motivated by artificial intelligence),
- communication complexity (motivated by distributed computing), and
- query complexity (motivated by databases).

The analysis and relative power of the basic models of computation remains a major challenge, with intricate connections to all of the areas mentioned above. Several lower bound techniques for explicitly defined functions form the basis for results in propositional proof complexity. The structure that underlies the lower bounds has led to efficient sample-based algorithms for learning unknown functions from certain complexity classes. Close connections have been discovered between circuit lower bounds for certain uniform complexity classes and the possibility of efficient derandomization. The discovery of probabilistically checkable proofs for NP has revived the area of approximability and culminated in surprisingly tight hardness of approximation results for a plethora of NP-hard optimization problems.

Thus, new results that relate or separate different models of computation and new methods for obtaining lower and upper bounds on the complexity of explicit discrete problems are topics of general importance for the computer science community.

The seminar "Computational Complexity of Discrete Problems" has evolved out of the series of seminars entitled "Complexity of Boolean Functions", a topic that has been covered at Dagstuhl on a regular basis since the foundation of IBFI. Over the years, the focus on nonuniform models has broadened to include uniform ones as well. The change in title reflects and solidifies this trend.

1.8 Moderately Exponential Time Algorithms

Seminar No. **08431** Date **19.10.–24.10.2008**
Organizers: Fedor V. Fomin, Kazuo Iwama, Dieter Kratsch

The Dagstuhl seminar on **Moderately Exponential Time Algorithms** took place from 19.10.08 to 24.10.08. This was the first meeting of researchers working on exact and "fast exponential time" algorithms for hard problems. In total 54 participants came from 18 countries.

Moderately exponential time algorithms for NP-hard problems are a natural type of algorithms and research on them dates back to Held and Karp's paper on the travelling salesman problem in the sixties. However until the year 2000, papers were published only sporadically (with the exception of work on satisfiability problems maybe). Some important and fundamental techniques have not been recognized at full value or even been forgotten, as e.g. the Inclusion-Exclusion method from Karp's ORL paper in 1982.

Recently the situation has changed — there is a rapidly increasing interest in exponential time algorithms on hard problems and papers have been accepted for high-level conferences in the last few years. There are many (young) researchers that are attracted by moderately exponential time algorithms, and this interest is easy to explain, the field is still an unexplored continent with many open problems and new techniques are still to appear to solve such problems. To mention a few examples:

- There is a trivial algorithm that for a given SAT formula Φ with m clauses and n variables determines in time roughly $O(2^n + m)$ whether there is a satisfying assignment for Φ. Despite of many attempts, no algorithm of running time $O(c^n+m)$, for some $c < 2$ is known. So what happens here? Is it just because we still do not have appropriate algorithmic techniques or are there deeper reasons for our failure to obtain faster algorithms for some problems? It would be very exciting to prove that (up to some reasonable conjecture in complexity theory) there exists a constant $c > 1$ such that SAT cannot be solved in time c^n .
- One of the most frequently used methods for solving NP-hard problems is Branch & Reduce. The techniques to analyze such algorithms, that we know so far, are based on linear recurrences and are far from being precise. The question here is: How to analyze Branch & Reduce algorithms to establish their worst case running time?
- The algorithm deciding whether a given graph on n vertices has a Hamiltonian cycle has running time $2^n \cdot n^{O(1)}$ and it is known since the 1960s. Amazingly, all progress

in algorithms for the last 40 years did not have any impact on the solution of this problem. Are there new techniques which can be applied to crack this problem?

Despite of the growing interest and the new researchers joining the potential community there has not been a workshop on moderately exponential time algorithms longer than one day since the year 2000. The major goal of the proposed Dagstuhl seminar was to unite for one week many of the researchers being interested in the design and analysis of moderately exponential algorithms for NP-hard problems. The Dagstuhl seminar was a unique opportunity to bring together those people, to share insights and methods, present and discuss open problems and future directions of research in the young domain.

There were 27 talks and 2 open problem sessions. Talks were complemented by intensive informal discussions, and many new research directions and open problems will result from these discussions. The warm and encouraging Dagstuhl atmosphere stimulated new research projects. We expect many new research results and collaborations growing from the seeds of this meeting.

1.9 Structured Decompositions and Efficient Algorithms

Seminar No. **08492** Date **30.11.–05.12.2008**
Organizers: Stephan Dahlke, Ingrid Daubechies, Michael Elad, Gitta Kutyniok, Gerd Teschke

Summary

New emerging technologies such as high-precision sensors or new MRI machines drive us towards a challenging quest for new, more effective, and more daring mathematical models and algorithms. Therefore, in the last few years researchers have started to investigate different methods to efficiently represent or extract relevant information from complex, high dimensional and/or multimodal data. Efficiently in this context means a representation that is linked to the features or characteristics of interest, thereby typically providing a sparse expansion of such. Besides the construction of new and advanced ansatz systems the central question is how to design algorithms that are able to treat complex and high dimensional data and that efficiently perform a suitable approximation of the signal. One of the main challenges is to design new sparse approximation algorithms that would ideally combine, with an adjustable tradeoff, two properties: a provably good 'quality' of the resulting decomposition under mild assumptions on the analyzed sparse signal, and numerically efficient design.

The topic is driven by applications as well as by theoretical questions. Therefore, the aim of this seminar was to bring together a good mixture of scientists with different backgrounds in order to discuss recent progress as well as new challenging perspectives. In particular, it was intended to strengthen the interaction of mathematicians and computer scientists.

The goals of the seminar can be summarized as follows:

- Initiate communications between different focuses of research.
- Comparison of methods.
- Open new areas of applications.
- Manifest the future direction of the field.

Seminar Topics

To reach the seminar goals, the organizers identified in advance the most relevant fields of research:

- Geometric multiscale analysis.
- Compressed sensing.
- Frame construction and coorbit-theory.
- Sparse and efficient reconstruction.
- Probabilistic models and dictionary learning.

The seminar was mainly centered around these topics, and the talks and discussion groups were clustered accordingly. During the seminar, it has turned out that in particular 'compressed sensing' and 'sparse and efficient reconstructions' are currently the most important topics. Indeed, most of the proposed talks were concerned with these two issues. This finding was also manifested by the discussion groups. Although originally small discussion groups for all the above topics were scheduled, the interest was absolutely centered around 'compressed sensing' and 'sparse approximation', so that only these two topics were discussed in greater details in small groups. For a detailed description of the outcome of the results, we refer to the next section.

The course of the seminar gave the impression that *sparsity* with all its facets will probably be one of the most important techniques in applied mathematics and computer sciences in the near future. It was fascinating to see how sparsity concepts are by now influencing many different fields of applications ranging from image processing/compression to adaptive numerical schemes and the treatment of inverse problems. It seems that the sparsity concept allows to exploit a very similar fundamental structure behind all these different applications. Closely related with sparsity is the concept of *compressed sensing*. Originally developed as a tool for sparse signal recovery, ideas from compressed sensing are now flowing into related fields such as numerical analysis and provide a way to design, e.g., more efficient adaptive schemes.

During the seminar, the feeling emerged that in the future the most important progress will probably be made on an *algorithmic* level. The development of new building blocks by itself seems to be of limited use. Instead, more or less data driven algorithms are needed. Moreover, most of the participants had the impression that the time is ripe to deal with new

and challenging applications for the following reason. It seems that in many classical fields of applications such as denoising and compression the mathematical methods have almost reached the ceiling and important further progress cannot be expected. Nevertheless, there is an urgent need to derive efficient algorithms for complex, i.e., high-dimensional and/or multimodal data etc., and these problems should be energetically attacked. In doing so, in particular *geometric* information should be exploited. On an algorithmic level, more effort should be spend to take into account the geometric structures of the signals to be analyzed, to efficiently detect lower-dimensional manifolds on which the essential information is concentrated etc. Moroever, we need a deeper geometric understanding of dictionaries. The choice of the 'best dictonary' with respect to the overall sparsity criteria seems to be essential, and for this, concepts from dictionary learning will probably be very important.

Outcome of the Discussion Groups

As already mentioned, the seminar interest was mainly centered around the topics 'compressed sensing' and 'sparse approximation' and therefore only two discussion groups took place.

- The discussion on 'compressed sensing' (chaired by W. Dahmen) started collecting the following guiding questions and issues:

  - Possible further developments of sparsity models for analog signals and corresponding formulations that can be treated by compressed sensing techniques.
  - A second complex of questions centered around the precise meaning of compressed sensing and resulting limitations. This concerns, in particular, the possibility of progressively upgrading resolution.
  - The conceptual scope of applications of compressed sensing and its potential role in other application areas.
  - The construction of deterministic sensor matrices.

  The first part of the discussion was primarily around construction principles of deterministic sensor matrices, which was seen as a question of central importance. All currently known attempts appear to be based on the concept of coherence which significantly limits the feasible sparsity range to the square root of the number of measurements. A specific ansatz for a concrete sensing matrix was discussed in greater detail, requiring at some point sharp estimates for multiple Gauss sums that may bring in number theoretic aspects. Another direction (but still random) was a construction principle of matrices with less demanding properties than RIP that still give rise to meaningful estimates for coefficients of the signal that are substantial compared with the energy of the whole signal. The remaining discussion revolved around questions concerning compressed sensing techniques for analog signals. This

direction of research is only beginning now and the currently used models should be seen as first examples.

In summary, the general agreement is that compressed sensing has brought in starting new perspectives stressing in particular the beneficial effect of randomness that calls for further investigations and efforts concerning practical applications as well as conceptual clarification regarding its scope and bearing on other areas. One of the challenging questions may concern a proper embedding of such concepts into an infinite dimensional framework.

- The discussion on 'sparse approximation' (chaired by Alfred Bruckstein) started by an attempt to define the state-of-the-art in this field, as a stepping stone for defining future challenges. The ingredients of the state-of-the-art the discussion brought up include

  - Sparsity – an established concept with widespread use and understanding. This is the major achievement that drives the rest.
  - Solid mathematical results in this field that add to its beauty and maturity. Specifically, proofs for the successful behavior of pursuit algorithms of various sorts to solve what seems as NP-hard problems.
  - Advances in numerical tools that support the practicality of this field. Those include iterative shrinkage algorithms for practical pursuit.
  - Success stories in applications in signal processing, image processing, and other fields.
  - Emergence of new dictionaries that serve the sparsity model better, and thus enable state-of-the-art performance in various applications.

  Armed with the above definition of the state-of-the-art, the discussion then turned to future work. First, considering where this field is going next, the following core points were mentioned and discussed:

  - Towards new dictionary design procedures, and ways to incorporate structure into them.
  - Deeper geometric understanding of good dictionaries in an attempt to bridge between the new area of dictionaries and past-used frames.
  - More analytical results (extensions of the theoretical analysis of pursuit methods mentioned above), defining new and related problems.
  - Nonlinear and adaptive designed dictionaries, and ways to extend the sparsity model to realistic data.

The final part in the discussion opened the stage for the dreams the participants have in terms of their research goals and desires. Topics mentioned include:

- Geometry and its relation to Gestalt.
- Operators and function spaces, and using these connections to obtain good dictionaries.
- 3D and beyond for representation systems.
- Divide and conquer and online algorithms.
- Inverse problems and recognition.
- Bilateral and nonlinear connection between coefficients.
- More probabilistic approaches/arguments/analysis.
- More nonlinearity and adaptivity.
- Size of dictionary issues.
- What is the correct way to measure errors in representation?

Final Concluding Remarks

This seminar was regarded by the participants as a very productive and inspiring meeting. Many intense discussions took place throughout the week, and several new cooperations were initiated. Especially, the interactions between computer scientists and applied mathematicians has been extremely fruitful and will certainly be continued in the future. Also, the major future directions of this research area were manifested and initial steps towards solutions undertaken. Concluding, this seminar can be regarded as a milestone in the development of the new, rapidly evolving research area of Structured Decompositions and Efficient Algorithms.

Chapter 2

Verification, Logic, Semantics

2.1 Types, Logics and Semantics for State

Seminar No. **08061** Date **03.02.–08.02.2008**
Organizers: Amal Ahmed, Nick Benton, Martin Hofmann, Greg Morrisett

Introduction

The combination of dynamically allocated, mutable data structures and higher-order features is present in almost all programming languages, from C to ML, Java and C♯. The search for good models and reasoning principles for, and language features that tame the complexity of, this combination goes back many decades. Recent years have seen a number of significant advances in our semantic understanding of state and encapsulation, including the development of separation logic, substructural type systems, models using parametric logical relations, and new bisimulation-based reasoning methods.

At the same time, concern about reliability, correctness and security of software has led to increased interest in tool and language support for specification and verification of realistic languages (for example JML and Spec♯), certified and certifying compilation, proof-carrying code, safe systems programming languages (such as Cyclone and CCured), and practical type systems capturing and controlling subtle aspects of state, such as ownership, effects, information flow and protocol conformance.

The seminar brought together researchers working on all aspects of state in programming languages, with the aim of developing closer links between the theoretical and applied lines of work, and laying groundwork for advances in the state of the art in both.

This is an exciting and important research area. Mathematically sound reasoning principles for state, combined with recent advances in program analysis and machine-assisted proof, have the potential to lead to improved programming languages, compilers, verification technology and even to new approaches to software deployment and operating system design. Using flexible, certified language-based approaches to encapsulation and security in place of hardware protection is a promising idea; systems such as Singularity and House (and, indeed, the JVM and CLR) have already taken steps in this direction.

Among the research challenges addressed at the seminar were:

- How do we integrate state and effects into dependently typed languages and how that might inform the design of tools and specification languages such as JML?
- What are the semantic foundations of existing logics and type systems for ownership, confinement, effects and permissions, and how may such foundations be used not only to understand and improve these systems, but also to relate them formally to one another?
- How should we model and reason soundly about concurrency, locks and transactions?
- How can we reason about controlled use of state at multiple levels of abstraction, for example in relating high-level, language-enforced restrictions to low-level guarantees on the behaviour of compiled code?
- What is the right mix of approaches to the control of state and other effects? How do we balance language design and type systems, automated verification tools and machine assisted proof?

Participation and Programme

The seminar brought together 45 researchers from Europe, the United States and Israel with interests and expertise in many different aspects of modelling and reasoning about mutable state. There were about 40 talks over the course of the week, comprising invited overview talks on particular topics, ordinary contributed talks (mostly on recent, completed work) and shorter, more informal talks on open problems and issues that arose during the week.

A major goal of the seminar was to forge links between researchers working on related problems from different perspectives. This was certainly achieved, with talks covering almost all combinations of semantic models, program logics, reasoning principles, program analyses and rich type systems for state in high-level and low-level, functional, imperative and object-oriented, sequential and concurrent, programming languages.

There was a clear sense of significant recent progress being made, on both the theoretical and applied aspects of reasoning about state. Separation logic, step-indexing and FM-domains are all recent developments with the first, in particular, having had a significant impact on much of the research discussed at the seminar. Fully automatic program analyses (such as pointer analyses) and machine-assisted proof have both advanced tremendously this century, to the extent that a number of projects are working on formal verifications of systems code that would have been considered impractical just a decade ago.

At the same time, some very interesting and useful technical interactions at the seminar concerned more elementary methodological questions. The definitions of, distinctions between and necessity of ghost variables, auxiliary variables, model variables and logic variables generated much discussion, as did the question of prescriptive versus descriptive readings of preconditions in triples and intensional versus extensional program properties.

This was an intense and productive week. With a relatively large number of particpants, most of whom wanted to speak, scheduled talks took up most of the days, including part of the traditional 'afternoon off' on Wednesday. Informal discussions continued into the night throughout the week.

2.2 Beyond the Finite: New Challenges in Verification and Semistructured Data

Seminar No. **08171** Date **20.04.–25.04.2008**
Organizers: Anca Muscholl, Ramaswamy Ramanujam, Michaël Rusinowitch, Thomas Schwentick, Victor Vianu

Objective: *Exploring the interaction of model checking and database static analysis techniques in the development of novel approaches to the verification of software systems handling data.*

The seminar ran as scheduled, from the morning of April 21 until mid-day on April 25. Since the seminar was centred on 7 themes (as outlined in the proposal, the event was structured accordingly: 7 invited talks (ranging from 60 to 90 minutes each) on each of the themes, and 25 presentations from participants (of 30 minutes each). An excursion to Trier on Wednesday afternoon provided a cultural interlude in a dense academic programme. With extensive interaction and discussions, the seminar was lively (with some heated debates) and highly educative. The central objective of the seminar was to look for common questions and techniques between research on verification and data, and to learn from each other; this was achieved satisfactorily.

Certain techniques emerged as being centrally useful. For instance, the use of well-quasi orderings on configurations of infinite space systems to prove termination of verification algorithms surfaced many times in discussions, whether it be in the context of parameterized systems, or in systems communicating by messages, or in software verification. Similarly, reachability studies on counter systems of various kinds was seen to have implications for a variety of contexts: analysis of data, reasoning about timed systems, and real arithmetic. The use of symbolic techniques and powerful theorems on term rewriting systems was illustrated in the study of security protocols as well as that of tree automata. Questions on logical characterizations and the need for logics over infinite alphabets was emphasized time and again, not only in the context of databases, but also in discussions on verification of parameterized systems and metric temporal logics. The need for decidable relational logics that combine quantification and order was talked of in the context of web services.

While logic and automata theory provided the lingua franca for the seminar, the discussions were not exclusively on these. There were presentations, especially in the context of software verification, on modelling issues as well as the use of (theorem proving and model checking) tools, and the pragmatics necessary in such a context.

The Dagstuhl setting, with its unique atmosphere of a castle set in an idyllic scene, backed by excellent organization and amenities, provided just the right tone for the seminar, allowing participants to focus on research interaction. With a mix of experienced and young researchers taking active part, we can confidently expect that the seminar will lead to new collaborations and applications of infinite-state systems benefiting both the verification and database areas.

List of talks:

Speaker	Title
Ahmed Bouajjani	Parameterized verification
Anca Muscholl	Communicating systems
Markus Müller-Olm	Some Aspects of Program Analysis Beyond the Finite
Axel Simon	No Strings Attached: Polyhedral Analysis of C String Buffers
Parosh Abdulla	Shape Analysis via Monotonic Abstraction
Philippe Schnoebelen	The complexity of lossy channel systems
Javier Esparza	SDSIRep : A reputation system based on SDSI
Luc Segoufin	Static analysis around XML
Claire David	Data Tree Patterns
Henrik Björklund	Conjunctive Query Containment over Trees
Victor Vianu	Verification of Data-Aware Web Services
Serge Abiteboul	Static Analysis of Active XML Systems
Mikolaj Bojanczyk	XPath with data
Patricia Bouyer	An Introduction to Timed Systems
Deepak D'Souza	Conflict-Tolerant Features
Felix Klaedtke	Runtime Monitoring of Metric First-order Temporal Properties
Stéphane Demri	Model checking memoryful linear-time logics over one-counter automata
Florent Jacquemard	Tree Automata Techniques for Infinite state verification of Security Protocols
S.P. Suresh	Unbounded data in security protocols
Stéphanie Delaune	Safely composing security protocols via tagging
Kumar Neeraj Verma	Single Blind Copying Protocols, and XOR
Michael Kaminski	Describing DTD by context-free grammars and tree automata over infinite alphabets
Christoph Löding	Tree automata with subtree comparisons
Pavel Krcal	R-automata
Florent Jacquemard	Closure of Hedge-Automata Languages by Hedge Rewriting
Bernard Boigelot	On the Sets of Real Numbers Recognized by Finite Automata in Multiple Bases
Martin Otto	Bisimulation Invariance over Transitive Frames
Jan Van den Bussche	A theory of stream queries
Shaz Qadeer	Back to the future: Revisiting precise program verification using SMT solvers
Supratik Chakraborty	Automatically Refining Abstract Interpretations for Program Analysis

2.3 Topological and Game-Theoretic Aspects of Infinite Computations

Seminar No. **08271** Date **29.06.–04.07.2008**
Organizers: Peter Hertling, Victor Selivanov, Wolfgang Thomas, Bill Wadge, Klaus Wagner

The theory of the infinite behaviour of continuously operating computing devices is of primary importance for several branches of theoretical and practical computer science. In particular, it is fundamental for the verification and synthesis of reactive systems like microprocessors or operating systems, for the understanding of data flow computation, and for the development of adequate mathematical foundations for exact real computation. The seminar brought together researchers from many different disciplines who are working on theoretical or practical aspects of infinite computations. In this summary we describe the topics, the goals, and the contributions of the seminar.

The theory of infinite computations develops different classifications of the corresponding mathematical objects, for example, languages of infinite words or trees as in automata theory, filters which transform infinite streams in dataflow computation, or functions on infinite words as in computability in analysis.

The study of computability with infinite objects, unlike that with finite objects, is intimately related to topology. It makes essential use of such classification tools as hierarchies as in descriptive set theory, reducibilities as in computability theory, and infinite games with perfect information. In the development of this area, ideas and techniques from different fields were employed, e.g. from topology, logic, game theory, Wadge reducibility, and domain theory. This work has produced, in particular, an impressive theory of the specification and verification of reactive systems, which is developing rapidly and is already of practical importance for the computer industry.

Despite the tight connections with several branches of mathematics, researchers from the cited fields often work independently and from time to time rediscover notions and techniques that have been used already in another field. The idea of this seminar is to bring together researchers from different fields, ranging from pure mathematics (descriptive set theory, domain theory, logic, symbolic dynamics) to theoretical and applied computer science (automata theory, computability in analysis, model checking), who work on topological and game-theoretic aspects of infinite computations, using similar notions and tools. The goal of the proposed seminar is to stimulate their work and establish ways for future cooperation and synergy. For example, while the theory of automata over infinite words already has direct applications in the theory of specification and synthesis of reactive systems, it would be very interesting to establish such direct applications also for the theory of infinite games.

The seminar was attended by 35 people from various scientific areas related to the topics of the seminar: by people working in computability, topology, descriptive set theory, automata theory, game theory, and by people working on applications in verification and synthesis of reactive systems and stream computations. In fact, many participants are

working in several of these areas and presented results connecting two or more of these areas. There were 5 overview talks of fifty minutes each and 26 contributed talks of thirty minutes each. Besides that, there were many discussions among smaller groups of people.

2.4 Distributed Verification and Grid Computing

Seminar No. **08332** Date **10.08.–14.08.2008**
Organizers: Henri E. Bal, Lubos Brim, Martin Leucker

The Dagstuhl Seminar on **Distributed Verification and Grid Computing** took place from 10.08.2008 to 14.08.2008 and brought together two groups of researchers to discuss their recent work and recent trends related to parallel verification of large scale computer systems on large scale grids. In total, 29 experts from 12 countries attended the seminar.

The computing power of computers has increased by a factor of a million over the past couple of decades. As a matter of fact, the development effort, both in industry and in academia, has gone into developing bigger, more powerful and more complex applications. In the next few decades we may still expect a similar rate of growth, due to various factors such as continuing miniaturization, parallel, and distributed computing.

With the increase in complexity of computer systems, it becomes even more important to develop formal methods for ensuring their quality and reliability. Various techniques for automated and semi-automated analysis and verification have been successfully applied to real-life computer systems. However, these techniques are computationally demanding and memory-intensive in general and their applicability to extremely large and complex systems routinely seen in practice these days is limited. The major hampering factor is the state space explosion problem due to which large industrial models cannot be efficiently handled unless we use more sophisticated and scalable methods and a balance of the usual trade-off between run-time, memory requirements, and precision of a method.

Recently, an increasing interest in parallelizing and distributing verification techniques has emerged, especially in the light of new multi-core architectures. On the other hand, grid computing is an emerging technology that enables large-scale resource sharing and coordinated problem solving within distributed systems. By providing scalable, secure, high-performance mechanisms for discovering and negotiating access to remote resources, Grid technologies promise to make it possible for scientific collaborations to share resources on an unprecedented scale that was previously impossible.

As such, grid computing promises to be an ideal partner to tackle huge verification tasks. However, while the verification community started to work out cluster based verification solutions, there is limited knowledge about grids. Similarly, while the grid community developed general grid infrastructure, highly optimized domain specific solutions first developed for the verification community might emerge to general solutions improving the state-of-the-art in grid computing.

The 17 talks given during this seminar were split into two main streams: distributed verification related topics and grid computing related presentations. Two overview talks

aimed at brief introductions to respective domains. The presentations were complemented by a huge number of informal discussions, in which, for example, a typical verification challenge for the GRID has been identified and addressed. Moreover, the scientific work was enriched also by several non-scientific activities like several table tennis matches and a traditional hike. The friendly atmosphere stimulated collaborations among the different communities that have already resulted in joint papers.

Chapter 3

Geometry, Image Processing, Graphics

3.1 Logic and Probability for Scene Interpretation

Seminar No. **08091** Date **24.02.–29.02.2008**
Organizers: Anthony G. Cohn, David C. Hogg, Ralf Möller, Bernd Neumann

From 25.2.2008 to 29.2.2008, the Dagstuhl Seminar 08091 "Logic and Probability for Scene Interpretation" was held in the International Conference and Research Center (IBFI), Schloss Dagstuhl. During the seminar, several participants presented their current research, and ongoing work and open problems were discussed.

The program consisted of 21 talks and discussions, attended by 40 participants from 12 countries. In accordance with the interdisciplinary nature of the workshop topic, the attendants represented distinct streams in Computer Vision, AI and Cognitive Science, in particular High-level Computer Vision, Logical Models in AI, Probabilistic Models in AI, Robotics, Multimedia Content Representation, and Cognitive Models.

The main goal of the workshop "to advance the use of logic-based knowledge representation and reasoning for scene interpretation (static and dynamic scenes), and explore possible ways for reconciling logics with probabilistic models" has been reached in several respects.

1. Several contributions showed the power but also the complexity of logical models for scene interpretation:

    - Brandon Bennett, Leeds University: Enhancing Tracking by Enforcing Spatio-Termporal Consistency Constraints
    - Francois Bremond, INRIA, Sophia Antipolis: Temporal scenarios for automatic video interpretation
    - Britta Hummel, Karlsruhe University: Description Logic for Intersection Understanding

- Ralf Möller, Technical University Hamburg-Harburg: A Logical Model for Multimedia Interpretation
- Fiora Pirri, University of Roma "La Sapienza": Lifting Models for Scene Interpretation
- Matthias Schlemmer, Technical University Vienna: Abstraction, ontology and task-guidance for visual perception in robots

2. Advanced combinations of probabilistic and logical models for scene interpretation were presented by

   - Bastian Leibe, ETH Zürich: Mobile Scene Understanding Integrating Recognition, Reconstruction, and Tracking
   - Dima Damen, Leeds University: Constrained Scene Interpretation - Sequences of Uncertain Events
   - Otthein Herzog, Bremen University: Qualitative Abstraction and Inherent Uncertainty in Scene Recognition
   - Pascal Hitzler, Karlsruhe University: Approximate Reasoning with OWL Ontologies
   - Manfred Jaeger, Aalborg University: Combining probabilistic graphical models and logic: the Markov Logic and Relational Bayesian Network approaches
   - Hans-Hellmut Nagel, Karlsruhe University: Toward Algorithmic Generation of Temporal-logic Representations for Driver Behavior from Legal Texts
   - Bernd Neumann, Hamburg University: Probabilistic Inferences in Compositional Hierarchies
   - Maria Petrou, Imperial College, London: Tower of Knowledge

3. The need for learning proved to be an important motivation for employing probabilistic models. Several contributions addressed learning aspects of scene interpretation:

   - Meg Aycinena, MIT, Cambridge: Learning Grammatical Models for Object Recognition
   - Luc De Raedt, Freiburg University: Statistical Relational Learning - a Logical Introduction
   - Paulo Santos, Centro Universitario di FEI, Sao Paulo: Assimilating knowledge from neuroimages in schizophrenia diagnostics
   - Sven Wachsmuth and Agnes Swadzba, Bielefeld University: Probabilistic modeling of spatial and temporal aspects of scenes for human-robot interaction

4. Cognitive aspects turned out to be less widely discussed. Many thanks go to Aaron Sloman who emphasized these apects in discussions and in his talk:

   - Aaron Sloman, The University of Birmingham: What are we trying to do, and how do logic and probability fit into the bigger picture? The participants were also reminded of the power of pattern-based approaches by the talks:

- Vaclav Hlavac, Czech Technical University, Prague: Image and Structure
- Diedrich Wolter, Bremen University: Qualitative Arrangement Information for Matching

In the final discussion, there was overwhelming agreement regarding the usefulness of the exchange of ideas and results in this workshop in general, and also regarding several insights in particular:

- A necessary third topic besides logics and probabilities in scene interpretation is learning. Scene interpretation appears to be an excellent topic for the learning community.
- There is urgent need for benchmarks, data sets and videos to be shared within the interdisciplinary scene interpretation community.

However, unanswered questions and unsolved problems by far outnumbered definite answers to the questions with which the workshop has started. Here are some of these open questions:Are there a priori roles for logics and probabilities in scene interpretation? What is the proper semantics for probabilities in scene inetrpretation? What is the impact of high-level and common sense knowledge in scene interpretation? What are the consequences of logics and probabilities for a scene interpretation system architecture? Thanks to the professional and friendly people of the Dagstuhl organisation! Many of the participants hope to convene again in Dagstuhl for discussions of a similar topic.

3.2 Geometric Modeling

Seminar No. **08221** Date **25.05.–30.05.2008**
Organizers: Gerald Farin, Stefanie Hahmann, Jörg Peters, Wenping Wang

The seminar succeeded in bringing together leading researchers from 17 countries to present and discuss radically different approaches to the challenge of modeling complex geometric phenomena on the computer. Acquisition, representation and analysis of 3-dimensional geometry call for the combination of technically complex and often interdisciplinary approaches that are grounded both in classical mathematics and computer science data structures and theory. Reflecting the dynamics of the field, the meeting included a number of junior researchers.

The presentations ranged from the application of graphics processing units, to graph techniques, to algebraic approaches, discrete geometry, combining clasical spline with new subdivision methods and leveraging the geometry of classical surfaces in areas as far as architecture and medical modeling.

The unique setting and combination of participants revealed and correlated a surprising number of new techniques and insights. New surface fitting methods addressed the intricate problem of resolving the shape where several different primary geometric features

merge (the multisided fair surface blending problem), and of representing thicker layers of surfaces (shells) as well as support functions on surfaces to support computations on manifolds. A key point, both for industrial applications and fundamental scientific inquiery, is the topological correctness of surfaces. In particular, topological and metric guarantees are needed when reconstructing and matching objects or extracting surfaces (geometric processing); for example to avoid singularities and self-intersections. Consistency leads to a reexamination and extension of difference (discrete) geometry. The need for interpreting and modifying existing geometry also motivates the search for finding parameterizations, both globally and via partition of space, The (highly nonlinear) challenge of not only measuring but of controlling intrinsic geometry was laid out in several talks. While some approaches concentrate on the mathematical challenge, an alternative is to emphasize the interface and human intervention. For simple geometry, determining geometry from constraints is another viable approach. Subdivision is an intriguing technique to flexibly represent geometry via refinement. This method bridges discrete and spline-based geometry. While key components have matured to the point where a survey talk laid out the fundamental structure, talks exposing work in progress and posing unresolved questions highlighted the need for further work. The detailed findings will appear in two refereed volumes, of full length research papers on Geometric Modeling to be published in a Springer journal.

Fittingly, the workshop was the setting for the *John Gregory* Award. The award, named after a pioneer of the field, honored the longterm innovative contributions to the field of Hartmut Prautzsch, Helmut Pottmann and Tom Sederberg.

3.3 Virtual Realities

Seminar No. **08231** Date **01.06.–06.06.2008**
Organizers: Guido Brunnett, Sabine Coquillart, Greg Welch

Virtual Reality (VR) is a multidisciplinary area of research aimed at interactive human-computer mediated simulations of artificial environments. Typical applications include simulation, training, scientific visualization, and entertainment. An important aspect of VR-based systems is the stimulation of the human senses—typically sight, sound, and touch—such that a user feels a sense of *presence (or immersion)* in the virtual environment. Different applications require different levels of presence, with corresponding levels of realism, sensory immersion, and spatiotemporal interactive fidelity.

To improve the sense of immersion, developers typically build *multi-modal systems*, i.e. systems that stimulate multiple human senses. The development of appropriate multi-modal devices (projection walls, head mounted displays, data gloves, force feedback arms, spatialized audio, etc.) to deliver multi-modal signals into the human sensory system is a basic aspect of VR technology. The fidelity and degree of immersion of these systems is a crucial factor in any user-perceived sense of presence.

In addition, the impression that can be achieved is limited by the richness of the model that describes the virtual environment. As such, another crucial aspect is VR modelling.

The scope of *VR modelling* typically goes beyond conventional (e.g. geometric) modelling in that it often integrates aspects related to the different modalities, for example visual, acoustic, or haptic characteristics.

If the environment includes dynamic objects, the movements of these objects must also be described. However, in contrast to conventional *simulation*, updates to the dynamic objects have to be computed in *real time*. To achieve high performance, today's VR systems often include some form of *parallel processing* of the simulation data. If compromises between the competing requirements of realistic behaviour and performance have to be made, *human factors* are taken into consideration.

Finally, a key ingredient of any VR system is the *interaction* between the human and the computer. For a sense of presence to be sustained, it is necessary that the objects of the virtual world appropriately react to the actions of the user. Some primary challenges in this context include the real-time tracking of the user's movements, and the design of intuitive devices for the user to control the virtual environment. In particular, many problems remain in the realization of VR systems that simultaneously support multiple users.

3.4 Statistical and Geometrical Approaches to Visual Motion Analysis

Seminar No. **08291** Date **13.07.–18.07.2008**
Organizers: Daniel Cremers, Bodo Rosenhahn, Alan Yuille

Motion analysis is central to both human and machine vision. It involves the interpretation of image data over time. It is crucial for a range of motion tasks such as obstacle detection, depth estimation, video analysis, scene interpretation, video compression and other applications. Motion analysis is difficult because it requires modeling the complicated relationships between the observed image data and the motion of objects and motion patterns (e.g. falling rain) in the visual scene.

This workshop was focused on critical aspects of motion analysis, including motion segmentation and the modeling of motion patterns. The aim was to gather researchers who are experts in the different motion tasks and in the different techniques used. These techniques include variational approaches, level set methods, probabilistic models, graph cut approaches, factorization techniques, and neural networks. All these techniques can be subsumed within statistical and geometrical frameworks.

We also involved experts in the study of human and primate vision. Primate visual systems are extremely sophisticated at processing motion so there is much to be learnt from studying them. In particular, we wanted to relate the computational models of primate visual systems to those developed for machine vision. Here, several researchers from the cognitive sciences and biologically inspired vision rounded off the overall high quality of presentations and discussions.

Another important component of the workshop was to develop datasets of image sequences with the associated motion ground truth. These datasets can be used as benchmarks to compare the performance of motion analysis models. They can also be used as data to train statistical models of motion analysis. Datasets with ground truth are being increasingly used in other aspects of machine vision but, at present, there are only very limited motion datasets (e.g. the Yosemite sequence). Here we would like to point out the empeda test sequences, available at `http://www.citr.auckland.ac.nz/6D/` or several discussions about the current optic flow benchmark at middlebury `http://vision.middlebury.edu/flow/`

We made the seminar interactive with plenty of time for discussion. The participants were also encouraged to exchange different modeling techniques and research experiences. We have also identified outstanding unsolved problems in motion analysis and discussed them during highly active group meetings. The participants of this seminar enjoyed the atmosphere and the services at Dagstuhl very much. The quality of this center is unique. It was also a pleasure for the organizers to be involved in a radio interview of the *Computer Club Zwei*. The interview about the Dagstuhl seminar (in German) is available at `http://www.audioads.de/files/14316/CC-Zwei-118-low.mp3`

There will be an edited book (within Springer's series on Lecture Notes of Computer Science, LNCS) following the seminar, and all seminar participants have been invited to contribute with chapters. The deadline for those submissions is in November 2008 (allowing to incorporate results or ideas stimulated by the seminar), and submissions will be reviewed (as normal). Expected publication date is 2009.

3.5 The Study of Visual Aestetics in Human-Computer Interaction

Seminar No. **08292** Date **13.07.–16.07.2008**
Organizers: Marc Hassenzahl, Gitte Lindgaard, Axel Platz, Noam Tractinsky

This seminar intended "to gather a group of about 25-30 participants who will exchange ideas, views, and case study results that address the seminar's themes." We aimed at discussing methodologies and measures in the study of visual aesthetics in HCI, to explore design antecedents of aesthetic interactive systems, as well as consequences of aesthetic design or aesthetic experience in HCI. We anticipated that the outcome of the seminar "will contribute to clarifying the concept, provide an overview of existing practical resources such as measurement scales, solidify the body of knowledge in this area, and generally spark interest in aesthetics in the HCI community."

21 people participated in the seminar. This seminar explored various aspects of the study of visual aesthetics in human-computer interaction (HCI). The growing attention that this field is gaining from the HCI community is manifested by the increasing rate of published-, and in-progress research, and by the emphasis espoused by usability- and Usability Experience (UX) practitioners on the importance of aesthetic design.

We identified a set of research challenges that this emerging field needs to discuss. These can be broadly classified into four categories, which can be depicted as belonging to two major axes. The first deals with theory building vs. measurement. It includes the development of theoretical and conceptual foundations of the field on the one hand, and the identification and development of measures and research methods that are appropriate for studying it on the other hand. The second axis contrasts antecedents of aesthetic design with its consequences. The context of the aesthetic experience and the contingencies that affect reactions to aesthetic interactive products and applications fall between those two axes.

One of the seminar's goals was to collect examples and case studies of visual aesthetics in interactive systems. Some participants have discussed the implications of aesthetics for the design of interactive systems and demonstrated the applications of aesthetic principles to design (Löwgren). Fishwick places aesthetics within the context of ubiquitous computing. He demonstrated the application of aesthetics to software representation (e.g., in Second Life) with the ultimate intent of popularizing software engineering.

Major debates and future directions

The seminar concluded with discussions of research topics that need to be addressed further. These include the following:

- Considering visual aesthetics as a dynamic process:
- Considering the transition of computing from an interaction paradigm to the computing-as-medium paradigm and the consequences of such a transition (Nake).
- The educational implications of the importance of visual aesthetics in HCI (Sutcliffe). Only very few in the HCI community are "visual designers". Who will teach aesthetics to most of the community and how (educational programs)?
- Is the aesthetic experience and resulting evaluation cognitive or emotional? Perceptual and sensory or reflective and intellectual? Necessarily wholistic or also decomposable?

Finally, it was recommended that in order to facilitate further research efforts and improve communication among group members, we should develop a web site and/or a wiki allowing people to share resources and ideas.

3.6 Representation, Analysis and Visualization of Moving Objects

Seminar No. **08451** Date **02.11.–07.11.2008**
Organizers: Wolfgang Bitterlich, Claus Brenner, Jörg-Rüdiger Sack, Monika Sester, Robert

Weibel

This workshop has been organized as a successor to five preceding ones that were centered around topics of computational cartography and geographic information systems. The major goal has been to bring together the still small, but fast growing, research community that is involved in developing better computational techniques for spatio-temporal object representation, data mining, and visualization of massive amounts of moving object data. The participants included experts from fields such as computational geometry, spatial databases, GIScience, photogrammetry, spatial statistics, and knowledge discovery and data mining. The majority of participants were from academic institutions, some from government agencies. Several industry representatives were invited, but unfortunately were unable to attend the seminar. However, one of the organizers is from ESRI Inc., the leading GIS company. The seminar has lead to a fruitful exchange of ideas between different disciplines, to the creation of new collaborations, and to recommendations for future research directions.

Mobility is key in a globalized world where people, goods, data and even ideas move in increasing volumes at increasing speeds over increasing distances. Understanding of mobility patterns and movement behaviors will increasingly become a key factor for success in many businesses such as location-based services (LBS), advertising, and logistics. It will be essential for the prediction and monitoring of individual and group behaviors in response to and mitigation of security threats over short and long time scales. Traffic simulation can greatly benefit from the analysis of movement data, for example through better estimates of key parameters. Finally, mobility patterns of endangered species are prerequisites to devising protective measures in nature conservation and successfully managing interactions between tourism and conservation.

Dynamic geographic objects may include phenomena as diverse as people, animals, vehicles or hurricanes, or the pathways of diseases such as SARS. Data recording the trajectories of MOs stem from a variety of sources, ranging from radio telemetry and GPS to mobile telecommunication devices and geo-sensor networks. Despite the diversity of sources and moving objects involved, movement data have in common that they represent joint recordings of spatial and temporal dimensions and capture motion in trajectories. Hence, it is possible to develop powerful methods for knowledge discovery (KD), by data mining and visual analytics, in movement data that can be suited to the needs of diverse application domains. KD of movement patterns may be performed in real-time (e.g. in geo-sensor networks) or off-line and a posteriori. Furthermore, the purpose may be explanation – e.g., discovering and explaining behavioral patterns of an animal to devise better protection measures. Or it may be prediction - e.g., predicting the next move of an object to trigger personalized information feeds in a mobile information system or issue a warning.

The problem is that we are only at the beginning of the evolution of a new research domain. Hence, despite the collection of increasing amounts of data in recent years which can be used in movement tracking, only few methods exist today which have been demonstrated to effectively exploit very large volumes of movement data at different spatial and temporal scales. Recent years have however seen increased interest in the development of such methods, which are currently being developed in a rather piecemeal fashion, and have yet

to migrate from research to demonstrate convincing social and commercial benefits.

Outcomes of the seminar include a collection of abstracts, presentations (slides) and some papers surveying the current state of the art in this field and latest research initiatives (available on the website `http://www.dagstuhl.de/08451/Materials/`). Similar to the previous seminars of this series, it is expected that new partnerships and collaborations between multi-disciplinary groups will form, further advancing this field with the inclusion of emerging topics. As a first concrete initiative in this direction, a proposal for a European COST Action on "Knowledge Discovery from Moving Objects" was launched (`http://www.cost.esf.org`). The preliminary proposal has since been accepted and the full proposal will be submitted in January 2009. The Dagstuhl seminar provided plenty of useful inputs for the full proposal, as well as links to researchers who are interested to participate in the proposed COST Action. Most of the Europeans among the participants of Dagstuhl seminar 08451 will also participate in the COST Action.

Another important result of the seminar is the list of past and future research directions that was compiled during the final round of breakout discussions on the last day of the seminar.

Chapter 4

Artificial Intelligence, Computer Linguistic

4.1 Recurrent Neural Networks – Models, Capacities, and Applications

Seminar No. **08041** Date **20.01.–25.01.2008**
Organizers: Luc De Raedt, Barbara Hammer, Pascal Hitzler, Wolfgang Maass

Artificial neural networks (FNNs) constitute one of the most successful machine learning techniques with application areas ranging from industrial tasks up to simulations of biological neural networks. Recurrent neural networks (RNNs) include cyclic connections of the neurons, such that they can incorporate context information or temporal dependencies in a natural way. Spatiotemporal data and context relations occur frequently in robotics, system identification and control, bioinformatics, medical and biomedical data such as EEG and EKG, sensor streams in technical applications, natural speech processing, analysis of text and web documents, etc. At the same time, the amount of available data is rapidly increasing due to dramatically improved technologies for automatic data acquisition and storage and easy accessibility in public databases or the web, such that high-quality analysis tools are essential for automated processing and mining of these data. Since, moreover, spatiotemporal signals and feedback connections are ubiquitous when considering biological neural networks of the human brain, RNNs carry the promise of efficient biologically plausible signal processing models optimally suited for a wide area of industrial applications on the one hand and an explanation of cognitive phenomena of the human brain on the other hand.

However, simple feedforward networks without recurrent connections and with a feature encoding of complex spatiotemporal signals which neglects structural aspects of data are still the preferred model in industrial or scientific applications, disregarding the potential of feedback connections. This is mainly due to the fact that traditional training of RNNs, unlike FNNs and backpropagation, faces severe for RNNs suffers from numerical barriers, a formal learning theory of RNNs in the classical sense of PAC learning does hardly exist,

RNNs easily show complex chaotic behavior which is complicated to manage, and the way how humans use recurrence to cope with language, complex symbols, or logical inference is only partially understood.

The aim of the seminar was to bring together researchers who are involved in these different areas and recent advances, in order to further the understanding and development of efficient, biologically plausible recurrent information processing, both in theory and in applications. Although often tackled separately, these aspects, the investigation of cognitive models, the design of efficient training models, and the integration of symbolic systems into RNNs, severely influence each other, and they need to be integrated to achieve optimum models, algorithmic design, and theoretical background.

Overall the presentations and discussions revealed that RNNs constitute a highly diverse and evolving field which opens interesting perspectives to machine learning for structures in the broadest sense. It still waits with quite a few open problems for researchers, a central problem being efficient and robust learning methods for challenging domains.

4.2 Perspectives Workshop: Theory and Practice of Argumentation Systems

Seminar No. **08042** Date **20.01.–23.01.2008**
Organizers: Jürgen Dix, Simon Parsons, Henry Prakken, Guillermo Simari

Executive Summary

The group (28 participants from 12 countries) convened in Dagstuhl in January 2008, for a three day meeting.

The main objective of this Perspectives Workshop was to bring together researchers and developers involved in *Argumentation Systems and Related Areas* who have been producing significant contributions towards the progress in theoretical and pragmatic aspects of this form of reasoning.

During the preparation of the meeting the organizers decided to separate the group into four areas:

- Argumentation and the Semantic Web,
- Argumentation and Decision Support in Application Areas,
- Argumentation and Multi-Agent Systems, and
- Argumentation and Social Networks.

For each of these areas, two participants were requested to give talks to the plenary as a foundation for the discussions that followed.

After completing the presentations, the participants attended separate meetings according to the areas mentioned. The groups reported the conclusions to the plenary after each meeting.

During the last plenary, because of the results obtained sofar and the enthusiasm of the participants, it was decided that a publication will be produced for each of the four areas.

The seminar meetings enabled the interaction, cross-fertilization, and mutual feedback among researchers and practitioners from the different, but related, areas providing the opportunity to discuss diverse views and research findings.

Introduction

The first articles on argumentation in computer science appeared circa 20 years ago. Since then we have seen great advances, establishing a solid theoretical basis, a broad canvas of applications, and, most recently, some realistic implementations. The field has gone from infancy to maturity, and the initial questions that researchers posed—"how do we do this?", "what is it good for?" and "how do we implement it'?"—are mostly answered.

We believe that argumentation systems are at a point where they can be of practical use and industrial importance. Whereas once argumentation systems only existed as *theoretical* models, there are now a number of software *implementations* which make it practical to build software systems that have argumentation at their core. Such implementations are not only research tools. The EU-funded ASPIC project (`http://www.argumentation.org/`) (on which Prakken and Parsons have been working) has developed industrial-strength Java components that implement an argumentation system, and which will be provided under an open-source license.

Such components will make it possible to easily construct software systems that make use of the power of argumentation in the service of other functionality (rather than as an end in itself). Such a move is currently being made by the EU-funded *ArguGRID* project (`http://www.argugrid.eu/`), on which Kakas and Toni are working. This project uses argumentation at the core of its mechanisms to handle GRID computing resources.

There are two follow-up publications for this seminar:
Jürgen Dix: Dagstuhl Manifesto: Perspectives Workshop Theory and Practice of Argumentation Systems 08042,
Informatik Spektrum 32.2009,2, p. 181-182.
Jürgen Dix, Simon Parsons, Henry Prakken, Guillermo Simari: Research challenges for argumentation.
Computer Science: Research and Development 23.2009,1, p. 27-34.

4.3 Programming Multi-Agent Systems

Seminar No. **08361** Date **31.08.–05.09.2008**
Organizers: Rafael Bordini, Mehdi Dastani, Jürgen Dix, Amal El Fallah-Seghrouchni

Intelligent agents and multi-agent systems (MAS) play an important role in today's software development. In fact, they constitute a new and interesting paradigm to implement complex systems, by offering relevant abstractions for the engineering of such intricate type of software. Several application domains, some at industrial level, take benefit from MAS technology. For almost two decades, the MAS community has developed and offers a large and rich set of concepts, architectures, interaction techniques, and general approaches to the analysis and the specification of MAS.

One of the main challenges of the MAS community in recent years has been to combine the existing practical tools and theoretical approaches in order to provide practitioners with mature programming languages and platforms that help them to design, implement, and deploy efficiently a new generation of complex software built as MAS. The organisers of this seminar, and also the participants, share the conviction that the success of agent-based systems can only be guaranteed if expressive programming languages and well-developed platforms are available, so that the concepts and techniques of multi-agent systems can be easily and directly used in practice.

The aim of this seminar was to bring together researchers from both academia and industry for bridging this gap and identifying interesting lines of research within multi-agent systems. In this respect, the seminar topic is also very relevant for industrial research and development.

The seminar concluded with very positive results, both in regards to the scientific content as well as a networking opportunity. Meanwhile, we have secured a special issue of the Journal of Autonomous Agents and Multi-Agent Systems for the best papers based on talks given by the participants of this Dagstuhl seminar. An internal call for papers has been issued and we aim for publication in the second half of 2009.

In summary, it is our impression that the participants enjoyed the great scientific atmosphere offered by Schloss Dagstuhl, and the technical programme of the seminar. We are grateful for having had the opportunity to organise this fruitful seminar, specially because it was another Dagstuhl seminar which helped us, six years ago, to start an outstanding international cooperation in the domain of Multi-Agent Programming. Special thanks are due to the whole Dagstuhl staff for their assistance in the organisation and the running of the seminar.

4.4 Planning in Multiagent Systems

Seminar No. **08461** Date **09.11.–14.11.2008**
Organizers: Jürgen Dix, Edmund H. Durfee, Cees Witteveen

Planning in Multiagent Systems, or Multiagent Planning (MAP for short), considers the planning problem in the context of multiagent systems. It extends traditional AI planning to domains where multiple agents are involved in a plan and need to act together.

Research in multiagent planning is promising for real-world problems: on one hand, AI planning techniques provide powerful tools for solving problems in single agent settings;

on the other hand, multiagent systems, which have made significant progress over the past few years, are recognized as a key technology for tackling complex problems in realistic application domains.

The motivation for this seminar is thus to bring together researchers working on these different fields in AI planning and multiagent systems to discuss the central topics mentioned above, to identify potential opportunities for coordination, and to develop benchmarks for future research in multiagent planning.

The seminar addressed, through presentations, panel discussions, and break-out sessions, the following topics:

- Coordination and Task Allocation
- Dynamic and Temporal Planning
- Robust Planning

It is our impression that the participants enjoyed the great scientific atmosphere offered by Schloss Dagstuhl, and the scientific program which offered them ample opportunities for discussion. We are grateful for having had the opportunity to organize this fruitful seminar. Special thanks are due to the whole Dagstuhl staff for their assistance in the organization and the running of the seminar.

Chapter 5

Programming Languages, Compiler

5.1 Scalable Program Analysis

Seminar No. **08161** Date **13.04.–18.04.2008**
Organizers: Florian Martin, Hanne Riis Nielson, Claudio Riva, Markus Schordan

Motivation and Introduction

As the volume of existing software in the industry grows at a rapid pace, the problems of understanding, maintaining, and developing software assume great significance. A strong support for analysis of programs is essential for a practical and meaningful solution to such problems. Static analysis tools can make a huge impact on how software is engineered but in an industrial context research must be properly balanced with a focus on deployment of analysis tools.

Our goal was to bring together researchers from academia and industry to discuss the strengths and weaknesses of state-of-the-art program analysis technology for industrial-sized software. To achieve that goal the seminar gathered 38 participants from 9 companies and 23 academic/research institutions.

Proceeding of the Seminar

The seminar started with two sessions on technological research challenges in industry and continued with more theoretical sessions, covering scalable shape and pointer analysis, concurrent program analysis, source-level and scalable instruction-level analysis, scalable path conditions, state-of-the-art techniques in abstract interpretation for scalability, and type systems. Overall we had 28 talks, with a good diversity in topics, showing the many research directions and applications of program analysis, also covering program synthesis with sketching, using machine learning for scalable analysis, analysis for architecture reconstruction, and data-flow analysis for multi-core architectures.

A high number of participants, about 50 %, had also implemented their analysis technique in a tool. Some participants had already expressed some interest in tool sessions before

the seminar, but there were also concerns mentioned that tool sessions can be a real show-stopper if too lengthy. To encourage lively discussions, the tool sessions were preceded by discussion groups starting on Tuesday. Each group consisted of 4–7 people, of which at least two people had already indicated their interest in presenting a tool. Each group was asked to define a set of challenging questions to be asked about a program analysis tool. On the the next day, Wednesday, we selected in a one hour discussion session in a democratic process a subset of the proposed questions to be answered by every tool presenter on Wednesday/Friday.

The idea was that each presentation of a tool would start by answering those selected 7 questions, providing some basis for comparing the tools, and to create a common frame for each tool presentation. The presentations were kept to a minimum in time, about 15 mins, and the presenters were asked to focus on the most impressive analysis feature of the tool.

This format with preceding working groups worked out well, as it further increased the interest in the tool sessions. We eventually scheduled 15 tool & infrastructure presentations on Thursday and Friday.

The following 10 tools were presented:

- AiT (Florian Martin - AbsInt)
- ASTREÉ (Patrick Cousot - ENS - Paris)
- CodeSonar (David Melski - GrammaTech Inc.- Ithaca)
- Columbus (Arpad Beszedes - University of Szeged)
- EspC Concurrency Toolset (Jason Yue Yang - Microsoft Corp. - Redmond)
- Havoc (Thomas Ball - Microsoft Corp. - Redmond)
- Parfait (Cristina Cifuentes - Sun Microsystems Laboratories - Brisbane)
- PluggableTypes (Michael D. Ernst , MIT - Cambridge)
- SAFE (Eran Yahav - IBM TJ Watson Research Center - Hawthorne)
- Space Invader (Dino Distefano - Queen Mary College - London)

The presented 5 infrastructures were interesting to the participants because they provided a basis for building analysis tools:

- Bauhaus (Rainer Koschke - Universität Bremen)
- LLNL-ROSE (Daniel J. Quinlan, Lawrence Livermore National Laboratory)
- LLVM (Vikram Adve - Univ. of Illinois - Urbana)
- SATIrE (Markus Schordan, TU Vienna)

- WALA (Steve Fink, IBM TJ Watson Research Center - Hawthorne)

An interesting experiment during the seminar was that the entire C++ source code of the infrastructure LLNL-ROSE was used as test case for the Columbus tool and the analysis results were presented in a tool session on the next day.

The tool presentations were mixed, some were hands-on with live presentations, others gave an in-depth view on the tool's capabilities without actually running it. The presenters were asked to keep the presentation short, 10–15 minutes, and focus on some specific interesting feature of the tool. Overall, the format worked well in keeping the audience's attention with every tool. It also encouraged people from academia and industry to get into discussions after the sessions regarding possible future applications and extensions of the tools.

Fun and Art

On Wednesday evening the participants took a well deserved break and visited the nearby winery. A tour on the beautiful hills of the winery with an overview of the history and state-of-the-art of wine making preceded the dinner. Discussions about the selection of the picture that the seminar participants would try to donate a share of, started at the dinner in the winery. Eventually Werner Rauber's picture "Rotkohl 2" ("Red Cabbage") was chosen, and 2 shares were donated by the participants. Since the juice of red cabbage can be used as a home-made pH indicator, turning red in acid and blue in basic solutions, it was considered to be a good analogy to having an indicator for the tradeoff between precision and run-time of an analysis.

Achievements of the Seminar

The seminar showed how broad the field of program analysis has grown over the years. Traditionally used in optimizing compilers, program analysis has turned into a major discipline with techniques and commercial tools supporting understanding, maintaining, and engineering of software. It often turned out that an in-depth discussion of scalability requires further investigations. The scalability of analysis techniques is a major issue as the size of software systems is rapidly growing and the automatic analysis of those systems is becoming yet more important in future. Many questions were raised about scalability - to address the scale of today's systems, analyses will have to be run as parallel programs in future, posing themselves as problem of being scalable on multiple cores, but also whether it can be applied to multiple parts of a system which may differ in structural properties of the code or even in used programming languages.

Raising awareness about the many faces of scalability is the major achievement of the seminar. As the seminar progressed, increasingly more questions about scalability were raised, mostly asking for more extensive evaluations of the methods & tools in future. Discussions about the need and development of specific benchmarks for scalability were started and agreed to be continued past the seminar by different groups of the seminar. Carefully systematically designed sets of test cases, accompanied by test cases from industry, was considered a good setting for evaluating scalability.

Participation Statistics

The seminar was attended by 38 people from 11 countries, which were: Australia (1), Austria (4), Denmark (1), Finland (1), France (4), Germany (7), Great Britain (4), Hungary (1), Ireland (1), Korea (1), U.S.A. (13).

From Academia/Research Institutions 26 people were attending the seminar, whereas from industry 12 people participated, representing 9 companies, which were: AbsInt (Saarbrücken), Airbus (Toulouse), Berner & Mattner Systemtechnik (Berlin), Google Inc. (Mountain View), Grammatech Inc. (Ithaca), IBM Research Center (Hawthorne), Microsoft (Redmond), Nokia (Helsinki), Robert Bosch GmbH (Stuttgart), Sun Microsystems Lab (Brisbane).

Conclusion

The Dagstuhl seminar on "Scalable Program Analysis" was a tremendous success with many fruitful discussions and new questions being raised. Several connections between industry and academia were formed and showed all signs that they will find their continuation after the seminar. The seminar also showed that program analysis and the question about scalability is cross-cutting many different communities. This was also reflected by the diversity of the techniques presented in talks and the tools as well. The cooperation of industry and academia, as being encouraged by many funding programs these days, will further help both sides, to focus on new methods for addressing, characterizing, and comparing scalability.

Chapter 6

Software Technology

6.1 Software Engineering for Self-Adaptive Systems

Seminar No. **08031** Date **13.01.–18.01.2008**
Organizers: Betty H. C. Cheng, Rogerio de Lemos, Holger Giese, Paola Inverardi, Jeff Magee

The simultaneous explosion of information, the integration of technology, and the continuous evolution from software- intensive systems to ultra-large-scale (ULS) systems requires new and innovative approaches for building, running and managing software systems. A consequence of this continuous evolution is that software systems must become more versatile, flexible, resilient, dependable, robust, energy-efficient, recoverable, customizable, configurable, or self-optimizing by adapting to changing operational contexts and environments. The complexity of current software-based systems has led the software engineering community to look for inspiration in diverse related fields (e.g., robotics, artificial intelligence) as well as other areas (e.g., biology) to find new ways of designing and managing systems and services. In this endeavour, the capability of the system to adjust its behaviour in response to its perception of the environment and the system itself in form of self-adaptation has become one of the most promising directions.

The topic of self-adaptive systems has been studied within the different research areas of software engineering, including, requirements engineering, software architectures, middleware, component-based development, and programming languages, however most of these initiatives have been isolated and until recently without a formal forum for discussing its diverse facets. Other research communities that have also investigated this topic from their own perspective are even more diverse: fault-tolerant computing, distributed systems, biologically inspired computing, distributed artificial intelligence, integrated management, robotics, knowledge-based systems, machine learning, control theory, etc. In addition, research in several application areas and technologies has grown in importance, for example, adaptable user interfaces, autonomic computing, dependable computing, embedded systems, mobile ad hoc networks, mobile and autonomous robots, multi-agent systems, peer-to-peer applications, sensor networks, service-oriented architectures, and ubiquitous computing.

It is important to emphasise that in all the above initiatives the common element that enables the provision of self-adaptability is software because of its flexible nature. However, the proper realization of the self-adaptation functionality still remains a significant intellectual challenge, and only recently have the first attempts in building self-adaptive systems emerged within specific application domains. Moreover, little endeavour has been made to establish suitable software engineering approaches for the provision of self- adaptation. In the long run, we need to establish the foundations that enable the systematic development of future generations of self-adaptive systems. Therefore it is worthwhile to identify the commonalities and differences of the results achieved so far in the different fields and look for ways to integrate them.

The development of self-adaptive systems can be viewed from two perspectives, either top-down when considering an individual system, or bottom-up when considering cooperative systems. Top-down self-adaptive systems assess their own behaviour and change it when the assessment indicates a need to adapt due to evolving functional or non-functional requirements. Such systems typically operate with an explicit internal representation of themselves and their global goals. In contrast, bottom-up self-adaptive systems (self- organizing systems) are composed of a large number of components that interact locally according to simple rules. The global behaviour of the system emerges from these local interactions, and it is difficult to deduce properties of the global system by studying only the local properties of its parts. Such systems do not necessarily use internal representations of global properties or goals; they are often inspired by biological or sociological phenomena. The two cases of self-adaptive behaviour in the form of individual and cooperative self-adaptation are two extreme poles. In practice, the line between both is rather blurred, and compromises will often lead to an engineering approach incorporating representatives from these two extreme poles. For example, ultra large-scale systems need both top-down self-adaptive and bottom-up self-adaptive characteristics (e.g., the Web is basically decentralized as a global system but local sub-webs are highly centralized). However, from the perspective of software development the major challenge is how to accommodate in a systematic engineering approach traditional top-down approaches with bottom-up approaches.

6.2 Combining the Advantages of Product Lines and Open Source

Seminar No. **08142** Date **02.04.–05.04.2008**
Organizers: Jesus Bermejo, Björn Lundell, Frank van der Linden

From April 2 to 5, the Dagstuhl Seminar 08142 “Combining the Advantages of Product Lines and Open Source” was held in the International Conference and Research Center (IBFI), Schloss Dagstuhl. During the seminar, several participants presented their current research, and ongoing work and open problems were discussed. The first section describes the seminar topics and goals in general.

Practitioners and researchers have already identified the potential cross-fertilisation benefits of software product lines and open source software development.

Product line development is being established as an important way of producing software in companies. This ensured an efficient way to obtain a variety of products. It is also an important contemporary research issue, as indicated by the recent special issue of Communications of the ACM, December 2006.

Using open source software appears to be a profitable way to obtain good software. This is a result of several of its properties, ranging from effective feedback to the openness of the source. At a first glance open source and product line practices are conflicting.

This workshop aims to find ways to overcome these conflicting practices and how to profit from both approaches. Product line engineering can improve in agility and fast feedback improving the quality of the result. Open source software development can profit for variability management techniques, developed in product line engineering to improve the efficiency to deal with a diversity of configurations.

In-depth description of the topic

Product line engineering sets up several processes to enable efficient management of reuse and variability. As a consequence, a large initial investment is done, and a lot of rules are put down to enable ease of traceability and the insurance that different sets of requirements lead to a fast production of diverse products. In principle product line engineering is a top-down process.

This process is often executed in large developments within companies involving distributed development. This distribution factor makes product-line engineering complex. How to ensure that the rules are followed in a distributed organisation.

Development of open source software typically starts bottom-up. As such, it has in many cases shown to be an effective way to get a large group of people working together on the same topic, building high quality software. Often, the people did not met before, but they share the interest for the code they are dealing with.

Open source development is intrinsically distributed, and its practices may a lot to offer to traditional distributed development. In fact, the open source development model has been adopted by companies as "inner source development", as a way to improve their distributed development.

Product line practices involving the planning and management of variability and reuse may be useful for open source communities and companies adopting open source practices. Many tools available within the open source world deal with versions and configurations. However they still do not perform effectively and friendly at the level of a user wanting to configuring the software.

The seminar was visited by 15 participants from the product line world and from open source communities. Most of them presented a position paper giving rise to a lot of discussion.

6.3 Perspectives Workshop: Model Engineering of Complex Systems

Seminar No. **08331** Date **10.08.–13.08.2008**
Organizers: Uwe Aßmann, Jean Bezivin, Richard Paige, Bernhard Rumpe, Douglas C. Schmidt

Complex systems are hard to define. Nevertheless they are more and more frequently encountered. Examples include a worldwide airline traffic management system, a global telecommunication or energy infrastructure or even the whole legacy portfolio accumulated for more than thirty years in a large insurance company. There are currently few engineering methods and tools to deal with them in practice. The purpose of this Dagstuhl Perspectives Workshop on Model Engineering for Complex Systems was to study the applicability of Model Driven Engineering (MDE) to the development and management of complex systems.

MDE is a software engineering field based on few simple and sound principles. Its power stems from the assumption of considering everything – engineering artefacts, manipulations of artefacts, etc. – as a model.

Our intuition was that MDE may provide the right level of abstraction to move the study of complex systems from an informal goal to more concrete grounds.

In order to provide first evidence in support of this intuition, the workshop studied different visions and different approaches to the development and management of different kinds of complex systems.

6.4 Evolutionary Test Generation

Seminar No. **08351** Date **24.08.–29.08.2008**
Organizers: Holger Schlingloff, Tanja Vos, Joachim Wegener

The "Evolutionary Test Generation" Dagstuhl seminar was held from September 24th to September 29th 2008. The organisation of the seminar was initiated by the EvoTest project, a project funded by the European Commission under the contract number IST-33472. The goal of our seminar was to bring together researchers from the software testing and evolutionary algorithms communities for the discussion of problems and challenges in evolutionary test generation. This goal has been satisfactorily met and has led to a comprehensive list of open problems and challenges identified and discussed during the seminar. The seminar has been attended by 33 people: 30 were researchers from all over the world working on evolutionary testing, test generation and/or evolutionary computing; 3 were industrial participants with experience and feedback from real-life challenges: Microsoft, IBM and Berner & Mattner.

Systematic testing is the most widely used method to ensure that a program meets its specification. The effectiveness of testing for quality assurance largely depends on the chosen test suite. Currently, test suites are constructed either manually or semi-automatically from the program code or program specification. For large systems, however, manual test case construction is tedious and error-prone, whereas semi-automatic procedures often achieve only insufficient coverage. Therefore, new methods for the automated generation of "good" test suites are necessary.

Evolutionary adaptive search techniques offer a promising perspective for this problem. Genetic algorithms have been investigated for complex search problems in various fields. Their basic principles are selection, mutation, and recombination. These principles can be beneficially applied to the automated generation and optimisation of test suites, both from code (white-box testing) and specification (black-box testing). However, to make this approach successful in practice, a lot of problems remain to be solved: the question of adequate testing objectives, coverage and reliability measures, representation issues for test cases and test suites, seeding, recombination and mutation strategies, and others.

The future challenges identified at the Dagstuhl seminar have been categorized as follows:

- Theoretical foundations
- Search Technique improvements
- New testing objectives
- Tool environment/testing infrastructure
- New application areas

6.5 Perspectives Workshop: Science of Design: High-Impact Requirements for Software-Intensive Systems

Seminar No. **08412** Date **08.10.–11.10.2008**
Organizers: Matthias Jarke, Kalle Lyytinen, John Mylopoulos

NSF-funded Science of Design (SoD) Initiative in North America has tried to establish principles of a science of design between the traditional natural science and social science methodologies. This is highly relevant in a scientific climate whose publications tend to accept formal results or formally conducted empirical studies of existing systems much more easily than design- and innovation-oriented research. It is therefore necessary to define more clearly what is "good" design research and, in particular, to align better the research strategies and the current and future needs of practice.

Within the design science initiative, two workshops in the US and Europe will discuss the relationship between the practice and the research in *requirements engineering* with the aim of identifying "high impact requirements" research, i.e. areas which have high

importance and relevance in current and future practice but have received little research interest in the past. Another goal is to bring together representatives of the broad variety of fragmented disciplines in which requirements play an important role, but where little communication exists.

The starting point of the workshops was a large-scale empirical study in over thirty organizations concerning the present relationships between research and practice in RE. The results were in part quite encouraging: Many RE research results have found their way at least partially into practice. But while uptake of scientific RE results has been more successful than often perceived, the transfer of new practice problems to research has been less successful. While the above-mentioned RE results were derived from problems of the 80's, many new challenges have arisen in the meantime for which relatively little RE research exists. Moreover, isolated ideas from different disciplines have not yet found their way into coherent theories. It is in these areas, where RE research with high impact could emerge in the next few years.

As a result of the first workshop held in Cleveland, Ohio, June 3-6, 2007, four key requirements principles have been identified that need deeper investigation: intertwine requirements and contexts, evolve designs and ecologies, manage through architectures, and recognize and mitigate against design complexity. These issues have been used as anchors for the second workshop held in Europe, namely Dagstuhl. The idea was to deepen the discussion on some particularly important challenges, but also to include more strongly the perspectives of European companies and researchers, such as the stronger focus on enterprise software architectures and formal semantics of service-oriented business software architectures, but also inter-cultural management aspects ranging from eInclusion aspects to offshoring. Accordingly, the Dagstuhl workshop has again brought together 22 representatives from research and industry with various different backgrounds in requirements engineering, software and IS development, human computer interaction (HCI), computer-supported cooperative work (CSCW), organization science, business processes, and service orientation. The program has interleaved a number of plenary keynotes and panel discussions from various disciplines (including case studies from industry) with working groups dedicated to five special topics including, but not limited to the ones mentioned above: "multiple concepts of design", "evolution and management of requirements", "stakeholder issues and economics of requirements", "intertwining requirements and design", and "requirements, architecture, and complexity".

While the details of the results of these working groups are given in a more detailed document of its own, from an overall perspective the following conclusions can be drawn. We argue that current and future design requirements are shaped by the rapid change in implementation capabilities and platforms, new application demands, and rapidly evolving environments. Over its thirty-year history, the idea of design requirements has changed from single, static and fixed-point statements of desirable system properties into dynamic and evolving rationales that mediate change between the dynamic business environments and the design and implementation worlds. As Fred Brooks noted in the Dagstuhl workshop: "Design is not about solving fixed problems; it is constant framing of solution spaces". This evolution has now probably reached a new turning point characterized by unprecedented scale, complexity, and dynamism. This calls for new ways to think about requirements and their role in the design. Like earlier turning points, such as the software

crisis in the 1970s, it will demand a resolute and careful intellectual response. The four requirements principles have numerous implications for research questions of which only a minority has been addressed yet.

Overall, the good news is that the importance of RE continues to grow as the arguments for it are broadening. But the bad news is that besides the need for making a business case for a decent return on investment within a shorter time frame, in the future RE we need to consider additional arguments such as the need for the alignment with business process, understanding your own capabilities, systematizing the customer expectation managements, ensuring legal protection against IP loss or contract violation suits, creating user buy-in, and minimized training costs when justifying your next RE project. We need to therefore expand RE research into new fields—including complexity science, industrial design, organization design, and economics, among others. One challenge is the void of interdisciplinary intellectual exchange between diverse communities that have a stake at software requirements given the observed need for increased diversity in the design of software-intensive systems.

6.6 Emerging Uses and Paradigms for Dynamic Binary Translation

Seminar No. **08441** Date **26.10.–31.10.2008**
Organizers: Bruce R. Childers, Jack Davidson, Koen De Bosschere, Mary Lou Soffa

Software designers and developers face many problems in designing, building, deploying, and maintaining cutting-edge software applications–reliability, security, performance, power, legacy code, use of multi-core platforms, and maintenance are just a few of the issues that must be considered. Many of these issues are fundamental parts of the grand challenges in computer science such as reliability and security.

As an example, consider reliability. No serious software expert would claim that the current state-of-the-art in software engineering allows the construction of even moderately complex software that does not contain flaws or bugs. Perhaps a more realistic approach is to acknowledge that, no matter what process is used (i.e., rigorous design practices, thorough software testing, etc.), the delivered software will inevitably contain imperfections. If we accept this fact, it is then expedient to consider approaches that address these flaws after delivery. Certainly a natural approach would be to consider how to build software that adapts to flaws as it runs.

Similarly, consider the hardware evolution. As technology scaling continues, hardware devices exhibit more variability in their crucial performance metrics. There is both variability over multiple devices that leave the fab, and variability within single devices over their lifetime. Some components can behave erratically or completely fail. Software must be able to adapt to this changing computational substrate.

The underlying theme is that modern software must be able to adapt–whether to an existing bug, a security attack, changing power availability, or a failing core. No longer

do we build software with the notion of constancy, rather we fully embrace dynamism and develop the theories and capabilities that allow software to adapt dynamically. The resulting software is more robust, can adapt to a changing computational environment, and is potentially cheaper and faster to build.

Dynamic binary translation (DBT) has gained much attention as a powerful technique for the run-time adaptation of software. It offers unprecedented flexibility in the control and modification of a program during its execution. Some of the uses of DBT include emulating and simulating instruction sets, monitoring and optimizing performance at run-time, providing resource protection and management, virtualizing resources, and detecting malware. In many cases, virtualization looks like a very promising paradigm, and in such cases DBT is the primary means for achieving and implementing virtualization.

This Dagstuhl Seminar brought together experts from research and industry to discuss and identify the problems and promising directions for the new software requirements and hardware architectures described above. During our seminar, we identified significant challenges and opportunities in six areas of DBT research.

Chapter 7

Distributed Computation, Networks, VLSI, Architecture

7.1 Numerical Validation in Current Hardware Architectures

Seminar No. **08021** Date **06.01.–11.01.2008**
Organizers: Annie Cuyt, Walter Krämer, Wolfram Luther, Peter Markstein

The major emphasis of the seminar concentrated on numerical validation in current hardware architectures and software environments. The general idea was to bring together experts who are concerned with computer arithmetic in systems with actual processor architectures and scientists who develop, use and need techniques from verified computation in their applications. Topics of the seminar therefore included

- The ongoing revision of the IEEE 754/854 standard for floating-point arithmetic
- Feasible ways to implement multiple precision (multiword) arithmetic and to compute the actual precision at run-time according to the needs of input data
- The achievement of a similar behaviour of fixed-point, floating-point and interval arithmetic across language compliant implementations
- The design of robust and efficient numerical programs portable from diverse computers to those that adhere to the IEEE standard.
- The development and propagation of validated special purpose software in different application areas
- Error analysis in several contexts
- Certification of numerical programs, verification and validation assessment.

Computer arithmetic plays an important role at the hardware and software level, when microprocessors, embedded systems or grids are designed. The reliability of numerical software strongly depends on the compliance with the corresponding floating-point norms. Standard CISC processors follow the 1985 IEEE-754 Standard which is actually under revision, but the new highly performing CELL processor is not fully IEEE compliant. The draft standard IEEE 754r guarantees that "systems perform floating-point computation that yields results independent of whether the processing is done in hardware, software, or a combination of the two. For operations specified in this standard, numerical results and exceptions are uniquely determined by the values of the input data, sequence of operations, and destination formats, all under user control." There was a broad consensus that the standard should include interval arithmetic.

The discussion focused on decimal number formats, faithful rounding, longer mantissa lengths, higher precision standard functions and linking of numerical and symbolic algebraic computation. This work is accompanied by new vector, matrix and elementary or special function libraries (i.e. complex functions and continued fractions) with guaranteed precision within their domains. Functions should be correctly rounded and even last bit accuracy is available for standard functions. Important discussion points were additional features like fast interval arithmetic, staggered correction arithmetic or a fast and accurate multiply and accumulate instruction by pipelining. The latter is the key operation for a fast multiple precision arithmetic. An exact dot product (implemented in pipelined hardware) for floating point vectors would provide operations in vector spaces with accurate results without any time penalty. Correctly rounded results in these vector spaces go hand in hand with correctly rounded elementary functions. The new norm will be based on solid and interesting theoretical studies in integrated or reconfigurable circuit design, mathematics and computer science.

In parallel to the ongoing IEEE committee discussions, the seminar aimed at highlighting design decisions on floating-point computations at runtime over the whole execution process under the silent consensus that there are features defined by the hardware standard, language defined or deferred to the implementation for several reasons.

Hardware and software should support several (user defined) number types, i.e. fixed width, binary or decimal floating-point numbers or interval arithmetic. A serious effort is actually made to standardize the use of intervals especially in the programming language C. Developers are encouraged to write efficient numerical programs that are easily portable with small revisions to other platforms. C-XSC is a C++ Library for Extended Scientific Computing to develop numerical software with result verification. This library is permanently enlarged and enhanced as highlighted in several talks.

However, depending on the requirements on speed, input and output range or precision, special purpose-processors also use other number systems, i.e. fixed-point or logarithmic number systems and non-standard mantissa length. Interesting reported examples were a 16-bit interval arithmetic on the FPGA (Field Programmable Gate Array) based NIOS-II soft processor and an online-arithmetic with rational numbers. Therefore, research on reliable computing includes a wide range of current hardware and software platforms and compilers.

Standardization is also asked for inhomogeneous computer networks. The verification step

should validate the partial results coming from the owners of parcels before combining them to the final result.

Our insights and the implemented systems should be used by people with various numerical problems to solve these problems in a comprehensible and reliable way and by people incorporating validated software tools into other systems or providing interfaces to these tools.

So we want to create an increasing awareness of interval tools, the validated modeling and simulation systems and computer-based proofs in science and engineering.

There is a follow-up publication for this seminar:
Numerical Validation in Current Hardware Architectures: International Dagstuhl Seminar, revised papers,
Lecture Notes in Computer Science 5492, Springer 2009.

7.2 Perspectives Workshop: Telecommunication Economics

Seminar No. **08043** Date **23.01.–26.01.2008**
Organizers: Burkhard Stiller

This Dagstuhl Seminar on "Telecommunication Economics" was organized to discuss and develop partially a strategic research outline among key people in order to enhance the competence in the field of telecommunication economics and respective network management tasks for integrated Internet and telecommunication networks. The view on respective guidelines and recommendations to relevant players (end-users, enterprises, operators, regulators, policy makers, and content providers), focusing on the provision of new converged wireless services and content delivery networks to people and enterprises determined an aspect of relevance.

The main objective was to allow business partnering to drive networking services and their sustainable provisioning for consumers and enterprises alike. This includes in more specific detail the following four objectives:

- The support of engineering leadership gained in mobile, broadband, digital TV, and wire-line communications, and selected media fields, by new sustainable business models in a fully deregulated and diversified demand framework.
- The study and identification of business opportunities throughout the value chain, especially for enterprises, content, and specialized services.
- The contribution to a strategy relative to socio-economic needs by increasing the motivation for deployment of cost effective and flexible solutions using networks and content.

- The provisioning of guidelines and recommendations for utilizing different types of technologies and quantify necessary actions. These results will potentially supply regulators and standardization bodies with analysis and guidelines for creating conditions for fast growing competitive mobile, broadband, and content markets while speeding up business.

Thus, in a nutshell, the "Telecommunications Economics" Seminar shall help researchers to guide the development of (a) applicable and usable prototypes and (b) services from the socio-economic and business perspectives, and not put them in the position of first waiting for customers to adopt their technologies.

There is a follow-up publication for this seminar:
Burkhard Stiller: Telecommunication economics: overview of the field, recommendations, and perspectives.
Computer Science: Research and Development 23.2009,1, p. 35-43.

7.3 Transactional Memory: From Implementation to Application

Seminar No. **08241** Date **08.06.–13.06.2008**
Organizers: Christof Fetzer, Tim Harris, Maurice Herlihy, Nir Shavit

A goal of current multiprocessor software design is to introduce parallelism into software applications by allowing operations that do not conflict in accessing memory to proceed concurrently. The key tool in designing concurrent data structures has been the use of locks. Unfortunately, course grained locking is easy to program with, but provides very poor performance because of limited parallelism. Fine-grained lock-based concurrent data structures perform exceptionally well, but designing them has long been recognized as a difficult task better left to experts. If concurrent programming is to become ubiquitous, researchers agree that one must develop alternative approaches that simplify code design and verification.

The transactional memory programming paradigm is on its way to becoming the approach of choice for replacing locks in concurrent programming. Combining sequences of concurrent operations into atomic transactions seems to promise a great reduction in the complexity of both programming and verification, by making parts of the code appear to be sequential without the need to program fine-grained locks. Transactions will hopefully remove from the programmer the burden of figuring out the interaction among concurrent operations that happen to conflict when accessing the same locations in memory. There has been a flurry of work on transactional memory systems, hardware implementations (HTM), purely software based ones, i.e. software transactional memories (STM), and hybrid schemes (HyTM) that combine hardware and software.

The need for techniques such as transactional memory has become ever so urgent with the shift in our basic computing paradigm from single core chips to multi-core chips: we

are soon to see a multiprocessor on every desktop, and we need ways of programming them efficiently. Many of the big hardware and software vendors are scrambling to invest in transactional memory research and development, and academia is in a great need to coordinate the many world-wide research efforts on the topic.

This Dagstuhl seminar brought together leading researchers working on transactional memory (HTM, STM, and HyTM) from both industry and academia, in order for the dialog to help to create a concerted effort in pushing forward this important research agenda. We discussed several ongoing research directions in transactional memory:

- What are the fundamental approaches in development of STM platforms for the new class of multi-core machines?
- What support do transactions need in hardware, compilation, library support, and how would HTM and STM systems interoperate?
- How do we build TM-based applications and what should the STM/HTM implementations provide to simplify their design given application development experience?
- What programming language transactional constructs would best help in programming, and what are the issues with their implementation?
- Can we put together a complete transactional programming environment given current knowledge, that is, hardware, software, and programming language support?

In summary, our seminar addressed various issues of STM design from both the implementation and application perspectives, with a meaningful and intense discussion among the researchers that improved our understanding of both. Hopefully, this will eventually result in better transactional memory platforms that make for faster and simpler-to-program concurrent applications.

7.4 Perspectives Workshop: End-to-End Protocols for the Future Internet

Seminar No. **08242** Date **08.06.–11.06.2008**
Organizers: Jari Arkko, Bob Briscoe, Lars Eggert, Anja Feldmann, Mark Handley

A critical momentum is building to push the current architecture of the Internet forward. Several substantial "Future Internet" initiatives have started in Europe, the US and Asia, and vendor and network operator communities have started to actively discuss the topic. Several proposals ranging from incremental changes to the current Internet to completely new architectures are under consideration.

Within these communities, however, a disconnect exists between the people designing the internetworking layers and the people designing methods for supporting end-to-end

data transport. This situation is problematic, because in the end, the benefits visible to average users depend on the robust interoperation of the end-to-end transports used by applications in concert with the behavior of the internetwork layer. For instance, identifier-locator separation may cause locators to be changed in a manner that causes TCP to treat associated effects as congestion. It is hence critical to engage in a discussion on how to maximize the synergies between a future internetwork architecture and future end-to-end transport.

This perspectives workshop brings the community of researchers and engineers experienced in current and next-generation internetworking architectures together with the community developing the Internet's end-to-end transport protocols. The goal of the workshop is to begin a dialog between the communities that will allow a Future Internet to deliver real performance and service-quality benefits to its users and services.

Prospective workshop participants should be interested in:

- new communication paradigms and network architectures
- cross-disciplinary motivations for new architecture – social, commercial and economic
- applications enabled or improved through advanced networking and transport mechanisms
- performance characteristics of new communication protocols and network architectures
- availability and robustness aspects of new architectures
- handling of unwanted traffic
- transition and deployment aspects of new network architectures
- formal principles of architecture and transport

The focus of this perspectives workshop is on discussing these issues among the participants, rather than conference-style presentations. Among the desired results of this workshop are the identification of main architectural concepts and building blocks, the identification of the main remaining research issues, the identification of erroneous assumptions and misunderstandings between the involved research communities and an attempt at a synthesis of ideas into one thought framework.

There is a follow-up publication for this seminar by the organizers:
"Dagstuhl perspectives workshop on end-to-end protocols for the future internet",
Computer communication review: 39.2009,2: p. 42–47.

7.5 Fault-Tolerant Distributed Algorithms on VLSI Chips

Seminar No. **08371** Date **07.09.–10.09.2008**
Organizers: Bernadette Charron-Bost, Shlomi Dolev, Jo Ebergen, Ulrich Schmid

The Dagstuhl seminar 08371 on *Fault-Tolerant Distributed Algorithms on VLSI Chips* was devoted to exploring whether the wealth of existing fault-tolerant distributed algorithms research can be utilized for meeting the challenges of future-generation VLSI chips. Participants from both the distributed fault-tolerant algorithms community, interested in this emerging application domain, and from the VLSI systems-on-chip and digital design community, interested in well-founded system-level approaches to fault-tolerance, surveyed the current state-of-the-art and tried to identify possibilities to work together. The seminar clearly achieved its purpose: It became apparent that most existing research in Distributed Algorithms is too heavy-weight for being immediately applied in the "core" VLSI design context, where power, area etc. are scarce resources. At the same time, however, it was recognized that emerging trends like large multicore chips and increasingly critical applications create new and promising application domains for fault-tolerant distributed algorithms. We are convinced that the very fruitful cross-community interactions that took place during the Dagstuhl seminar will contribute to new research activities in those areas.

Description

Shrinking feature sizes and increasing clock speeds are the most visible signs of the tremendous advances in VLSI design, which will accommodate billions of transistors on a single chip in the near future. This comes, however, at the price of increased system-level complexity: In today's deep submicron technology with GHz clock speeds, wiring delays dominate transistor switching delays, and signals cannot traverse the whole die within single clock cycle any more. In fact, a modern VLSI chip can no longer be viewed as a monolithic block of synchronous hardware, where all state transitions occur simultaneously. Rather, VLSI chips are nowadays considered as systems of interacting subsystems — the advent of Systems-on-Chip (SoC)and Networks-on-Chip (NoC).

In addition, ever-increasing manufacturing variabilities increase the defect ratio, and the reduced voltage swing needed for high clock speeds and low power consumption also increases the adverse effects of -particle and neutron hits during operation, as well as crosstalk and ground-bouncing sensitivity. The resulting increase of the transient failure rate (soft-error rate), which was negligible in most former-generation chips, has hence raised general concerns about the dependability of future generation VLSI chips. Consequently, suitable fault-tolerance mechanisms with respect to timing errors or value errors are vital for such devices: Fine-grained fault-tolerance like radiation-hardening, fault masking at transistor or gate level, error-correcting codes or error detection and recovery are the primary methods of choice here.

Due to the above trends, however, modern VLSI chips have much in common with the loosely-coupled distributed systems that have been studied by the fault-tolerant dis-

tributed algorithms community for decades. System-level fault tolerance based on replication and distributed agreement is the dominant approach here, and a wealth of different computing and failure models, algorithms & protocols, and theoretical results regarding solvability of problems and achievable performance have been established in the past.

The purpose of our Dagstuhl seminar was to explore whether fault-tolerant distributed algorithms research can indeed be utilized for meeting the challenges of future-generation VLSI chips: Just as Temporal Logic, established in the distributed computing scope decades ago, found its way to the VLSI domain, other radically new solutions and methods may also find their way. And indeed, some recent research suggested a positive answer to this question: For example, demonstrated that distributed fault-tolerant clock generation algorithms can be adapted to the very special requirements of VLSI chips, and demonstrated that self-stabilization is a very promising approach for designing robust VLSI chips.

Fifteen participants from the distributed fault-tolerant algorithms community (and related fields, like verification), interested in the new application domain of VLSI chips, and twelve participants from the VLSI community, interested in system-level approaches to fault-tolerance, joined at Dagstuhl in order to survey the current state-of-the-art and identify possibilities to work together.

The presentations and the unique setting of Dagstuhl, with its relaxed and stimulating atmosphere, fully achieved their purpose: Long discussions during the official seminar, and many fruitful cross-community interactions during the free times were stimulated, which even exceeded the amount of available time.

Chapter 8

Modelling, Simulation, Scheduling

8.1 Scheduling

Seminar No. **08071** Date **10.02.–15.02.2008**
Organizers: Jane W. S. Liu, Rolf H. Möhring, Kirk Pruhs

Scheduling is a form of decision making that involves allocating scarce resources to achieve some objective. The study of scheduling dates back to at least the 1950's when operations research researchers studied problems of managing activities in a workshop. Computer systems researchers started studying scheduling in the 1960's in the development of operating systems and time-critical applications.

Several of these scheduling problems proved to be closed linked to fundamental concepts in theoretical computer science (such as approximability, intractability, and NP-completeness), thereby attracting the attention of theoretical computer scientists as well. Today, scheduling is widely studied as parts of the disciplines of mathematical programming and operations research; algorithmics and theoretical computer science; and computer systems, particularly real-time and embedded systems. The specific scarce resources, as well as the objectives to be optimized, differ in the different disciplines; nevertheless, there are remarkable similarities (as well as significant differences) in the general framework adopted by researchers in scheduling theory in these disparate disciplines.

The primary objectives of the seminar were to bring together leading and promising young researchers in the different communities to discuss scheduling problems that arise in current and future technology; to expose each community to the important problems addressed by the other communities; and to facilitate a transfer of solution techniques from each community to the others.

There were approximately fifty participants at the seminar, approximately evenly split between the three areas. The first morning consisted of three introductory talks by Ted Baker on real-time scheduling, Andreas Schulz on mathematical programming techniques and by Nikhil Bansal on online scheduling. During the the first afternoon each participant gave a short talk about their research interests. The participants were then broken up into about ten clusters, of six people each, by the organizers based on research interests. Some of the clusters were:

- Game Theoretic Scheduling: This research line considers situations each user is motivated to try to optimize the performance for this user's task. So for example, the user might not truthfully state the resources required by his/her task. The general goal is to study the effect of such selfish behavior and to design mechanisms to align players incentives with the global good.

- Stochastic Scheduling: This research line assumes that only probabilistic information is known about the resource requirements for a job. The goal is to try to find schedules that are good in expectation.

- Power Aware Scheduling: Modern processors can change their speed as a way of managing power. This research line focuses on designing speed changing policies that will be the most effective.

- Periodic Scheduling: This line of research is distinguished by the fact that jobs must be executed periodically over a long period of time. Periodic scheduling occurs frequently in real-time applications. The resulting scheduling problems are notoriously difficult.

These clusters met periodically through the week to discuss research problems, and to organize talks from within the cluster. The clusters organized 32 medium length talks on various recent research results. There was a plenary session on Thursday afternoon to discuss interesting directions for future research and future collaborations.

This seminar was essentially a first meeting of the real-time community with the operations research and theoretical computer science community. The general consensus was that both communities learned a lot about the other communities. And at least some collaborations between the communities were born from this seminar.

8.2 The Evolution of Conceptual Modeling

Seminar No. **08181** Date **27.04.–30.04.2008**
Organizers: Lois Delcambre, Roland H. Kaschek, Heinrich C. Mayr

The seminar took place at Dagstuhl from 27 – 30 April 2008. It was organized by Roland Kaschek, Lois Delcambre and Heinrich C. Mayr. The seminar's purpose was looking into conceptual modeling from different perspectives, and along different dimensions: we wanted to achieve a better understanding of conceptual modeling issues in various domains of discourse, from a historical perspective and from a view beyond individual (modeling) projects. Consequently we did not focus on a particular application area or development project.

In total 33 colleagues attended the seminar and 26 presentations were given. Many attendees expressed their satisfaction with the superior working and meeting conditions at Dagstuhl, as well as with its beautiful environment. It was understood that the attendees

were interested in documenting in a Springer LNCS volume the common effort for better understanding conceptual modeling. Related editorial work as well as preparations is ongoing. Springer has in the meantime committed to making this book.

Many attendees explicitly mentioned to us that they highly valuated the breadth of subjects discussed in the seminar. It would, however, not be a true description of the seminar to keep quiet about the critique of some of the attendees of exactly that breadth of the discussion. Certainly they have a valid point here: to some extent, of course, the breadth goes at expense of the depth. On the other hand, considering smaller and smaller areas of knowledge for being capable of going into more and more depth of these small areas also has its problems. Overall the attendees evaluated the seminar positively. The project was launched to organize a continuing seminar at Dagstuhl in April 2013. Maybe a repeated seminar will be more successful at discussing the evolution of conceptual modeling. Maybe more explicitly acknowledging that modern computing already has a history will help to focus on the succession of ways to do conceptual modeling, the concepts and notations used throughout, as well as the problems to be solved and the degree to which one actually can do so.

The discussion was colored by contributions of a number of colleagues from smaller companies who attended the seminar; unfortunately they did not give presentations.

About 25 abstracts have been provided by seminar attendees for the above mentioned LNCS volume. We are currently confident to publish that volume in late 2009.

Chapter 9

Cryptography, Security

9.1 Perspectives Workshop: Network Attack Detection and Defense

Seminar No. **08102** Date **02.03.–06.03.2008**
Organizers: Georg Carle, Falko Dressler, Richard Kemmerer, Hartmut König, Christopher Kruegel

The increasing dependence of human society on information technology (IT) systems requires appropriate measures to cope with their misuse. The growing potential of threats, which make these systems more and more vulnerable, is caused by the complexity of the technologies themselves and by the growing number of individuals which are able to abuse the systems. Subversive insiders, hackers, and terrorists get better and better opportunities for attacks. In industrial countries this concerns both numerous companies and the critical infrastructures, e.g. the health care system, the traffic system, power supply, trade (in particular e-commerce), or the military protection.

In today's Internet there is a ubiquitous threat of attacks for each user. Most well-known examples are denial of service attacks and spam mails. However, the range of threats to the Internet and its users has become meanwhile much broader. It ranges from worm attacks via the infiltration of malware till sophisticated intrusions into dedicated computer systems. The Internet itself provides the means to automate attack execution and to make them more and more sophisticated. The protection against these threats and the mitigation of their effects has become a crucial issue for the use of the Internet. Complementary to preventive security measures, reactive approaches are increasingly applied to counter these threats. Reactive approaches allow detecting ongoing attacks and to trigger responses and counter measures to prevent further damage.

Network monitoring and flow analysis has been developed as complementary approach for the detection of network attacks. They aim at the detection of network anomalies based on traffic measurements. Their importance arose with the increasing appearance of denial of service attacks and worm evasions, which are less efficient to detect with intrusion detection systems.

Reactive measures comprise beside the classical virus scanner intrusion detection and flow analysis. The development of intrusion detection systems began already in the eighties. Intrusion detection systems possess a prime importance as reactive measures. They pursue two complementary approaches: anomaly detection, which aims at the exposure of abnormal user behavior, and misuse detection, which focuses on the detection of attacks in audit trails described by patterns of known security violations. A wide range of commercial intrusion detection products has been offered meanwhile; especially for misuse detection. The deployment of the intrusion detection technology still evokes a lot of unsolved problems. These concern among others the still high false positive rate in practical use, the scalability of the supervised domains, and explanatory power of anomaly-based intrusion indications. In recent years intrusion detection has received a wider research interest which increased the efficiency of the technology, in particular in connection with other approaches e.g. firewalling, honeypots, intrusion prevention.

In recent years network monitoring and flow analysis has been developed as complementary approach for the detection of network attacks. Flow analysis aims at the detection of network anomalies based on traffic measurements. Their importance arose with the increasing appearance of denial of service attacks and worm evasions which are less efficient to detect with intrusion detection systems. The flow analysis community developed two approaches for high speed data collection: flow monitoring and packet sampling. Flow monitoring aims to collect statistical information about specific portions of the overall network traffic, e.g. information about end-to-end transport layer connections. On the other hand, packet sampling reduces the traffic using explicit filters or statistical sampling algorithms.

There is an urgent need to coordinate the research activities in intrusion detection and network monitoring. For example, sampling and flow monitoring have been developed as important methods in the network monitoring field (for accounting, charging, and security). They are more and more applied for attack detection (anomaly detection, flow based signatures). This, however, requires a close cooperation of the two communities. The same applies to the intrusion detection community for the detection of worm epidemics and denial of service attacks. Here traffic analysis can help to make the detection procedure more effective. This objective makes the subject of the seminar to be rather cross-disciplinary.

There are two follow-up publications for this seminar:
Hartmut König: Dagstuhl Manifesto: Perspectives Workshop Network Attack Detection and Defense 08102.
Informatik Spektrum 32.2009,1, p. 70-72.
Georg Carle, Falko Dressler, Richard Kemmerer, Hartmut König, Christopher Kruegel: Network Attack Detection and Defense: Manifesto of the Dagstuhl Perspective Workshop.
Computer Science: Research and Development 23.2009,1, p. 15-25.

9.2 Countering Insider Threats

Seminar No. **08302** Date **20.07.–25.07.2008**
Organizers: Matt Bishop, Dieter Gollmann, Jeffrey Hunker, Christian W. Probst

Introduction

The "insider threat" or "insider problem" has received considerable attention, and is cited as the most serious security problem in many studies. It is also considered the most difficult problem to deal with, because an "insider" has information and capabilities not known to other, external attackers. However, the term "insider threat" is usually either not defined at all, or defined nebulously.

The difficulty in handling the insider threat is reasonable under those circumstances; if one cannot define a problem precisely, how can one approach a solution, let alone know when the problem is solved? It is noteworthy that, despite this imponderability, definitions of the insider threat still have some common elements. For example, a workshop report defined the problem as malevolent (or possibly inadvertent) actions by an already trusted person with access to sensitive information and information systems. Elsewhere, that same report defined an insider as someone with access, privilege, or knowledge of information systems and services. Another report implicitly defined an insider as anyone operating inside the security perimeter—while already the assumption of only having a single security perimeter may be optimistic.

The goal of this Dagstuhl seminar was to bring together researchers and practitioners from different communities to discuss in a multi-national setting what the problems are we care about, what our response is, which factors influence the cost of dealing with insider threats and attacks, and so on. In a time where we barely understand which factors cause insider threats, and our solutions are scattered all over communities, areas, and instruments, this coordinated action between the involved communities seems to be needed more than ever.

This Dagstuhl seminar was, to our knowledge, the first European seminar focusing on insider threats bringing together US and European researchers and practitioners. The five days of the seminar allowed not only for a rich assortment of presentations, but even more importantly for extended discussions, both formal and informal, among the participants. We even had the opportunity for a structured exercise that challenged participants to define specific insider threats, develop the appropriate responses, and critique each other's problem-solution formulation.

Who is an Insider?

One of the most urgent quests of the seminar was to try to identify the characteristic features of an insider. To this end two research groups presented their taxonomies for identifying insider threat, or insiders per se, and several recent cases of insider actions were discussed—Binney vs. Banner, a message flood created as consequence of a security bulletin, spies that stole secrets for the Chinese Army, and a tax authority employee who used her influence to embed backdoors into taxation software. While these cases could not differ more, they served the purpose of illustrating the widely differing characteristics of insider threats, and initiated an intense discussion of how one possibly could aim at detecting and inhibiting them.

[This section is shortened by the editor of Dagstuhl News. The following sections of the report are left out: Taxonomies, Issues in Detection/Forensics/Mitigation, Policies, Human Factors and Compliance, Surveys and the Real World.]

Conclusion

The Dagstuhl seminar on Countering Insider Threats provided a week-long base for discussions of what the relevant aspects of the problem are for different communities. In motivating the seminar, we asked the seminar participants to jointly emerge with a definition of what or who an insider is, and how to deal with the threat posed by it. It is our believe that we succeeded in getting a better understanding of what these different communities mean by "insider". As stated above, this knowledge has already during the seminar been used to develop integrated approaches towards qualitative reasoning about the threat and possible attacks. Beyond this shared definition of what constitutes an insider, the most prominent outcome of the seminar is the beginning of a taxonomy or framework for categorising different threats. While this fell short of the ambitious goals we originally had formulated [webpage], the process of reaching this definition was highly enlightening and is documented in this article. In the process, the seminar identified the need for:

- A framework or taxonomy for distinguishing among different types of insider threats;
- Methodologies for assessing the risk and impact of insider threat incidents;
- Incorporating human factors into the development of solutions;
- Better formulations for specifying useful policy at both systems and organizational levels—policy that would be meaningful and applicable to the insider threats deemed most important.

There were some cross-cutting conclusions that emerged from the seminar. The role of trust was discussed in a number of different contexts. In one sense, the ideal security framework for addressing insider threats would eliminate the need for trust — all behaviours would either be defined permissible, or else made impossible to execute. But this model ignores two realities. In any but the simplest settings, context of actions is highly determinative in shaping what is appropriate or needed behaviour. Further, many (most?) organizations would not accept a working environment so rigidly defined as to eliminate the need for trust. Hence, we emerge with the conclusion that trust relationships will be present in most organizations; how to best factor trust into security policies and frameworks remains, however, unclear.

Security, moreover, is context dependent. Security is not achieved by deploying generic (context free) controls. However, the importance of context in addressing insider threats poses a number of challenges. Capturing qualitatively the various situations that might arise in an organization is itself probably impossible, though effective dialogue between those defining security controls and those working as insiders in the organization will

certainly help. Hence, insider threat prevention and response has to deal with the reality that controls will not adequately capture all of the behaviours that might be appropriate in a given context. Even if all contexts could be qualitatively described, policy languages and controls are inadequate at the current time to fully capture the range of contexts identified.

Motivation and intent clearly are important in defining insider threats and defining appropriate detection/forensics/mitigation strategies. While intent (the purpose of actions) is at least partially observable, motivation (the incitement to action) is not. The intent to, for instance, obtain certain data may reflect malicious motives, or may reflect positive motives (as in a hospital emergency where certain information is desperately needed regardless of legitimate access). Devining motivation highlights the need for context aware policies, but even with context motivations may be difficult to determine. We conclude that approaches for understanding motivation a priori are still highly immature.

Each of these observations emphasize the conclusion that security will not be achieved solely by deploying security technology. Most people are not entirely logical or consistent in their behaviour, and this confounds our ability to formulate measures to reliably prevent or detect malicious insider behaviour.

As we write this report we are in the middle of preparing an application for a follow-up seminar. We would like to thank all participants of the seminar for making it a fruitful and inspiring event—and especially Dagstuhl's wonderful staff, for their endless efforts, both before and during the seminar, to make the stay in Dagstuhl as successful as possible.

As stated above we believe that the week in Dagstuhl has been influential in heightening awareness among communities for activities and developments. During the seminar many participants expressed the wish for a community website to establish a central focal point, both for communication between communities, but also to the outside, governmental agencies, and companies. This web portal is currently under construction: `http://www.insiderthreat.org`

9.3 Geographic Privacy-Aware Knowledge Discovery and Delivery

Seminar No. **08471** Date **16.11.–21.11.2008**
Organizers: Bart Kuijpers, Dino Pedreschi, Yucel Saygin, Stefano Spaccapietra

The Dagstuhl-Seminar on Geographic Privacy-Aware Knowledge Discovery and Delivery was held during 16 - 21 November, 2008, with 37 participants registered from various countries from Europe, as well as other parts of the world such as United States, Canada, Argentina, and Brazil. Issues in the newly emerging area of geographic knowledge discovery with a privacy perspective were discussed in a week to consolidate some of the research questions. The Dagstuhl program included plenary sessions and special interest group meetings which continued even late in the evening with heated discussions. The plenary sessions were dedicated for the talks of some of the participants covering a variety

of issues in geographic knowledge discovery and delivery. The reports on special interest group meetings (SIG) were also presented and discussed during the plenary sessions.

The topics of the talks presented during the plenary sessions could be summarized as:

- Consolidation of the notion of trajectories
- Semantic aspects of trajectories
- Warehousing of trajectories
- Mining of trajectories
- Applications
- Privacy and legal aspects

The special interest groups were formed to discuss in detail the (1) Definition and Semantics of Trajectories, (2) Applications of Anonymized Geographic Data. After each SIG meeting, a plenary meeting was organized to harmonize the results of the discussions and to share them with other interested researchers. A summary session was held for the SIG meetings to discuss open questions even basic ones regarding what a trajectory is and how do we define the specific area of knowledge discovery on geographic data, and application areas of privacy in trajectory data publication and analysis. The issues discussed during the SIG meetings are summarized in the following subsections.

Privacy SIG: Applications of Anonymized Data

In this working group privacy issues in trajectories were discussed around the popular topic of anonymity. The aim was to find some killer applications for the anonymized data so that anonymization techniques could be adopted and widely used. The questions thrown into the discussion arena were the following:

- What are the Killer Applications?
- What information is needed to support these applications?
- What structure of data best supports these applications?

The psychological as well as the economical aspects of privacy were discussed during the meetings. The relationship between the psychological vs quantifiable privacy risks were questioned in the context of geo-referenced data. The rich background information and other sources which could be linked to the location data were pointed out. The need to move from the needed information content to the anonymization techniques to support it was proposed as the main research direction of anonymization. The Economic aspects of privacy questioning wether privacy could be treated as a "good" and if we can sell or buy privacy was discussed.

The applications of geo-referenced data were categorized into groups:

- health
- traffic and transportation
- edutainment
- public safety emergency response
- recommender systems

The types of geo-spatial self-published data in the context of the above applications were identified as:

- Geo-referenced media such as photos in Flickr and similar sites which are tagged with spatial information.
- Blogs - including textual location Calendar - event location
- Personal GPS tracks obtained during activities like running, walking, or hiking which could be combined with health data
- SMS-based microblogging like in VANET

The typical uses of the geo-referenced data were listed as:

- Direct geo-marketing
- Social science
- Investigation, public safety, and law enforcement
- Event detection
- Dating
- Recommender systems
- Information dissemination - geocast (instead of broadcast)
- Targeted health information

The possible risks or misuses of the collected georeferenced data were identified as:

- Geospam
- Harassement or stalking
- Profiling: Insurance risks
- Suppressing political dissent

- Archived data - changing standards
- Travel restrictions
- Dissemination of misinformation

The possible precautions against the misuse of the above data were identified as:

- Educating the society about the risks, and demonstrate the inference capabilities.
- Regulations
- Technology for scrubbing,generalization, cloaking for anonymity. Enabling multiple virtual identities and authentication through virtual ID.
- Continuous and dynamic Risk Assessment adaptive to change with increasing information

SIG for Trajectories: Definition and Semantics

The first meeting of this SIG was on the basics of trajectories. Two views of movement which are traffic or trajectory oriented are discussed. An important communication or terminology problem appeared during this first SIG meeting to identify the transition from the data to patterns. Data Mining Query Language (DMQL) was considered to be a transition from the data world to the pattern world. It was stated that trajectory mining needs a language with the capability to express spatio-temporal constraints over sequences and sets of objects.

The second meeting of this SIG was on the semantic aspects. The main question was weather semantics of movement is anything that goes beyond raw spatio-temporal positioning. Multiple Semantic Layers were identified which are

- Movement only, no semantics: sequence of ST points
- Trajectories as meaningful movement segments (meaningful: from the application viewpoint)
- Trajectories as sequences of moves between "places" (geo-places or anything else that changes over time/space)
- Trajectories attached to objects (persons, cars, birds, etc.)

The third SIG meeting was on aggregation and warehousing issues of trajectories. Conceptual models(EER) DW model, spatial hierarchies, spatial OLAP, implementation, temporal issues were discussed.

9.4 Theoretical Foundations of Practical Information Security

Seminar No. **08491** Date **30.11.–05.12.2008**
Organizers: Ran Canetti, Shafi Goldwasser, Günter Müller, Rainer Steinwandt

Introduction and Motivation

Designing, building, and operating secure information processing systems is a complex task, and the only scientific way to address the diverse challenges arising throughout the life-cycle of security criticial systems is to consolidate and increase the knowledge of the theoretical foundations of practical security problems. To this aim, the mutual exchange of ideas across individual security research communities can be extraordinary beneficial. Accordingly, the motivation of this Dagstuhl seminar was the integration of different research areas with the common goal of providing an integral theoretical basis that is needed for the design of secure information processing systems.

Coping with the full spectrum of challenges in information security is far beyond the scope of a single seminar, and thus participants were selected from a number of different, but still related, fields, so that a common scientific language or similarity of theoretical tools can facilitate an efficient exchange of ideas. Ideally, the seminar would help in identifying possibilities of cross-fertilization among seemingly different research directions within information security.

In addition to senior experts from academics, an effort was made to include participants with experience in industry, and also to include young researchers in the field who had already demonstrated a strong research potential.

Atmosphere, Organization, and Participation

It is fair to say that the seminar brought together some of the world experts on theoretical foundations of information security. The organizers are indebted for excellent presentations that were delivered by participants at this Dagstuhl seminar. More than 30 participants from several countries came together, and the additional flexibility offered by a Dagstuhl seminar in comparison to traditional conferences turned out to be of invaluable help:

Talks of different lengths were scheduled, and lively technical discussions during talks were the norm. This resulted in various program changes and the scheduling of additional talks. The possibility to discuss results in technical depth when needed was of great benefit and together with the infrastructure offered by Schloss Dagstuhl resulted in an extremely fruitful reasearch atmosphere with rather long working hours. Feedback of seminar participants to the organizers was extremely positive, and it is no exaggeration to consider this Dagstuhl seminar as a world class research meeting.

Summary of Topics

Owing to the nature of the workshop, the topics presented covered quite different subjects of information security. Because of the significance of cryptographic techniques, it is not surprising that many talks made use the the technical machinery offered by research on theoretical foundations of cryptography. These talks were supplemented by presentations on several other aspects of information security and privacy.

Given the topic of the workshop, it is not surprising that presentations often focused on theoretical models and provable constructions. However, the techniques used in different presentations varied greatly, therewith giving seminar participants the possibility of experiencing techniques not typically encountered in their own line of research. This type of crossing the boundaries of individual subareas of research on the theoretical foundations of information security was hoped for in the organization of this seminar and greatly added to the diversity of the discussions.

Chapter 10

Data Bases, Information Retrieval

10.1 Ranked XML Querying

Seminar No. **08111** Date **09.03.–14.03.2008**
Organizers: Sihem Amer-Yahia, Divesh Srivastava, Gerhard Weikum

This paper is based on a five-day workshop on "Ranked XML Querying" that took place in Schloss Dagstuhl in Germany in March 2008 and was attended by 27 people from three different research communities: database systems (DB), information retrieval (IR), and Web. The seminar title was interpreted in an IR-style "andish" sense (it covered also subsets of Ranking, XML, Querying, with larger sets being favored) rather than the DB-style strictly conjunctive manner. So in essence, the seminar really addressed the integration of DB and IR technologies with Web 2.0 being an important target area.

DB and IR have evolved as separate communities for historical reasons. They were spawned in the sixties with focus on very different application areas: accounting and reservation systems on the DB side, and library and patent information on the IR side. Consequently, they have emphasized different methodological paradigms: precise querying over schematized data, based on logic and algebra (DB), vs. keyword search and ranking over text and uncertain data, based on statistics and probability theory (IR). However, there are now many applications that require managing both structured and unstructured data and thus mandate serious consideration on how to integrate the DB and IR worlds at both foundational and software-system levels. These applications include Web and Web 2.0 use cases as well as more corporate-oriented scenarios such as customer support and health care. All three communities that participated in the seminar (DB, IR, Web) agreed on the importance of the general direction and came up with ten tenets, from different viewpoints, on why DB&IR integration is desirable.

All three of the participating communities – DB, IR, and Web – felt that looking across the fence paid off very well, and that the communities should continue learning from each other. Challenges are ahead in areas like Web 2.0, personal information management, and entity-relationship search; these will remain difficult and rewarding areas for a while. Combining the different and quite complementary expertises from DB and IR would be vital towards well-founded and practically viable solutions.

10.2 Software Engineering for Tailor-made Data Management

Seminar No. **08281** Date **06.07.–11.07.2008**
Organizers: Sven Apel, Don Batory, Goetz Graefe, Gunter Sake, Olaf Spinczyk

Tailoring data management components is not only important for the domain of embedded systems. A database management system (DBMS) that provides exactly the functionality that is needed by the platform, the application scenario, and the stakeholder's requirements, yields several benefits, e.g., a lean and well-structured code base, robustness, and maintainability.

The desire for tailor-made data management solutions is not new: concepts like kernel-systems or component toolkits have been proposed 20 years ago. However, a view on the current practice reveals that nowadays data management solutions are either monolithic DBMS such as ORACLE and DB2 or special-purpose systems developed from scratch for specific platforms and scenarios, e.g., for embedded systems and sensor networks.

While monolithic DBMS architectures hinder a reasonable and effective long-term evolution and maintenance, special-purpose solutions suffer from the conceptual problem to reinvent the wheel for every platform and scenario or to be too general to be efficient. A mere adaptation of present solutions is impossible from the practical point of view, e.g., it becomes too expensive or simply impractical, which is confirmed by the current practice. Especially in the domain of embedded and realtime systems there are extra requirements on resource consumption, footprint, and execution time. That is, contemporary data management solutions have to be tailorable to the specific properties of the target platform and the requirements and demands made by the stakeholders and the application scenario.

The question that arises is what is different now from the attempts made some years ago that makes us believe that something can change in future. This question and possible answers from the field of modern software engineering have been discussed in the Dagstuhl Seminar "Software Engineering for Tailor-made Data Management", July 6h to July 11th, 2008.

In 12 sessions, 21 talks have presented recent work and new ideas regarding trends in software engineering and their application to the development of tailor-made data management solutions. On the one hand, there were researchers from the database community who stressed the necessity of tailoring data management solutions, e.g., in order to be able to adapt them at runtime in a service-oriented world or to provide different variants for query processing and optimization. On the other hand, several researchers from the software engineering and programming languages communities presented new ideas to build, manage, and evolve complex systems such as database systems. Continuous discussions have connected the different views of both communities. A recurring idea was to think in terms of families of DMS instead of individual systems. Members of a DMS family share common "features" with the goal of reuse. Several researchers presented mechanisms for identifying, implementing, and managing features or reported from success stories in the field of DMS or related domain.

In two special sessions the participants have been split into three discussion groups addressing the specific topics of tailor-made data management solutions:

- "Software Engineering or C for Embedded Systems?"
- "Dynamic Adaptation in DBMS - Is There a Use Case?"
- "DBMS in 10 Years - How Can Software Engineering Help?"

In a wrap-up session, the participants have reflected the program and discussions of the seminar. The general opinion was that both communities have learned much from each other. Individual joint projects have been planned, a mailing list has been installed, and a continuation of a related workshop series has been encouraged.

10.3 Uncertainty Management in Information Systems

Seminar No. **08421** Date **12.10.–17.10.2008**
Organizers: Christoph Koch, Birgitta König-Ries, Volker Markl, Maurice van Keulen

Topic and Aims of the Seminar

Computer science has long pretended that information systems are perfect mirror images of a perfect world. Database management systems, e.g., work under the assumption that the data stored represent a correct subset of the real world. Of course, this idealized assumption is rarely true. Information systems contain

- wrong information caused, e.g., by data entry errors: This is a common problem for instance in genomic databases
- imprecise or falsely precise information, e.g., a measuring device will provide information with a certain precision only. Typically, information systems store the measured date, but do not store information about the conditions under which this data is true and the precision achieved.
- incomplete information. A certain piece of information may not be available to the information system.
- inconsistent information. Different information systems may contain contradictory information.

In the past information systems have worked around these flaws by extensive consistency checking, plausibility checks, or human discovery and correction. These solutions are bound to fail as systems become ever more distributed, the information more globalized, and the individual systems more autonomous. Hence, we need to find ways for our information systems to directly deal with the uncertainty induced by them.

Nor is imperfection necessarily a bad thing. Inconsistent or unknown information has been addressed by more or less ad-hoc concepts like "NULL" values in SQL. Take inconsistent information. It may reflect information collected under different circumstances or in different contexts, i.e., it may represent different views on the same phenomenon, and the sum total may very well carry more information than any single one.

The challenge, then, is to make system operation resilient to imperfect data. Resilience is not simply a matter of correction but more so of reconciling what appears contradictory information.

Meeting the challenge becomes particularly pressing when we consider the modern development of the computing environment into large-scale, open, mobile, extremely widely distributed systems. Even if everything works correctly, it will no longer be possible to guarantee consistency across such systems. Consider as an example a large-scale peer-to-peer organization. Each peer observes part of the environment only. Even collectively the peers will never observe the environment at the same instance in time. Hence, there is neither individually nor collectively a consistent image of the environment. This reminds one of the uncertainty principle in physics which states that predictions can be made only within certain probabilities. Consequently, what such systems need to incorporate is what is referred to as uncertainty management. We will need mechanisms that allow the individual components to function despite the fact that they have incomplete and maybe incorrect knowledge, and that the system as a whole reaches its goals by limiting the collective uncertainty of collaborating subsystems to acceptable levels of uncertainty.

Uncertainty is an issue that appears in many disciplines. The aim of the seminar was to bring together researchers from all these communities. Some of these communities have a long history of dealing with this problem, for others, it is a new challenge. These are in particular:

- Database Management Systems
- Multi-Agent Systems
- Peer to Peer Systems
- Sensor Networks
- Data Stream Management Systems
- Reputation Systems
- Context-Aware Systems
- Artificial Intelligence
- Information Retrieval
- Self-organizing Systems
- Semantic Web

- Market Economics, Decision Science
- Fuzzy Systems

Each of these communities needs to deal with the issues described above, and many of them do in their own ways. Unfortunately, up to now, there has been little exchange between the communities on their approaches. The outcome of this seminar was a classification of the different types of uncertainty, an overview of the state of the art on dealing with them in the different communities, the applicability of these solutions to other types of systems, and an identification of promising avenues of research.

The seminar was roughly structured along the following three areas:

- Fundamentals, e.g., models for representing uncertain data, impact of uncertainty on (database) operations, consistency models, error correction
- Methods for uncertainty reduction and inconsistency tolerance
- Applications, e.g., obtaining information from sensor networks or stream management systems, structuring unstructured data, object localization, belief revision in agent and reputation systems, personal information management, data integration.

[The following section "Organization of the Seminar" was deleted by the editor of Dagstuhl News.]

Talks

We had scheduled beforehand a number of plenary talks with the aim to provide participants with a wide overview of topics in uncertainty management. In addition, participants had the opportunity to give shorter talks on their research topics in parallel sessions.

During the seminar, the following plenary talks were given:

- Anish Das Sarma, Stanford University: Trio: A System for Data, Uncertainty, and Lineage
- Amol Deshpande, University of Maryland - College Park: Uncertain Data Management for Sensor Networks
- Norbert Fuhr, Universität Duisburg-Essen: Vague Predicates, Probabilistic Rules and 4-Valued Logic for Probabilistic Databases
- Peter Haas , IBM Almaden Center - San José: A Monte Carlo Approach to Managing Uncertain Data
- Ihab Ilyas , University of Waterloo: URank: Ranking Uncertain Data
- Christoph Koch, Cornell University: MayBMS: A System for Managing Large Uncertain and Probabilistic Databases

- Christopher Re , University of Washington: An Overview of the Mystiq Probabilistic Database
- Maurice van Keulen , University of Twente: Probabilistic Data Integration

[The following sections "What is special about uncertainty management in information systems?" and "Demos" were deleted by the editor of Dagstuhl News.]

Workgroups

Based on the selection process described above, workgroups were formed. The first set of groups met on Tuesday and Wednesday and then reported back to the plenary, the second set on Thursday. Details on the workgroup results can be found in the individual reports by the workgroups. We include here just a very brief description based on these reports on what the workgroups were about, hoping to encourage the reader to take a closer look at these reports.

Uncertainty and Trust. The aim of the working group was to analyze the relationship between trust and uncertainty in distributed reputation systems. Starting from the identification of sources and types of uncertainty in such systems, the group discussed their relationship to trust. Afterwards, a list of desirable properties of trust representations was compiled and finally open research challenges in the area were identified by the participants.

Lineage/Provenance in Probabilistic Databases. The group discussed the different usages of lineage information in probabilistic databases, different ways to represent lineage depending on the use case, as well as a number of open issues including approximation of lineage, uncertainty in lineage information, the relationship between lineage and privacy, and the implications non-independent input data have on lineage.

Explanation. This group's discussion was structured along three main questions: Why is explanation of results needed in (uncertain) information systems? What should such an explanation contain? How can it be provided, more precisely how should uncertainty be represented?

Probabilistic Databases Benchmark. This group was attended by representatives of groups working on probabilistic databases and discussed first steps towards a common benchmark that shall allow to compare different approaches. A number of concrete steps that will be taken in the near future were agreed upon.

Imprecision, Diversity, and Uncertainty. The main goal of this workgroup was to discuss how to measure uncertainty in the data and the model and how to determine the quality of uncertain data.

Classification, Representation, and Modeling. The discussion in this group was split up into two subgroups: the first subgroup studied how different representation and modeling alternatives currently proposed can fit in a bigger picture of theory and technology interaction, while the second subgroup focused on contrasting current system implementations and the reasons behind such diverse class of available prototypes.

Conclusion

The seminar confirmed our belief that uncertainty management is an extremely important area of computer science research that will need contributions from a number of disciplines to be successfully tackled. The seminar identified a number of potential killer applications and many advantages of incorporating uncertainty management in information systems. The seminar provided an excellent basis for the initiation of such interdisciplinary work, but also for the exchange of ideas and the organisation of future collaboration among groups working in the same area, as is evidenced for instance by the probabilistic databases benchmarking initiative. Here, the "database-heavy" nature of the group turned out to be very beneficial to achieving a concrete outcome of the seminar. For a more detailed description of the results, please refer to the workgroup reports included in the seminar proceedings.

Chapter 11

Bioinformatics

11.1 Computational Proteomics

Seminar No. **08101** Date **02.03.–07.03.2008**
Organizers: Christian Huber, Oliver Kohlbacher, Michal Linial, Katrin Marcus, Knut Reinert

The field of Computational Proteomics has grown rapidly and gained a lot of momentum over the last years. Computational Proteomics was previously a field with a small and specialized community. Over the last few years, however, it has been recognized by experimental groups, that the analysis of the increasingly complex proteomics studies has become intractable without efficient algorithms implemented in easy-to-use tools. Conversely, the computer science and bioinformatics communities started to realize the wealth of interesting problems in this area. Both sides are thus eager to come together and work on these problems. To initiate this close collaboration of researchers from different fields (biology, medical sciences, biochemistry, analytical chemistry, bioinformatics, and computer science), we need opportunities to bring both communities together in an inspiring and relaxed atmosphere. Past experience from our 2005 seminar has shown that Dagstuhl is an ideal place for this.

Currently, computational proteomics faces a number of challenges. The increasing speed and accuracy of the new instrument generation yields datasets that are up to an order of magnitude larger than datasets seen a few years ago. This implies new algorithmic techniques and data analysis capabilities. The computer science problems to be tackled range from data management, over optimization problems, to machine learning. On the experimental side, the development of high mass accuracy and better separation techniques pose new challenges.

The seminar brought together a mixed audience from proteomics and bioinformatics: At the beginning we took a pool showing that there were 10 people who declared themselves as “wet lab” and 26 as “computer science”. Moreover, there were 21 people “holding a PhD degree” and 15 “working on it”. We are happy to see that the communities really grow together.

The talks and discussions were arranged on a daily basis featuring related topics:

- Systems Biology and novel experimental techniques.
- Quantitative analysis.
- Identification.
- Pathway analysis and biomarkers.

In conclusion, the workshop was very successful. It sparked interesting discussions, research collaborations, several joint grant proposals (e.g. for the BMBF program QuantPro), and joint publications. Other publications sparked by the seminar will certainly follow in the near future. The seminar also initiated the implementation of a webpage for interchanging proteomics data (www.computationalproteomics.net). The success of the seminar and the positive feedback of the participants encourage us to organize a follow-up for this style of meeting.

11.2 Ontologies and Text Mining for Life Sciences: Current Status and Future Perspectives

Seminar No. **08131** Date **24.03.–28.03.2008**
Organizers: Michael Ashburner, Ulf Leser, Dietrich Rebholz-Schuhmann

Researchers in Text Mining and researchers active in developing ontological resources provide solutions to preserve semantic information properly, i.e. in ontologies and/or fact databases. Researchers from both fields tend to work independently from each other, but there is a shared interest to profit from ongoing research in the complementary domain. The relatedness of both domains has led to the idea to organize a workshop that brings together members of both research domains.

Life Science researchers deliver their findings in scientific publications. These documents are nowadays distributed electronically and increasingly processed by automatic means to also incorporate those findings and the data into structured, scientific databases. Methods for this purpose are generally subsumed under the term "Text Mining", encompassing techniques belonging to the fields of machine learning, information retrieval and natural language processing. Text Mining-based solutions have, for instance, been developed for the identification of protein-protein interactions, of gene regulatory events, for the functional annotation of proteins, for the identification and prioritization of disease-related genes, and for the analysis of results from high-throughput experiments.

Text Mining for the Life Sciences has received considerable interest over the last years and is now an established area for conferences and workshops (e.g., ISMB, KDD, ECCB, Coling, ACL, PSB) and has lead to international large-scale challenge events (KDD-Cup, Genomics track at TREC, BioCreative2&2, BioNLP). The cause for this interest is the ever increasing amount of publications imposing an unbearable work burden on the individual researcher and the promising advances in natural language processing and machine learning that form the solution to the problem, if they are integrated into biomedical applications.

Text Mining has to cope with a large semantic gap between the raw textual data and the representation of meaningful results in databases, e.g., normalization of events in the text to conceptual representations of events according to "textbook" knowledge. It is hoped that ontologies fill this gap delivering a structured representation of biomedical knowledge. Although large and increasingly comprehensive biological ontologies are now available for many relevant topics (e.g. Gene Ontology, Sequence Ontology, Phenotype Ontologies etc.), it has not yet been proven what type of resources are ideally suited for Text Mining solutions.

Investigating on the aims of research in Text Mining and in ontological design, we find that ontologies are not designed to support Text Mining but rather to improve the annotation of database content. Although, Text Mining solutions intend to fill data-bases with content, it is not the case that Text Mining solution find ontological concepts easily in the literature, and, even more, ontological resources are not designed to support Text Mining solutions in the sense that the ontological terms fit to the demands of a natural language processing system. However, the Text Mining community exploits ontological resources to link generated evidence from the literature to the ontological concepts. Furthermore, the ontologies are not only a tool, but also a target for Text Mining research. Plenty of methods have been devised that automatically or semi-automatically construct ontologies or enrich existing ontologies by extracting terms and relationships from biomedical text collections.

These areas are researched by a community of researchers working in a highly interdisciplinary way in the domains of biology, biochemistry, chemistry, medicine, machine learning, formal ontologies, natural language processing, bioinformatics and others. It was the aim for this seminar to bring together researchers from all those areas to investigate on the state-of-the-art in both research fields, to discuss the suitability and progress of available resources, to identify areas where we are lacking tools, standards, or resources, and to foster joint opportunities for Text Mining and ontological research for the benefits of life science research.

In preparation of the seminar and prior to the meeting, the organizers identified three areas that best highlight the achievements and challenges in bringing together ontologies, Text Mining, and biological research:

1. exploring the benefits resulting from improved relations between Text Mining and biological ontologies,
2. technical advances in Text Mining and their application to life science research, and,
3. impact of advanced natural language processing (NLP) methods, and
4. success stories of Text Mining solutions with and without ontological support.

The seminar brought together more than 40 internationally renowned researchers from all domains mentioned beforehand. The ambience of the seminar is best described with the concept of a prolonged, lively and heated discussion. The discussion was mainly driven by the divergence of requirements, goals, and expectations between the Text Mining and

the ontology community. On the other side, a number of talks have pointed out the successful integration of Text Mining solutions into research in ontological design and the exploitation of ontological resources for successful Text Mining solutions.

11.3 Group Testing in the Life Sciences

Seminar No. **08301** Date **20.07.–25.07.2008**
Organizers: Alexander Schliep, Amin Shokrollahi, Nicolas Thierry-Mieg

Group testing AKA smart-pooling is a general strategy for minimizing the number of tests necessary for identifying positives among a large collection of items. It has the potential to efficiently identify and correct for experimental errors (false-positives and false-negatives). It can be used whenever tests can detect the presence of a positive in a group (or pool) of items, provided that positives are rare. Group testing has numerous applications in the life sciences, such as physical mapping, interactome mapping, drug-resistance screening or designing DNA-microarrays, and many connections to computer science, mathematics and communications, from error-correcting codes to combinatorial design theory and to statistics. The goal of the seminar is to bring together researchers representing the different communities working on group testing and experimentalists from the life sciences. We plan to address the following topics:

- Generalized group testing, where the choice of pools is constrained
- The decoding problem of inferring positives from the pool outcomes, to try to reconcile the stochastic and combinatorial formulations
- The real-world design problem of assigning items to pools, where the focus is on average-case performance rather than worst-case
- Applications in the life sciences, taking into account application-specific constraints on the design and decoding problems

The desired outcome of the seminar is a better understanding of the requirements for and the possibilities of group testing in the life sciences. Computer scientists should gain an increased understanding of the constraints imposed by the realities of wet lab experiments and the novel theoretical challenges arising from them. Biologists should obtain a clear view of the various smart-pooling methods and solutions that are available.

Chapter 12

Applications, Multi-Domain Work

12.1 Organic Computing – Controlled Self-organization

Seminar No. **08141** Date **30.03.–04.04.2008**
Organizers: Kirstie Bellman, Michael G. Hinchey, Christian Müller-Schloer, Hartmut Schmeck, Rolf Würtz

Organic Computing (OC) has become a challenging vision for the design of future information processing systems: As they become increasingly powerful, cheaper and smaller, our environment will be filled with collections of autonomous systems equipped with sensors and actuators to be aware of their environment, to communicate, and to organize themselves in order to perform the actions and services that seem to be required. However, due to increasing complexity we will not be able to explicitly design and manage all intelligent components of a digitally enhanced environment in every detail and anticipate every possible configuration. Therefore, our technical systems will have to act more independently, flexibly, and autonomously, i.e., they will have to exhibit life-like properties. We call such systems "organic". Hence, an "Organic Computing System" is a technical system, which adapts dynamically to the current conditions of its environment. It will be self-organizing, self-configuring, self-healing, self-protecting, self-explaining, and context-aware. After the successful initial Dagstuhl Seminar on Organic Computing in January 2006 with its emphasis on "Controlled Emergence" this seminar focused on controlled self-organization (SO). The major objective of the seminar was to explore the question "How can we build useful self-organizing systems?" This was expressed by three main topics for the seminar:

1. Basic understanding of self-organization
2. Organization of technical SO systems
3. Design of SO systems

The seminar was attended by 32 participants with the majority coming from Germany and a strong fraction from the United States. Starting with an extensive introductory

session, the seminar was organized as a sequence of couples of short presentations followed by intensive discussions, triggered by the presentations and by explicit questions on their overall topic. This more or less created a sequence of panels. The emphasis on discussions inspired a lively exchange of ideas. The first session on "Distributed self-organizing applications" presented generic distributed architectures (reconfigurable hardware, middleware), and applications like air traffic control (with highly strict security and safety requirements) and smart camera systems.

More general talks were presented on Generic Organic Computing architectures and wrappings as a form of test environment for complex systems. It became clear, that the application of self-organizing systems is not confined to toy applications. Rather, they are required to be built around legacy systems to keep these under control. This requirement is particularly strong in hardware design. The need for learning at design time and runtime was emphasized. There are significant commonalities between complex hardware and software systems. Self-organized scheduling for the parallelization of optimization procedures was a new example for this. Thursday afternoon was devoted to working groups.

As a direction for future applications multi-application test-beds were envisioned that would make rapidly changing objectives tractable. This will probably be robot playgrounds and surveillance scenarios. The talks of the seminar clearly demonstrated a range of applications where principles of OC have been used successfully. But, definitely, there is an urgent need for more investigations on how we can find adequate methods for managing the complexity of self-adaptive and self-organizing systems. The demand is obvious, and good partial solutions are already there.

12.2 Contextual and Social Media Understanding and Usage

Seminar No. **08251** Date **15.06.–20.06.2008**
Organizers: Susanne Boll, Mohan S. Kankanhalli, Gopal Pingali, Svetha Venkatesh

Traditional multi-media research focused on signal analysis to capture the semantics of media content: video, audio, and image analysis led to semi-automatic understanding of media. However, telling a sunset from a sunflower is still difficult and it is obvious that looking at information inherent in media is not enough. Instead, taking into account synergies between different media and contextual information could help to close the semantic gap. This is exemplified by Flickr that introduced a new way for communities to share and tag photos. While tagging does not solve the problem completely it introduced a new perspective on the situational usage of media, as well as the co-presence of objects and persons. We believe that we are now at the doorstep to a new decade of contextual and social understanding of media.

In this seminar we will explore how content, context, and social community influence media understanding. We will examine potentials and challenges, research directions, and application domains that range from personal media and news to biomedicine and

robotics. The seminar intends to bring together the research community (from academia, labs, and industry) to make possible that capturing, storing, finding, and using digital media becomes an everyday activity in our computing environment. We aim to uncover the role of context and social aspects for tomorrow's multimedia content understanding and media usage.

The objective of this seminar is to build a solid foundation for incorporating contextual information to aid media understanding, analysis, and interpretation. Some contextual data is symbolic while other is in signal form. It is not clear how to reconcile both. Likewise, it is a challenge to characterize the minimal contextual information necessary for disambiguation. Furthermore, we will look at the system infrastructure (bandwidth, connectivity, computing resources) that are implied by our approach. As part of this we are interested in the following topics:

- Multimedia content analysis, indexing, and retrieval exploiting user and usage context
- The role of context in the multimedia life cycle
- Studying social behavior and social networks in and for media usage
- Media search, delivery and consumption to the individual user's context and situation
- Smart acquisition, indexing, distribution, sharing, and exploitation of media context
- Browsing, mining, searching media content in contextual and social spaces
- Ubiquitous experiential environments
- Including the human into media understanding and usage Systems support for social media

12.3 Computer Science in Sport – Mission and Methods

Seminar No. **08372** Date **07.09.–10.09.2008**
Organizers: Arnold Baca, Martin Lames, Keith Lyons, Bernhard Nebel, Josef Wiemeyer

From September 7 to 10, 2008 about 30 experts from computer science and sport science met at the Leibniz-Zentrum für Informatik in Dagstuhl to discuss interdisciplinary issues in the area of computer science in sport. Five topics were selected for discussion: doping, modeling and simulation, pervasive computing, robotics and sport technology. A total of 17 presentations dealt with selected projects and issues in the above-mentioned fields.

- Doping – an individual decision or a social phenomenon?

- Modeling and simulation – between reduction and abundance
- Robotics – two directions of transfer: from humans to robots and back
- Pervasive computing – technology meets human needs
- Sport technology – the view of practice
- Mission statement

The evaluation of the seminar showed that the seminar

- inspired new ideas for further work (research, development or teaching),
- inspired joint projects, joint development, or joint publications,
- led to insights from neighboring fields or communities and
- identified new research directions.

Finally, there was a clear consensus of the participants to continue this kind of exchange.

12.4 Social Web Communities

Seminar No. **08391** Date **21.09.–26.09.2008**
Organizers: Harith Alani, Steffen Staab, Gerd Stumme

Blogs, Wikis, and Social Bookmark Tools have rapidly emerged on the Web. The reasons for their immediate success are that people are happy to share information, and that these tools provide an infrastructure for doing so without requiring any specific skills. At the moment, there exists no foundational research for these systems, and they provide only very simple structures for organising knowledge. Individual users create their own structures, but these can currently not be exploited for knowledge sharing. The objective of the seminar was to provide theoretical foundations for upcoming Web 2.0 applications and to investigate further applications that go beyond bookmark- and file-sharing. The main research question can be summarized as follows: How will current and emerging resource sharing systems support users to leverage more knowledge and power from the information they share on Web 2.0 applications? Research areas like Semantic Web, Machine Learning, Information Retrieval, Information Extraction, Social Network Analysis, Natural Language Processing, Library and Information Sciences, and Hypermedia Systems have been working for a while on these questions. In the workshop, researchers from these areas came together to assess the state of the art and to set up a road map describing the next steps towards the next generation of social software.

Topic of the Seminar

Within the last two years, social software on the Web, such as Flickr, Delicious, Bibsonomy, Facebook, etc., has received a tremendous impact with regard to hundred of millions of users. A key factor to the success of social software tools in the Web is their grass-roots approach to sharing of information between users: there are no limitations on the kind of tags users may select. The resulting structures are often called 'folksonomies', that is, 'taxonomies' created by 'folks'.

Such systems are also considered to realize a Web version 2.0. The reason is that the initial use of the Web could be characterized by many users consuming what a comparatively small set of producers had developed, whereas with social software on the Web, everyone becomes a prosumer, i.e. someone who produces and consumes content. The success of this approach is visible with applications like flickr, which had approximately 250,000 users in April 2006. In the reference sharing systems CiteULike and Connotea researchers and others insert, tag, and recommend scientific references in a shared knowledge space. This indicates a currently ongoing grass-root creation of knowledge spaces on the Web which is closely in line with "the 2010 goals of the European Union of bringing IST applications and services to everyone, every home, every school and to all businesses" .

The reason for the apparent success of the upcoming tools for web cooperation (wikis, blogs, etc.) and resource sharing (social bookmark systems, photo sharing systems, etc.) lies mainly in the fact that no specific skills are needed for publishing and editing. As these systems grow larger, however, the users will feel the need for more structure to better organize their resources and enhance search and retrieval. For instance, approaches for tagging tags, or for bundling them, are currently discussed on the corresponding news groups. We anticipate that resource sharing systems, together with wikis and blogs, are only first appearances of an emerging family of Web 2.0 tools.

12.5 Perspectives Workshop: Virtual games, interactive hosted services and user-generated content in Web 2.0

Seminar No. **08393** Date **24.09.–27.09.2008**
Organizers: Thomas Hoeren, R.K. Murti Poolla, Gottfried Vossen

Recently, almost everything seems to have become "2.0", be it music, gadgets, health, entertainment, business, Silicon Valley, countries such as India, the family, and, most notably, the Web. 10GB of "user-generated content" is created in the World-Wide Web daily (see Ramakrishnan and Tomkins, 2007), that is, more than five times the amount of content created by professional Web editors. Web 2.0 has rapidly become a label that everybody using the Internet and doing business through it seems to be able to relate to; what it primarily stands for is the transition of the Web from a medium where people just read information to a medium where people both read and write; in other words, the Web

meanwhile heavily benefits from user contributions and user-generated content (UGC) in a variety of media forms. This has been enabled by technological advances that nowadays make it possible for users to easily employ services offered on the Web and to embark on tasks that have previously been reserved for specialists.

UGC can primarily be observed in the consumer area, but is also entering enterprises. Especially in the former, numerous legal issues arise, which is demonstrated by the large number of cases from this field that courts of laws have to deal with recently. This situation is due to a number of reasons, including the fact that legal restrictions are often ignored, or that users are unaware of the laws they may be or are violating. The goal of this manifest, which contains the findings of a Dagstuhl Perspectives Workshop held at Schloss Dagstuhl, Germany in September 2008, is to shed some light on the interplay between law and Web 2.0 and to discuss a number of questions and issues that urgently deserve clarification.

There is a follow-up publication for this seminar:
Thomas Hoeren, Gottfried Vossen: Manifest: The role of law in an electronic world dominated by Web 2.0.
Computer Science: Research and Development 23.2009,1, p. 7-13.