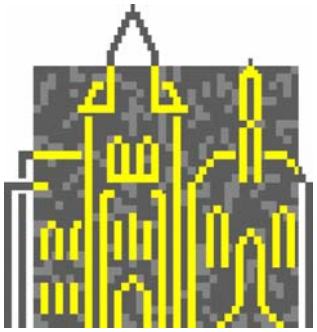

SCHLOSS DAGSTUHL

INTERNATIONAL
CONFERENCE
AND RESEARCH CENTER
FOR COMPUTER SCIENCE

Dagstuhl News

January - December 2000

Volume 3
2001

ISSN 1438-7581

Period: January - December 2000

Frequency: 1 per year

The International Conference and Research Center for Computer Science 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.

Associates: Gesellschaft für Informatik e.V., Bonn
Technische Universität Darmstadt
Universität Frankfurt
Universität Kaiserslautern
Universität Karlsruhe
Universität Stuttgart
Universität Trier
Universität des Saarlandes

Funding: The state governments of Saarland and Rhineland Palatinate

Address: IBFI Schloß Dagstuhl
Octavieallee
D-66687 Wadern
Tel.: +49 - 6871 - 905127
Fax: +49 - 6871 - 905130
E-mail: service@dagstuhl.de
Internet: http://www.dagstuhl.de/

Welcome

You have in your hands the third edition of the "Dagstuhl News", a publication for the members of the Foundation "Informatikzentrum Schloss Dagstuhl", the *Dagstuhl Foundation* for short.

The big event in the year 2000 was our 10th Anniversary Conference, "Informatics – 10 Years Back, 10 Years Ahead". The conference offered very interesting talks by leading scientists of our discipline. They leaned back, looked what has happened in their respective areas during the existence of Dagstuhl and tried to speculate what will happen in the near future. No report about this conference is contained in these News, but a proceedings volume has been published as LNCS volume 2000 by Springer Verlag.

The main part of this leaflet consists of collected resumees and other valuable information taken from the Dagstuhl-Seminar Reports. 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 2000 are on the Web under URL:
`www.dagstuhl.de/DATA/Seminars/00`

The State and the Activities of the *Dagstuhl Foundation*

The foundation currently has 42 personal members and 8 institutional members.

In 2000, the foundation has supported a few guests with travel grants and a reduction of the Seminar fees. According to German law only the interests earned can be used to support the aims of a foundation.

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)

Contents

1 Constraint Programming and Integer Programming

Seminar No. **00031** Report No. **262** Date **16.01.–21.01.2000**
Organizers: Krzysztof Apt, Michael Jünger, Pascal van Hentenryck, Laurence A. Wolsey

In the **Mathematical Programming Community** (consisting mainly of mathematicians and operations researchers) **Integer Programming (IP)** started flourishing in the 1950's when cutting plane and branch & bound algorithms were proposed for the solution of

$$\text{(IPP)} \qquad \begin{array}{ll} \text{maximize} & c^T x \\ \text{subject to} & Ax \leq b \\ & x \text{ integer} \end{array}$$

where c and b are n- and m-dimensional vectors, respectively, and A is an m by n matrix of rational numbers. Refinements and hybrids of both techniques led to powerful algorithms for IPP as well as combinatorial optimization problems with linear objective function

$$\text{(COP)} \qquad \begin{array}{ll} \text{maximize} & \sum_{e \in F} c_e \\ \text{subject to} & F \in \mathcal{F} \end{array}$$

where for a finite set E a collection of subsets $\mathcal{F} \subseteq 2^E$ defines the feasible solutions. Prominent successes are, e.g., the most powerful software packages for the traveling salesman problem.

In the **Artificial Intelligence Community** (consisting mainly of logicians and computer scientists) **Constraint Programming (CP)** started flourishing in the 1980's after constraint satisfaction problems had been a focus of research in the 1970's. Today's state of the art includes several systems that support programming language constructs whose expressive power goes far beyond linear equations and inequalities to include logical and higher-order constraints, global constraints, as well as specific support for scheduling and resource allocation problems. These languages also let users specify and control a search procedure appropriate for a given application. Prominent successes are, e.g., in the area of production scheduling.

Many areas of applications of constraint programming techniques are beyond the scope of integer programming techniques, but Integer Programming and Constraint Programming are complementary techniques

for combinatorial optimization problems of the form COP. Depending on the problem, one or the other technique is more appropriate.

Software for general integer programming, and special purpose software for structured problems such as the traveling salesman problem or the max cut problem, almost always depends on the formulation (IPP). Common features of the algorithms are a priori preprocessing, enumeration in which the bounds provided by linear programming play a crucial role in pruning the enumeration tree, and cutting planes that are used to tighten the linear programming bounds, and push the linear programming solution closer to integrality.

Using an IPP formulation is an advantage in terms of generality, but as concerns flexibility,

1. modelling certain simple situations/expressions may require a very large number of variables,

2. the preprocessor must maintain an IPP formulation, and therefore only a restricted set of transformations can be carried out,

3. certain IPP formulations have very weak linear programming relaxations so the bounds provided are useless, and

4. the choice of how to branch is also restricted by the IPP formulation.

CP is also based on enumeration, but typically has a richer language to represent a problem, allowing for more refined preprocessing and branching. The basic idea behind CP systems is to enforce some consistency notions on the constraints (e.g., arc-consistency) at each node of the search tree. These consistency notions define in fact relaxations of the problem. In general, these relaxations are specific to classes of applications (e.g. edge finder in scheduling) and are used to reduce the set of possible values that the decision variables can take. In so doing, they also produce lower bounds for optimization problems. However, CP systems have not exploited so far the wealth of results provided by linear programming relaxations and their associated cuts.

Only very recently, first scientific links between IP and CP were installed, but still, both subjects develop independently to a large extent. Reasons include the different scientific contexts in which the two disciplines developed. This results in language barriers. Integer programmers

use mathematical concepts that are unfamiliar to constraint programmers who in turn use computer science concepts that are unfamiliar to integer programmers.

After the seminar, the organizers are even more convinced that combinations of IP and CP techniques will have an increasing impact on combinatorial problem solving, and that the potential of joint efforts is largely unexplored. On Monday and Tuesday morning we tried to overcome the language barriers by asking prominent researchers to give introductions for the other community into the most relevant topics. This first part required a thorough preparation of the lectures and was be planned in advance (which is unusual for a Dagstuhl seminar). From Tuesday afternoon on, the seminar followed the usual scheme: The talks were scheduled "on-line" according to need, and gradually, more and more interaction between the participants of the two communities developed. On Thursday evening, we had a panel discussion with two panelists of either community in which the situation was reflected and it was tried to identify problems for which a fruitful interaction/competition of constraint programming techniques and integer/combinatorial optimization techniques appears challenging and is likely to enhance the interaction of both communities. We are confident that the seminar was an important step towards this goal.

2 Linear, Semidefinite Programming and Randomization Methods for Combinatorial Optimization Problems

Seminar No. **00041** Report No. **263** Date **23.01.–28.01.2000**
Organizers: Klaus Jansen, Jose Rolim, Madhu Sudan

During the week of January 23 - 28, 2000, the Seminar on Linear, Semidefinite Programming and Randomization Methods for Combinatorial Optimization Problems was organized by K. Jansen (Kiel), J. Rolim (Geneva) and Madhu Sudan (MIT Cambridge). 41 Participants came from universities or research institutes from Austria, Israel, Greece, France, Germany, Netherlands, Russia, Switzerland, United Kingdom and United States America.

The objective of the seminar was to focus on algorithmic and complexity aspects arising in the development of efficient solution techniques for computationally difficult optimization problems. The seminar promoted an exchange of different methods used in approximation algorithms. A general technique for efficient approximation algorithms is to formulate an optimization problem as an integer linear program and then to relax the integrality conditions. Recently, there has been striking success in obtaining also approximation algorithms on more general mathematical programming such as semidefinite programming. In this and other context, randomization has proved to be a powerful algorithmic method: it yields to simple and easy to analyse algorithms for many optimization problems, and it leads to a better performance guarantee.

Main topics of interests were:

- extending of randomization and semidefinite programming techniques to other optimization problems,
- improved approximation algorithms and structural insights by studying linear programming, semidefinite programming and randomization,
- development of approaches to solve (approximatively) large linear and semidefinite programs,
- complexity questions for randomization and semidefinite programming,
- practical implementation of the used techniques (randomization, semidefinite programming),
- exchange of informations on recent research and stimulation of further research.

The seminar was intended to bring together researchers from different areas in combinatorial optimization and from applications. Different algorithmic methods and techniques have been covered by 25 lectures. There was also an open problem session on Thursday evening and a lively discussion on problems from different fields of application. Talks were given studying practical problems like facility location, scheduling, load balancing, data broadcasting, discrete tomography, and curve reconstruction and also theoretical oriented problems on satisfiability, polymatroids,

lattices , semidefinite relaxations, metarounding, probabilistic methods, discrepancies, sorting by reversals and structured optimization.

3 Theory of Evolutionary Algorithms

Seminar No. **00071** Report No. **265** Date **13.02.–18.02.2000**
Organizers: H.-G. Beyer, K. De Jong, D. Fogel, I. Wegener

During the last ten years computers have become fast enough to support evolutionary algorithms and a lot of applications to real-world problems have been developed. This has led to a great deal of empirical knowledge on the behavior of evolutionary algorithms and to many heuristics for choosing their associated parameters. There is also a developing theory of evolutionary algorithms based on tools from the analysis of randomized algorithms, of Markov processes, and of dynamical systems. The aim of this workshop was to contribute to this theory and to allow a discussion between researchers with different backgrounds.

The organizers are happy to report that 45 researchers accepted an invitation to Dagstuhl. They came from Germany (21), England (8), USA (5), France (2), Netherlands (2), Romania (2), Austria (1), Belgium (1), India (1), Mexico (1), and Poland (1). The 31 talks captured all the aspects of a theory of evolutionary algorithms, among them statistical dynamics, time-varying landscapes, convergence issues, complexity results, fitness landscapes, models of evolutionary algorithms, analysis of the run time of evolutionary algorithms, self-adaptation, new variants of evolutionary algorithms, and genetic programming. The discussion was extremely vivid. There was almost no talk that evoked fewer than five questions and remarks. The schedule included an informal evening session where eight topics suggested by the participants were discussed.

Besides the official schedule the participants used unscheduled time for many discussions and some informal sessions with short talks, all inspired by the special Dagstuhl atmosphere. The special event of the week was the Wednesday hike where it has snowed heavily on the way out and the sun shone on the way back through the snow.

4 Specification of Distributed Information Systems

Seminar No. **00081** Report No. **266** Date **20.02.–25.02.2000**
Organizers: Hans-Dieter Ehrich, Ursula Goltz, José Meseguer, Amir Pnueli

The specification, development and use of distributed information systems is an important research area with many practical applications. Examples for existing such systems are those owned by banks, airlines or governments. Moreover, the explosive growth of the Internet is evidence that a global information structure is developing, and expansion in the use of large distributed information systems can be predicted.

Information systems are *reactive* systems. Unlike transformational systems—that is, systems whose function is to transform some inputs into output results—reactive systems have an unlimited number of interactions with their environment. Basic notions of program correctness do not apply, and techniques for designing transformational systems do not easily carry over to reactive systems. Non-distributed reactive systems, like conventional operating and database systems, are not easy to design, implement and restructure, a lot remains to be done to develop helpful high-level concepts, features, languages, methods and tools. However, existing products can already cope successfully with reasonably complex systems in practice. Formal methods from logic, algebra and process theory have been successfully applied, for instance in query and transaction processing, but they are not in general standard engineering practice yet.

Distributed reactive systems are still an order of magnitude more complex and difficult to specify, analyse, implement and reconfigure. They are not well understood, and methods for designing, implementing and reengineering them have not yet reached a level of mature and sound engineering practice. A great challenge is to cope with *concurrency, synchronization, and communication* at an appropriate level of abstraction. This can be achieved by applying concepts from concurrency theory for distributed information systems.

This seminar was the follow-up of the Dagstuhl Seminar 98071 "Information Systems as Reactive Systems" which took place in February 1998. Until that seminar, there was little interaction between the communities concerned with information systems and concurrency theory. One out-

come of the seminar was that there is a definite need for interaction; new contacts have been established. The focus of the present seminar was again to bring experts from the respective fields together. In contrast to the previous seminar, we put even more focus on applications, seeing the design of distributed information systems not only as theoretically challenging but also as an engineering problem.

In particular, the following topics have been addressed:

- Modelling and design methods.
  Information systems involve data, objects and global processes. In conceptual modeling, these aspects have been addressed in isolation and various combinations, and the same holds for the conceptual modeling perspectives: system architecture, object classes, object behavior, and object communication. Diagrammatic modeling approaches like OOA&D, OMT, OOSE and their recent amalgamation UML employ a multitude of specification concepts and techniques that are not always well integrated and sometimes not well understood. Approaches for giving formal semantics for (subsets of) UML would give a more clear picture of the underlying system model.

  - Activity charts contain ambiguities and many fuzzy constructs like swimlanes. The formalization of activity charts in terms of clocked Kripke structures was discussed, for which ATCTL can be used as property specification language to do model checking.
  - UML sequence diagrams can be used to specify requirements. An approach was introduced in which sequence diagrams with timing contraints and a system modeled in Uppaal timed automata are checked for consistency. For this the sequence diagrams were translated in Uppaal automata and consistency is verified by the Uppaal model checker.
  - Validation of a formal specification is concerned with the question whether a specification precisely reflects the informal requirements. One way to validate process specifications is the generation and visualisation of process nets from a system model given by a Petri net.
  - Whenever the visual means of a modeling language will change, a re-implementation of supporting tools is necessary. In or-

der to avoid such re-implementation, a generator for graphic editors was presented. In this approach a visual language is specified by a visual grammar which is based on a visual alphabet.

- Semantics of object-oriented specification languages.
  Describing the semantics of a formal language amounts to giving a map from syntactic constructs into a suitable semantic domain. For object-oriented specification languages, possible domains are process algebras, Petri nets, event structures, the actor model, type algebra, and others. Studies on relating and evaluating such domains are necessary. Interesting language issues are (amongst others) module composition, refinement and atomicity.

  - The use of labelled transitions systems was considered for the purpose of specifying and describing the behaviour of object-based systems.
  - MSCs possess a standard semantics based on process algebra which is rather complicated. An alternative are the theory of traces and partially commutative monoids. The trace based semantics is bisimular with the operational semantics and allows to apply the rich theory of partially commutative monoids with associated algorithms and tools.
  - It is important that integrity and correctness of specifications on higher levels are not damaged by a translation by a compiler. It was shown how the result of a translation can be verified.

- Logics for distributed systems.
  An attractive model for the execution of a distributed system is a *distributed process*, which has a time line for each location of the system. Temporal logics based on this notion are particularly suited for the specification of distributed systems. Rewriting logic is another well known and quite successful paradigm in this field. There is a multi-purpose specification language, Maude, that is based on rewriting techniques. Mobile Maude is a mobile agent language extending Maude and supporting mobile computation. Mobile Maude uses reflection to obtain a simple and general declarative mobile language design and makes strong assurances of mobile agent behaviour.

- The question was discussed how global conditions for object systems can be split and assigned to a group of objects such that these objects together satisfy the global condition. One approach is to give the global statements in a distributed logic which allows for local assertions about the states of other objects in a local logic.
- A semantic model of both external and internal dynamics of agent systems was given which uses locality and temporal compatibility as basic concepts.
- A semantic approach was proposed which makes precise the reflective concept of composable services (security, faulttolerance, performance) in a distributed system. The approach is based on the executable formal semantics for distributed object-oriented systems provided by rewriting logic.
- To deal with open system specifications a talk uses tile logic, where both closed and open terms are managed analogously. This allows to analyse the *bisimularity as congruence* property for serveral tile formats that accomplish different concepts of subterm sharing.
- The recent developments in the area of automatic verification of logics referring to the distribution of a system were summerized. These logics refer to the causal order of events happening in a distributed system, either referring to infinite partial order of executions or to the prefix ordering of finite partial order executions.

• Mobility.
For modeling large systems of communicating objects, it is indispensible to deal with a dynamically changing communication structure. In process algebras, this issue has been formally dealt with in the π-calculus and, more recently, in the fusion calculus.

- To verify temporal and functional properties cast in terms of the π-calculus, a π-μ-calculus (extension of the modal μ-calculus) was introduced as an appropriate temporal logic for the π -calculus. Moreover a variant of the π-calculus, known as spi-calculus, was used for the specification of secure communication protocols and for verifying security properties by means of equivalence checking.

- Application-oriented concepts in specification languages.
  Among the issues to be addressed are real time, fault tolerance, and internet applications.

  - A profile for modeling multimedia applications was presented. The profile OMMMA-L (Object-oriented Modeling of Multimedia Application - the Language) introduces a so-called presentation diagram to model the spatial aspect (layout) of a multimedia application.
  - A framework was introduced which uses the codesign approach based on an abstraction layer model to develop information-intensive internet sites.
  - A talk discussed media as model to envision, to design and to formalize platforms for communities of collaborating human or artificial agents. In this approach the media concept envisions media as plattforms of multi-agent systems and the media reference model determines the main components of a medium and guides its application.
  - A new formalism for modeling real-time systems in different application areas and on various levels of abstraction was presented. For I/O interval structures in the model, efficient model checking and analysis techniques based on MTBDDs exist in this approach.
  - Information system objects are subject to frequently occurring modifications of behavior rules. A proposal was discussed for extending object specification techniques with capabilities to supporting evolving behavior descriptions. The semantics of these added language constructs can be expressed e.g. using an extended temporal logic framework.
  - The method Focus offers a formal framework for topdown-development of distributed, reactive systems. It was demonstrated how this method can be applied to the modeling of operating system concepts.

The programme of the seminar was intense and stimulating; it comprised 23 talks, the abstracts are recorded in this report in alphabetical order. Some of the basic principles of verification of concurrent systems are summarized in a new book (authors: Willem-Paul de Roever etal)

which was presented at the workshop. We had an interesting panel discussion in which the panelists discussed the question if and how results of the concurrency theory can help to develop distributed information systems and which are the most promising aspects from which distributed information systems can benefit.

5 Data Structures

Seminar No. **00091** Report No. **267** Date **27.02.–03.03.2000**
Organizers: Susanne Albers, Ian Munro, Peter Widmayer

The area of Data Structures continues to be an important and vibrant aspect of computer science. The topic is an essential component in the algorithmic solution of many problems. Although data structures have been studied for more than forty years, there is still a large research community working on exciting and challenging problems. The Fifth Dagstuhl Seminar on Data Structures was attended by over 50 people from 10 different countries; and so, was substantially larger than previous meetings. Attendees included many young researchers, e.g. there were 10 Ph.D. students. A third of the participants were attending the seminar for the first time, bringing new ideas and points of view.

The workshop presentations addressed classical data structuring problems as well as new problems arising in important applications. Interesting results were presented on classical issues such as search trees, priority queues, hashing and predecessor problems. Moreover, there were contributions on graph problems such as DFS and shortest paths computations. Several talks were concerned with geometric problems such as range searching and convex hull computations. There was an increase in the number of presentations on external memory algorithms, relative to previous meetings. Last but not least, there were several contributions investigating data structure problems in specific application areas such as Parallel and Distributed Computing, Computational Biology and Database Systems.

In an open problem session many challenging research problems were proposed. Indeed solutions to two of the open problems posed in this session were presented later in the week. As usual, most participants used the seminar to develop ongoing research collaborations and to form new ones.

6 Multi-Image Search, Filtering, Reasoning and Visualization

Seminar No. **00111** Report No. **268** Date **12.03.–17.03.2000**
Organizers: Alfred Bruckstein, Thomas Huang, Reinhard Klette, SongDe Ma

Many problems in computer vision involve multiple images. These could be images taken by multiple cameras of the same scene, video sequences or multispectral images, etc. This workshop was concerned with the theoretical, algorithmic and implementation issues in multi-image acquisition, storage, retrieval, processing, analysis, manipulation and visualization. 38 talks were presented in 10 technical sessions: active vision; motion detection, modeling and understanding; omnidirectional vision; terrain reconstruction and calibration; scene reconstruction and image separation; object and scene visualization; image and video databases; object shape recovery; pose estimation, object recognition and modeling; and image properties and reasoning.

In addition, 5 working groups (WG) were organized: learning, action and development; image and video databases; visualization; color and photometry; new camera technologies. Three talks were presented in these groups. The charge to each WG was to assess the state-of-the-art of the given topic and to identify challenging and promising research directions.

A refereed proceedings book has been published:
Reinhard Klette, Thomas S. Huang, Georgy L. Gimelfarb (eds.)
Multi-image analysis: 10th international workshop on theoretical foundations of computer vision
Springer 2001, LNCS 2032

WG I: Learning, Action and Development

Chairs: Claus-E. Liedtke, John Weng

The topic of learning requires a definition of the term. According to Michalski et.al. one possible definition is the following: “Learning denotes changes in the system that are adaptive in the sense that they enable the system to do the same task drawn from the same population more efficiently and more effectively the next time”.

The members of the working group work presently themselves or are interested in

- learning of robot actions using reinforcement strategies,
- learning receptive fields,
- learning object recognition from multiple views of objects,
- learning explicit formulated knowledge contents from images like the the rules of a rule-based system or the scene structure represented by semantic nets,
- real-time learning from sensors by active robots, etc.

Aspects of learning which have been discussed included the model representations of the models to be learned and learning paradigms. Technical challenges which were recognized included

- slow convergence of reinforcement learning,
- automated learning, direct from images,
- autonomous learning,
- automated learning of scene structure,
- learning using a small learning set, especially learning from one sample.

Machine learning is often seen in competition with human learning abilities. In this connection the working group discussed a new approach called developmental approach. It is motivated by human mental development from infancy to adulthood. Development includes learning but requires more. Some major differences between conventional learning and the automatic development for mental skills are:

1. In conventional machine learning, the task is given, but in automatic development, the tasks that the machine will learn are unknown to the system designer.
2. In the former, the human designer defines features, but in automatic development, the system automatically derives features from sensory inputs.
3. In the former, the internal representation is pre-designed by the human according to the task given but in the latter, the representation is automatically generated according to the developmental scheme and the sensory data.

4. In the former, invariance depends on the predefined features and the rules that the human designer has specified. In the latter, the invariance is automatically achieved through experience as a generalization capability.

More information about development is available at http://www.cse.msu.edu/dl/ on an upcoming Workshop on Development and Learning (WDL) funded by National Science Foundation (NSF) and Defense Advanced Research Project Agency (DARPA) to be held at Michigan State University.

A discussion about some basics of learning from vision sensors came to the following conclusion:

Learning is always related to actions to be taken or tasks to be performed. Therefore learning from images is related to a model, which maps visual images onto actions. The model can be structured into different layers, which are again substructured into different domains. This seems to be true for any recognition process. The bottom layer is the sensor layer, the highest layer is the intelligent action layer. For machine learning from images the question is how to identify the layers and domains and how to interrelate them. One approach which has been used is to model the layers and domains as a semantic net. There we have different abstraction levels from the sensor layer at the bottom, which is directly connected to the images, up to the most abstract layer at the top, which represents the understanding of the visualized scene. Within the layers there are the nodes of the semantic net which we call concepts. The concepts are linked by relations, which do not have merely some descriptive property but influence actively the automated interpretation process. Right now the levels of the semantic net, the concepts and the relations have to be determined by a human from his expert knowledge. Autonomous learning would mean to find the layers and concepts automatically, and this is still an unsolved problem.

As was mentioned, learning is based on a model which links sensor signals with tasks. The sensor information has to be processed such, that the task can be performed. If many vision based tasks have successfully been completed through many experiments, then one can observe, that certain processing steps and certain intermediate processing results appear again and again. It can further be observed, that some steps and some intermediate results differ from application to application and from situation to situation. We call the latter the variants and the first ones

the invariants. One objective of learning is to identify the variants and invariants on the different levels of abstraction. In this connection the question was raised to which extent should unsupervised learning be used and to which extent supervised learning. Improvement of learning, i.e. learning on a higher level, is the search for variants and invariants not only on the sensor level but on the level of previously identified "higher" invariants. So learning may become more efficient.

In order to implement vision based tasks by learning one need not learn everything from scratch. We have already lots of experience and know how to solve some sensor based action problems. Referring to the visual sensor the variants and invariants are related to spatial-temporal effects. So we know that some useful invariants are gradients, surfaces, vector-fields, flow, etc.. We know that a good approach to describe the invariants of the world is to assume a 3D geometrical representation of the world. So the use of 3D-transforms like translations and rotations might be useful elements in the interrelations between the concepts. If we want to develop a powerful vision based action tool, with a high degree of flexibility, we have probably to start out developing a toolbox of known invariants first.

WG II: Image/Video Databases: Browsing, Searching, Learning and Summarization

Chair: Kap Luk Chan

... The group identifies a few promising research directions. It is generally recognized that many low level descriptors for color, texture and shape exist. However, there are missing links between the low level descriptors and their association to some high level concepts. The clustering or structuring of large databases is an issue to be addressed. The approach to the structuring can vary from static pre-clustering to adaptive or evolutionary clustering. In here, the issue is not only on the perceptual feature sets and their association to high level concepts but also the approach to learning, such as learning from users. There is a whole spectrum of possibilities for dynamic clustering of databases. The important possibility of human intervention in the searching and browsing of the image/video database should be recognized. The system should have the intelligence of learning from query and continuously re-structuring the database. Filtering of information during a search should facilitate faster convergence

the satisfactory results. Relevance feedback is an important mechanism not only for filtering but also plays a role in evolutionary structuring of databases. The summarization of video databases is an important functionality of the system, and is also a very challenging problem. Summarization allows extraction and collection of key information embedded in video and facilitates fast search and retrieval. ...

7 Semantics for the Web

Seminar No. **00121** Report No. **269** Date **19.03.–24.03.2000**
Organizers: Dieter Fensel, Jim Hendler, Henry Lieberman, Wolfgang Wahlster

The World-wide Web (WWW) has drastically changed the availability of electronically available information. Currently there are around 300 million static documents in the WWW which are used by more than 100 million users internationally. In addition, this number is growing astronomically. In 1990, the WWW began with a small number of documents as an in-house solution for around thousand users at CERN. By 2002, the standardization committee for the WWW (called W3C) expects around a billion web users and an even higher number of available documents. However, this success and exponential growth makes it increasingly difficult to find, to access, to present, and to maintain the information of use to a wide variety of users. Currently, pages on web must use representation means rooted in format languages such as HTML or SGML and make use of protocols that allow browsers to present information to human readers. The information content, however, is mainly presented by natural language. Thus, there is a wide gap between the information available for tools that try to address the problems above and the information kept in human readable form.

- **Searching for information:** Already, finding the right piece of information is often a nightmare. One gets lost in huge amounts of irrelevant information and may often miss the relevant ones. Searches are imprecise, often returning pointers to many thousands of pages (and this situation worsens as the web grows). In addition, a user must read retrieved documents in order to extract the desired information — so even once the page is found, the search

may be difficult or the information obscured. Thus, the same piece of knowledge must often be presented in different contexts and adapted to different users' needs and queries. However, the web lacks automated translation tools to allow this information to automatically be transformed between different representation formats and contexts.

- **Presenting information:** A related problem is that the maintenance of web sources has become very difficult. Keeping redundant information consistent and keeping information correct is hardly supported by current web tools, and thus the burden on a user to maintain the consistency is often overwhelming. This leads to a plethora of sites with inconsistent and/or contradictory information.

- **Electronic commerce:** Automatization of electronic commerce is seriously hampered by the way information is currently presented. Shopping agents use wrappers and heuristics to extract product informations from weakly structured textual information. However, development and maintenance costs are high and provided services are limited.

There is an emerging awareness that providing solutions to these problems requires that there be a machine understandable semantics for some or all of the information presented in the WWW. Achieving such a semantics requires:

- Developing languages for expressing machine understandable meta information for documents.

- Developing terminologies (i.e., name spaces or ontologies) using these languages and making them available on the web.

- Integrating and translating different terminologies.

- Developing tools that use such languages and terminologies to provide support in finding, accessing, presenting and maintaining information sources.

Developing such languages, ontologies, and tools is a wide ranging problem that touches on the research areas of a broad variety of research communities. Therefore this seminar brought together colleagues

from these different research communities (who would typically not meet at area conferences or other workshops that are more methodologically driven). These include researchers in the areas of databases, intelligent information integration, knowledge representation, knowledge engineering, information agents, knowledge management, information retrieval, meta data, web standards (RDF, XML, XML-QL, XSL), and others. The goal of this meeting in Dagstuhl was to bring together the scientists and technologists working in these areas, and to thus allow the exchange of information about emerging tools and techniques.

The contents of the seminar was organized as follows. First, we discussed a number of arising new web standards that should improve the representation of machine processable semantics of information. Second, we discussed the use of ontologies for representation of semantics (in the sense of formal and real-world semantics). Third, these semantic annotations allow automatization in information access and task achievement. Therefore, we discussed intelligent information access based on these semantic annotations. Forth, we discussed a number of applications of these new techniques and had a number of exiting demonstrations. Last but not least we had some workshops dealing with emerging issues during the seminar. The presentations are available at
`http://www.semanticweb.org/events/dagstuhl2000/`

Conclusions

Tim Berners-Lee, director of the World Wide Web consortium, has referred to the future of the current WWW as the "semantic web" — an extended web of machine-readable information and automated services going way beyond current capabilities. The explicit representation of the semantics underlying data, programs, pages, and other web resources, will enable a knowledge-based web providing a qualitatively new level of service. Automated services will be better able to assist humans in achieving their goals by "understanding" more of the content on the web and thus providing more accurate filtering, categorization, and search of information sources. This process will ultimately lead to an extremely knowledgeable system with various specialized reasoning services that will support us in nearly all aspects of our daily life — making access to information as pervasive, and necessary, as access to electricity is today.

In the systems of the future, information will not be simply a set of passive entities residing in a repository. Instead, active information

sources will play a critical role accessed via network-enabled, information-provision services. These services will not only support better extraction and search, but will also more directly support human task achievement. To make this possible, machine-understandable representation of semantics is required for the automated selection and combination of these reasoning services.

A key enabler for the semantic web is on-line ontological support for data, information and knowledge exchange. Given the exponential growth of on-line information available, automatic processing becomes mandatory for keeping it managed and accessible. Being used to describe the structure and semantics of information exchange, ontologies will play a key role in areas such as knowledge management, B2B e-commerce and other such burgeoning electronic enterprises.

In this workshop, many cutting edge papers were presented describing the state of the art in this emerging new research area. Participants from a number of different organizations described research activities in academia, industry and government. In addition, a description of the US Defense Advanced Research Project Agency's DAML (DARPA Agent Markup Language) project, was presented, and possible EC funding plans for this work were discussed. Many follow-on activities to the Dagstuhl seminar are now being planned, including: setting up a scientific journal as part of the Electronic Transactions on AI, submitting IST proposal on research projects and thematic networks, and organizing follow-up Transatlantic workshops on web-semantics efforts.

More information on current and future efforts can be found at `www.ontoknowledge.org`, `www.semanticweb.org`, and `www.daml.org`.

8 Validation of Dependable Behaviour of Safety- and Mission-Critical Computing Systems

Seminar No. **00131** Report No. **270** Date **26.03.–31.03.2000**
Organizers: Andrea Bondavalli, Mario Dal Cin, Klaus Echtle, Erik Maehle

The success of many applications strongly depends on the correct operation of computing systems – either because their malfunction would lead

to a violation of safety or their non-availability would cause the failure of an important mission and thus end up in severe financial losses.

To prevent all of these cases special countermeasures against faults must be applied. Depending on safety, availability and performance requirements on one side and cost limitations on the other the designs of computing systems may extremely differ from each other. In the well-known field of fault-tolerant computing a variety of solutions have been developed and applied successfully. A lot of experience has been gained about the operation of fault-tolerant computing systems.

However, there are still challenges in the design and the evaluation of dependable computing systems. This Dagstuhl seminar was particularly focused on the following problems:

- How can one prove or at least demonstrate that a design exhibits in fact the desired dependable behaviour, even in the presence of any of the specified faults?
- How can the design process be controled or at least influenced by appropriate evaluation methods, in order to ensure dependable behaviour in early design phases – in later design phases redesign towards dependability is more expensive.

Usual testing methods cannot be applied because they are not able to cover realistic faults which are too rare, compared to the usual durations of testing (with respect to testing, the scarcity of faults causes problems, not benefits!). Moreover, testing requires the complete implementation which is only available at the end of the design process.

A general problem of the validation of dependable computing systems lies in their complexity. Efficiency requirements typically disallow straight-forward solutions to the dependability problem. Instead, tricky approaches must be chosen, which deal with different fault types, locations and times of occurrence in different ways. How can we make sure that no faults "slip through" the complex net of countermeasures?

Besides some exceptional contributions the solutions presented in the seminar fall into the following categories (and some combinations thereof):

- Structuring the design process
- Modeling and respective tools

- Strict formal verification (using first order predicate logic)
- Partial verification using reachability analysis
- Inspection of a design
- So-called pessimistic fault injection
- Combination of functional analysis and quantitative evaluation
- Efficient techniques for model-based dependability evaluation

As was expected there is no superior concept that solves all dependability problems at a time. Most promising is the combination of design principles and various analysis methods.

In general software and system development the usage of (partly) formal specification techniques like UML or SDL grows. As one of the main results of the seminar, it was concluded that the dependability-related specifications or models need not be different from their nature. Instead they can use the same modeling language. It only needs to be enriched by dependability parameters and conditions.

9 Verification of Infinite-state Systems

Seminar No. **00141** Report No. **271** Date **02.04.–07.04.2000**
Organizers: Ahmed Bouajjani, Javier Esparza

The development of our modern societies needs more and more involvement of computers in managing highly complex and (safety-)critical tasks, e.g., in telecommunication, chemical and physical process control, transportation systems, etc. It is essential to be able to produce reliable hardware and software systems, since any erroneous behaviour can have catastrophic (economical and human) consequences. This requires rigorous methods and techniques to conceive, analyze and validate these systems.

The verification problem consists in checking whether a system satisfies its specification. During the two last decades, significant achievements have been obtained in the case of finite-state systems (systems with finitely many states). One of the main actual challenges in the domain of

automated verification is the conception of methods and tools allowing to deal with verification problems beyond the finite-state framework. Such problems rise naturally as soon as we consider aspects like:

- real-time constraints: timed and hybrid systems,
- unbounded discrete data structures: counters, fifo-channels, stacks, etc.
- parametric reasoning about families of systems, e.g., networks of processes,
- process mobility, dynamic creation and destruction of processes (dynamic modification of the communication structure).

In the last two years the specification and verification of infinite-state systems have attracted the attention of more and more researchers belonging to a very broad range of research communities. Both process algebras (or term rewriting systems) and automata (or finite control machines) are being used as specification formalisms. Verification problems can be reduced to checking behavioural equivalence or implementation (simulation) relations, or to checking the satisfaction of properties described in temporal logics or fixpoint calculi (model checking problems). Verification methods can be deductive (based on the use of theorem provers), or algorithmic (based on decision or semi-decision procedures). Algorithmic methods can be based on fixpoint theory, automata theory or (constrained) logic programming.

Recent work has shown that different techniques can be combined with sometimes spectacular results. As a result, it is being acknowledged that only a combination of techniques can lead to methods and tools widely used in practice. The aim of this meeting is to bring together researchers working on different research fields in order to make a synthesis on the state of the art and evaluate the potential of combined methods. Concrete questions to be addressed are:

- Which results in logic, automata theory, rewriting systems, etc. are applicable to automatic verification?
- How should deductive and algorithmic techniques be combined?
- Which are the right techniques to deal with abstraction?

- Which are the most promising application fields (mobile systems, cryptographic protocols, static program analysis ...), and which are the most appropriate models, specification formalisms, and verification techniques to deal with them?

10 Stochastic Modelling in Telecommunications

Seminar No. **00151** Report No. **272** Date **09.04.–14.04.2000**
Organizers: U. Herzog, G. Latouche, V. Ramaswami, P. Tran-Gia

Motivation text:
Telecommunication systems have a long tradition as a source and inspiration for stochastic models which are used at every stage, from conception to planning, realization and operation. Hot topics at the present time concern mobile telecommunications and developments in internet-based services and systems.

Internet-based services have created an enormous interest when they were first proposed, and are still the subject of intense scientific activity. During the workshop which was held this year, we have had several presentations in that area, and it is clear that the subject is not closed.

Mobile telecommunications are still in full development, both technological and methodological. No week goes by without a new announcement for an international conference devoted to the subject. This is spurred by the adoption of cellular telephones by users throughout the world, the deployment of IRIDIUM and future low earth orbit satellite systems, and the integration of fixed networks with wireless systems. These developments, in turn, create more research on mathematical models of spatial traffic, etc.

We expect the strength of the seminar to reside in the mix of participants: a sizeable proportion are employed in industrial research centers, and, among those from the academic world, several have close working relationship with industry. We are confident that this will lead to extremely interesting interactions among the participants, a varied set of points of views and a rich collection of problems addressed and models analyzed.

11 Probabilistic Methods in Verification

Seminar No. **00181** Report No. **273** Date **30.04–05.05.2000**
Organizers: Marta Kwiatkowska, Ulrich Herzog, Christoph Meinel, Moshe Vardi

The established methodology of engineering computer systems, both hardware and software, involves first building a *model* of the system, and then its detailed *analysis*, before implementation takes place. The motivation for this is to increase the engineers' confidence in the design of the system, with respect to desirable characteristics such as functional correctness and performance requirements. Two related disciplines which are instances of this methodology are *verification* and *performance evaluation*:

- The first of these, **verification** through *model checking*, employs algorithmic methods to provide "Yes/No" answers to qualitative correctness requirements, primarily concerned with *system behaviour over time* (for example, ensuring the delivery of a message or safety of a particular activity). A model of the system is formulated using an appropriate formalism, and then supplied as input to a software tool which automatically checks if a given specification is satisfied. Such model checking tools are used widely in practical applications, particularly for analysing hardware and communication protocols. The term **probabilistic verification** refers to methods in which "Yes/No" answers are replaced with estimates of the *likelihood* of the system satisfying a specification. Two prevailing views of probabilistic verification exist. The first concerns probabilistic models of the system (for example, discrete-time Markov chains or Markov decision processes), and aims to model check these against probabilistic variants of temporal logics. The second is applied in the context of non-probabilistic systems, but those of a size which makes exhaustive model checking impractical or infeasible. The aim is then to establish that the required properties hold with high probability.

- The field of **performance evaluation** involves building a probabilistic model of a system, followed by analysis focused on the calculation of *performance measures*. Typically, the underlying model

of system description formalisms in this field is a continuous-time Markov chain, with the desired system requirements (throughput, mean time to failure, etc.) expressed in terms of steady-state probabilities. This relies heavily on the use of numerical methods and tools when analysing the models. Performance evaluation tools have been successfully used to predict the impact of changes to load and arrival characteristics of computer networks.

Though the fields of verification and performance evaluation have historically concentrated on analysing different aspects of the system (*qualitative* correctness requirements versus *quantitative* performance issues), they are complementary and have much in common. For example, both aim to build a representation of the model in the computer memory, and in fact the difficulties and challenges posed by representing very large models have been recognised in both communities (the verification community calls this the 'state explosion problem', whereas to performance evaluation practitioners it is known as 'largeness'). Therefore, the appeal of integration and cross-fertilisation of techniques between the two fields is immediate. The goal of this Dagstuhl meeting was to bring together researchers representing the different communities, who would not necessarily meet at other conferences or workshops, in order to provoke debate and to facilitate exchange of expertise. In all, 48 researchers from 11 countries participated in the meeting.

In an effort to bring the two distinct, yet closely related, fields together four keynote speakers were invited to give overview lectures. **Luca de Alfaro** gave an introductory talk on the algorithmic verification of probabilistic systems. **Rajeev Alur** then spoke about modular specification and simulation of hybrid systems. An introduction to the field of performance evaluation was presented by **Boudewijn Haverkort**. This was complemented by a talk from **Martin Reiser** giving an overview of performance evaluation of computer and communication systems over the past thirty years.

The remaining 31 presentations given by participants of the meeting covered a range of topics from probabilistic verification and performance evaluation. Some centred on the development of **modelling formalisms** for probabilistic systems, including stochastic variants of process algebras and the π-calculus, real-time extensions to probabilistic and stochastic systems, and continuous space Markov processes.

Other talks concerned **methods of analysis** for such systems: model

checking algorithms for probabilistic and stochastic temporal logics; and equivalences on probabilistic systems and their corresponding decision procedures.

A third group of talks centred on the **implementation** of probabilistic verification, describing recent or ongoing work on the development of efficient tools and techniques. These included symbolic, BDD-based, model checking of probabilistic algorithms and stochastic Petri net tools employing Kronecker-based techniques. While the former focuses on verification and the latter on performance evaluation, what they have in common is the use of BDDs and Kronecker, which should lead to fruitful exchange of ideas. A group of talks introduced BDD-based methods for representing and verifying logical circuits with high probability.

A number of interesting **application areas** were also highlighted, including security and fault-tolerant systems.

The selection of presentations was accompanied by a **panel discussion** chaired by **Moshe Vardi** held towards the end of the meeting. Six prominent researchers (**Boudewijn Haverkort**, **Ulrich Herzog**, **Radha Jagadeesan**, **Joost-Pieter Katoen**, **Marta Kwiatkowska** and **Frits Vandraager**) were invited to answer questions on the present and future relationship between the fields of probabilistic verification and performance evaluation, prompting lively, interesting and productive discussion summarised at `http://www.cs.bham.ac.uk/~mzk/Dagstuhl/`. The optimism for future cooperation was evident not only here, but in the numerous stimulating discussions between participants during the week.

12 Quality of Service in Networks and Distributed Systems

Seminar No. **00191** Report No. **274** Date **07.05–10.05.2000**
Organizers: Andrew Campbell, Domenico Ferrari, Stefan Fischer, Lars Wolf

Distributed multimedia systems are becoming more and more important in many situations of our daily life, for instance in office applications (video conferencing), learning environments (tele-teaching and tele-learning, virtual universities), or entertainment (online games, video-on-

demand). Usually, some of the media types used in such an application have specific requirements on their transmission and presentation. The notion of Quality of Service (QoS) plays a central role when discussing about how to fulfil these requirements of multimedia applications. Distributed multimedia systems need QoS support in order to function properly. Moreover, other applications such as certain simulation systems need QoS functionality as well.

For this reason, research in QoS has increased significantly during the past few years. For an end-to-end QoS, which is in most applications necessary (user to user), support has to be provided in all components of the participating systems, i.e., the end system components, the communication system and the application. Accordingly, there has been active QoS research in network hardware (switches, routers), protocol software (RSVP, RTP etc.), operating systems (CPU scheduling), user interfaces, etc. Today, some of the basic technical issues are understood, but a significant amount of work is still necessary. Furthermore, additional research is devoted to (partially) non-technical issues such as pricing for QoS, but also new technical developments such as Active Networks.

This seminar concentrated on these new issues. In 30 talks and a number of panel discussions, the following topics were covered: QoS Architectures/QoS Management, Integrated and Differentiated Services, Multicast and Routing Issues for QoS, QoS in Mobile and Wireless Environments, Pricing and Accounting for QoS, QoS in Heterogeneous Networks, Active/Programmable Networks and QoS, QoS in Middleware, User Level QoS, and Adaptive Applications.

13 Scientific Visualization

Seminar No. **00211** Report No. **275** Date **21.05.–26.05.2000**
Organizers: G.-P. Bonneau, G. Nielson, F. Post

Motivation text:
This seminar will emphasize four areas of Scientific Visualization:

Volume Modeling:

The development of techniques for representing 2D surface models, which started in the early seventies and continues today, has been one of the

most influential successes in graphics and CAGD research. The results of this research have influenced how we design cars and toasters and the way we are entertained (for example, the movies Toy Story and Jurassic Park). Research is just beginning on extending surface model techniques to volume models. That is, the dimension of the independent data is increased from 2D to 3D. To date, there has been very little research on volume models, but the potential benefits are tremendous. Presently, most visualization techniques deal directly with the volume data and only produce an image. The image is, of course, only 2D and so nothing that is 3D is modeled or extracted form the data. Rather than apply techniques (visualization, queries, feature extraction) directly to the data, volume modeling is concerned with the development of techniques for producing a 3D volume model and then applying these techniques to the model. A volume model will allow for the generation of multiresolution models which are extremely useful for many of the operations commonly needed in a scientific investigation. The volume model also serves as a means of compressing the original volume data and this has benefits with regards to archiving and transmission of large volumetric data sets. Volume modeling is a very important emerging topic in scientific visualization which will be emphasized at this seminar.

Information Visualization:

An increasing number of automated processes and scanning devices are producing voluminous data that cannot be represented using conventional 3-D computer graphics techniques. These include single monitoring systems, LAN (local area networks) with attached scanning devices and, of course, the WWW. Research into new methods for the representation, archiving, perusal and understanding of this data is currently underway. The need for new methods for handling these abstract data sets is extremely widespread and the potential impact is tremendous. Two particularly fruitful areas are emerging which will be emphasized at this workshop. These are “smart spaces” and “visual interaction”. Wireless technology and mobile computing will soon allow users to move within a variety of distributed environments while declining prices of computing power will enable an increasing number of common objects to possess an intelligence heretofore reserved for single purpose computing devices. The confluence of these two trends is allowing researchers the opportunity to explore innovative ways of connecting mobile users and

computers with intelligent objects embedded within smart spaces. A smart space may be stationary, such as an office or conference room with attending aids or it may be mobile such as a vehicle with sensing and scanning devices. The possibilities and opportunities are very exciting and potentially rewarding. The new area of visual interaction is based upon recent advances in immersed visualization and gestural input and new algorithms for analyzing and processing data consisting of images (also 3D), video and audio signals. This is allowing the possibility of new and more effective paradigms for interaction between users and the simulated and real worlds contained within or accessible through computer systems.

Flow Visualization:

Flow and vector fields are common data types in many scientific applications. The need and benefits of understanding 3D flows is well accepted. Visualization techniques are providing assistance in this area. New and exciting research is occurring in flow visualization. This is due in part to the tremendous success of some recently developed and highly useful techniques which have revealed the importance of this research area. Emerging topics which will be included in this seminar include: multiresolution and wavelet methods for curvilinear grids, the identification, classification and representation of higher(than linear)order critical points and associated topological graphs for vector fields, extension to 3D of the surface based, highly successful LIC (line integral convolution), and real-time, interactive methods applied in virtual environments.

Volume Visualization:

The topic of volume visualization has been one of the main research topics of scientific visualization from the very beginning days of the late 1980's. The importance and benefits of this technology has been the driving force for a high level of quality research in this area. Volume visualization is rapidly maturing to a very pervasive technology which is used on an every-day basis in many application areas. Still it is an active research area as the frontiers are pushed further and further. This seminar will focus on Volume Graphics, which is the subfield of volume visualization that deals with modeling, manipulating, and rendering of 3D scenes that are represented by volumetric data. The emphasis in

Volume Graphics is on the ability to use volume data to achieve or even surpass many of the abilities traditionally supported by surface-based approaches. Recent trends in the industry and the academia predict that volume graphics is poised to leap forward in the near future. It is the goal of this symposium to reflect and promote this trend. The seminar will focus on algorithmic and complexity aspects arising in the development of efficient solution techniques for computationally difficult optimization problems. A general technique for efficient approximation algorithms is to formulate an optimization problem as an integer linear program and then to relax the integrality conditions. Recently, there has been striking success in obtaining also approximation algorithms on more general mathematical programming such as semidefinite programming. In this and other context, randomization has proved to be a powerful algorithmic method: It yields to simple and easy to analyze algorithms for many optimization problems, and it leads to a better performance guarantee. The seminar is intended to bring together researchers from different areas in combinatorial optimization and from applications. It will support the collaboration between researchers in computer science, mathematics and operations research. The workshop will have the following aims: extending of randomization and semidefinite programming techniques to other optimization problems, improved approximation algorithms and structural insights by studying linear programming, semidefinite programming and randomization, development of approaches to solve (approximatively) large linear and semidefinite programs, complexity questions for randomization and semidefinite programming, practical implementation of the used techniques (randomization, semidefinite programming), exchange of informations on recent research and stimulation of further research.

14 Topology in Computer Science: Constructivity; Asymmetry and Partiality; Digitization

Seminar No. **00231** Report No. **276** Date **04.06–09.06.2000**
Organizers: Ralph Kopperman, Mike Smyth, Dieter Spreen

Topological notions and methods have successfully been applied in var-

ious areas of computer science. The seminar concentrated on the following aspects: constructivity, asymmetry and partiality, and digitization. These key words not only describe a central theme of present-day research in computer science oriented topology, but also reflect the various features in which topological spaces used in computer science applications differ from those classically studied in mathematics.

• Constructivity

In great part, topological research in computer science originates from work on the formal description of programming language semantics and on automated program verification. Because of this connection constructive (effective) approaches to the theory of semantic domains and the relation of such domains to logic have been considered from the very beginning.

Nowadays this work is continued in research on Effective Topologies, Locale Theory and Formal Topologies. In studies on Formal Topology the theories of Scott domains and of metric spaces (especially the real numbers) are developed in the framework of higher-order formal intuitionistic logic.

• Asymmetry and Partiality

Both notions are characteristic for topological spaces used in computer science, as opposed to those considered in classical mathematics. Typically, a space considered in theory of computation is a space of partially defined objects which represent stages of some computation process. As a result of this, given two points, in general only one of them is separable from the other by an observable (or positive) property: whenever a (partial) object satisfies such a property, then also every more fully defined object must satisfy it. Thus a topology based on such properties can satisfy only very weak separation axioms. In the traditional spaces coming from applications in analysis or physics, however, two points can symmetrically be separated by a pair of disjoint properties. In other words, the spaces are at least Hausdorff.

In recent years, early ideas of Scott and Weihrauch/Schreiber to embed the real numbers or, more generally, metric spaces in suitable Scott domains formed by intervals or spheres, respectively, have been taken up

again. The program is to develop an important part of analysis starting with domains which contain not only the real numbers, but also their "unsharp" approximations. In this way one will come up with data structures and algorithms for numerical computations that are superior to those based on floating point representation and arithmetic.

Another line of research which is characterized by the above notions is the work on generalized metrics and quasi uniform spaces. By relaxing the classical requirements for a metric space categories of spaces are defined that contain Scott domains as well as metric spaces. Both types of spaces are used in giving meaning to certain programming language constructs. As has been shown this approach not only leads to a unified theory to be used in programming language semantics, but also allows the study of quantitative aspects of computations, which is impossible when dealing merely with Scott domains.

- Digitization

This draws attention to the digital nature of most computer applications. Starting from classical applications in physics, most topological notions have been developed by having the continuum in mind. In computer science applications they very often turn out to be no longer applicable in their classical form, or even to be meaningless, which means that they have to be redefined. Work done here is also basic to the other subfields of topology in computer science.

Obviously, all aspects mentioned above are strongly intertwined.

The aim of the workshop was to bring together computer scientists and those mathematicians who work on similar problems but from a different perspective and who, often, are not aware of the computer science motivations, and to create a common forum for the exchange of ideas and results. 56 top scientists and promising young researchers accepted the invitation to participate in the challenging experience. They came from 15 countries, mostly European countries and the USA, but also Canada, New Zealand, Russia, South Africa and Turkey. The 41 talks covered all of the areas mentioned above.

The atmosphere was very friendly, but the discussions were most lively. During the breaks and till late in the night, participants also gathered in smaller groups for continuing discussions, communicating new results and exchanging ideas. Some participants gave additional

evening lectures to discuss their results in more detail. Moreover, there was a lively discussion on the relationship between the (theoretical) results presented in many talks and applications in computer science.

The success of the workshop exceeded our expectation. The particpants expressed high appreciation of this gathering and praised the extraordinary Dagstuhl atmosphere which made this success possible.

15 Multimedia for Multimedia: Learning and Teaching in the next decade

Seminar No. **00241** Report No. **277** Date **11.06.–16.06.2000**
Organizers: Wolfgang Effelsberg, Eduard Fox, Nicolas Georganas, Ralf Steinmetz

Since the beginning of the nineties learning and teaching has tried to take advantage of multimedia and Internet technology and applications. Various research approaches have been used:

- Multimedia Courseware implies a focus on the learning and teaching material itself: How shall the contents be structured, accessed, and presented? Which media shall be used? How can this be adapted to the teaching/learning situation?
- Teleteaching/Telelearning as well as Cooperative Learning Environments refers to learning and teaching at different locations making use of networking facilities to communicate audio, video and other data. How can we support the learning in a team (possibly distributed teams with cooperative environments)? How can teams be supported in the learning process? How can synchronous and asynchronous usage modes be integrated?

A wide range of conferences and workshops in this area shows that these issues are topics of interest in many research communities. However many of the contributions have been rather superficial and not focused. Therefore, in this Dagstuhl seminar we took a narrowly defined focus – multimedia courseware and tools for learning multimedia – but approach it from a number of different perspectives.

The seminar brought together researchers from different areas interested in and working on "multimedia for multimedia": Experts in multimedia technology and applications, pedagogical and cognitive aspects, curricula on multimedia, lectures of multimedia and related publishers. Also new/up-coming leading and visionary researchers have been invited to demonstrate the state-of-the-art and work in progress.

The participants discussed emerging requirements, exchanged experiences in teaching and learning in this area (including a large number of demonstrations), identified commonalities all over the world, and discussed the potential of new multimedia technologies. According to the goal of the seminar, not only to hear presentations, the participants elaborated on common experiences in teaching multimedia.

The seminar included presentations, demonstrations and discussions on the following topics:

- Existing curricula on multimedia
- Design of physical and virtual environments for the learning of multimedia
- Potential and problems stemming from the sharing of high quality media elements
- Demonstrations of tools
- Potential of existing and upcoming document formats (e.g. virtual environments)
- Usability testing of multimedia courseware
- Quality of service considerations
- Asynchronous vs. synchronous learning
- Next generation Internet and its impact
- Educating the multimedia educators
- Social economic issues of "multimedia learning"

In the seminar different approaches to the usage of multimedia for teaching multimedia were identified. Based on a comparison of these

approaches the need for common multimedia curricula were discussed. Such curricula will allow the easy pooling and exchange of multimedia learning units and systems in order to improve the quality of multimedia training as a whole. As an outcome of the seminar the participants started to build up a worldwide network of experience to be used in the future by the whole research community interested in "multimedia for multimedia".

16 Image Synthesis and Interactive 3D Graphics

Seminar No. **00251** Report No. **278** Date **18.06.–23.06.2000**
Organizers: Michael Cohen, Heinrich Müller, Claude Puech, Hans-Peter Seidel

The goal of digital image synthesis and interactive 3D graphics is the calculation of images or image sequences from three-dimensional scenes. The task has many applications and challenging research problems which are particularly caused by the quest for rapid rendering from the requirements of applications on one side, and the inherently complex computational task on the other side. A typical problem is the development of approximating approaches which allow to calculate the optical effects sufficiently precise, but practically efficient. For the treatment of the problem, methods of finite element calculation, Monte-Carlo-Integration, design of efficient data structures and algorithms, but also human visual perception are applied. Another emphasis lies on the development of solutions on specialized hardware architectures and their use by adapted algorithms and software. In the last years, significant progress has been achieved, but by far not all important problems have been solved yet.

This Dagstuhl Seminar has provided a forum for leading researchers in this area to present their ideas. It has particularly benefited from the active participation of a high number of young researchers who have been for the first time at a Dagstuhl Seminar. Staying one week together at the beautiful place of Dagstuhl has stimulated the scientific and private exchange between the more than 50 participants from 11 countries far beyond that what can happen at usual conferences.

The number of participants and the willingness of active participation

by giving a presentation caused 47 talks. By keeping the presentation time short there has been the necessary time for discussions which has been extensively and intensively used by the attendees. Although the presentations covered an extremely broad spectrum, it was surprising to always see a well-filled auditorium. The impression was that the chance was used to learn more on fields aside from the often very specialized own research. Hot topics, like e.g. point-based rendering, have found particular interest.

The positive feedback of the participants has shown again that there is a need of events of the type of Dagstuhl Seminars, and we have the hope to be able to have a follow-up in the future.

17 Dynamically Reconfigurable Architectures

Seminar No. **00261** Report No. **279** Date **25.06.–30.06.2000**
Organizers: Gordon Brebner, Karl-Heinz Brenner, Hossam El Gindy, Hartmut Schmeck

The Dagstuhl Seminar on *Dynamically Reconfigurable Architectures* brought together 42 participants from 11 different countries. As with its predecessor seminar held in 1998, the participants came from three distinct communities, concerned with:

- field programmable gate arrays for fast and flexible configurable computing, and the associated design tools;
- computational models of, and designing efficient algorithms for, processor arrays with reconfigurable bus systems;
- optoelectronic systems for communication with very high bandwidth.

A main aim of the seminar was to encourage cross-fertilisation between these communities, to work towards a new understanding of dynamically reconfigurable architectures and their possibilities. Some of the specific questions addressed by the seminar were:

- What are shared characteristics of, and possible interfaces between, machine-based computational models and circuit-based computational models?

- How will current "ASIC replacement" device architectures evolve to be genuine "computational component" device architectures?
- What are appropriate system-level architectures for systems composed of processors, configurable logic and memory, and how might these be implemented at chip level?
- What overall computational models are best to evaluate alternative system-level architectures?
- Which aspects of current hardware and software design practice are, and are not, relevant to the design of "soft" circuitry and "hard" programs?
- What are appropriate new design processes, and supporting tools, for systems composed of a mixture of circuit and program, implemented in hard and soft manners?
- How general purpose can system design be? Is specialisation inevitable to achieve best performance?
- How are reconfigurable systems best presented to algorithm designers?
- What are realistic aspirations for the practical benefits achievable from the use of reconfigurable systems?
- How can optical communication be used best to support run-time reconfiguration?

There were 34 talks given by the participants, each shedding some light on one or more of the above issues. A lively discussion session on the Wednesday evening provided an opportunity to synthesise the differing viewpoints of the participants. Interestingly, one topic that provoked much discussion was how to define the term 'dynamically reconfigurable architecture', with differing positions emerging on how dynamic the reconfiguration should be expected to be, and on what types of architecture might be included.

The seminar emphasised that technological advances have opened up new ways of implementing complex systems, ways that blur the traditional barriers between hardware and software components. Because of this, existing design tools do not seem to be adequate for the necessary

new design styles - in fact, at one extreme, it is possible to let hardware evolve by itself, learning the required functions. There are many applications, ranging from computing on spacecraft to the control of motors, that should be able to derive benefits from enabling novel types of configurable computing systems.

A recurrent theme was whether the traditional Von Neumann model of computation has now run out of steam, as sequential processors become just one type of component in parallel and distributed systems. Irrespective of opinion on this, the need for robust computational models for run-time reconfigurable systems and evolvable hardware was highlighted, possibly following the general philosophy of the RMESH model for reconfigurable meshes. A thorough understanding of the computational possibilities is of great importance in assessing the benefits of dynamic reconfiguration for real-life applications.

Based on such an understanding, it will then be possible to make progress on finding apt high-level languages and notations for expressing computations based upon dynamically reconfigurable architectures. While mature concepts from both hardware and software engineering will influence this exercise, it seems that these must be augmented by natural high-level ways of expressing features such as concurrency, topology and reconfiguration.

Bridges then have to be built between the high level descriptions of function and the physical underlying architectures, which may be based on a variety of technologies, including optoelectronics. Robust computational models will also enable the development of apt virtual machines, that mask the particular technical details of specific reconfigurable computing machines.

The seminar pointed the way to how progress might be made in these directions, and new collaborations were initiated. The pleasant atmosphere of Schloß Dagstuhl was an important incentive for the lively interaction between the participants. We thank all who contributed to the success of the seminar. Furthermore, we gratefully acknowledge financial support for the seminar from Xilinx Inc.

18 Stochastic and Dynamic Real-Time Systems

Seminar No. **00271** Report No. **280** Date **02.07.–07.07.2000**
Organizers: Günter Hommel, Lonnie R. Welch

In many existing real-time computing models, the execution time of a "job" is used to characterize its workload. Typically, it is assumed that an *integer worst-case execution time* (WCET) is known *a priori*. This is not without justification, since static engineering approaches based on non-stochastic models have utility in many application domains [Sha91]. Furthermore, the pre-deployment guarantee afforded by such approaches is highly desirable. However, there are numerous applications which must operate in dynamic environments, thereby precluding *accurate* characterization of the applications' properties by static models which are non-stochastic. Some real-time systems operate in environments which can be characterized *a priori* by a statistical distribution. Other control systems operate in environments which can not be modeled accurately with a time-invariant distribution; their time-variant stochastic characterizations must be repeatedly derived *a posteriori*.

A growing number of researchers in the field of real-time systems are aware of those problems. On the other side there are researchers in the field of stochastic modeling who are interested in modeling and analyzing non-Markovian stochastic systems including their partially deterministic behavior. The goal of this Dagstuhl-Seminar is now to bring together researchers of both fields in order to consider engineering approaches for real-time systems which cannot be characterized accurately by non-stochastic *a priori* models.

In typical real-time computing models (e.g., see [Liu73, Ram89, Xu90, Sha91, Bak91]), execution time is assumed to be an a priori integer "worst-case" execution time (WCET). While [Sha91] establishes the utility of a priori WCET-based approaches by listing some domains of successful application, others [Leh96, Jah95, Hab90, Kuo97, Sun96, Ram89, Tia95, Str97, Ste97, Liu91, Abe98, Atl98, Bra98] cite the drawbacks, and in some cases the inapplicability, of the approaches in certain domains. [Ram89, Tia95, Leh96, Hab90, Abe98] indicate that characterizing workloads of real-time systems using a priori worst-case execution times can lead to poor resource utilization, particularly when the difference be-

tween WCET and normal execution time is large. It is stated in [Ste97, Abe98] that accurately measuring WCET is often difficult and sometimes impossible. In response to such difficulties, techniques for detection and handling of deadline violations have been developed [Jah95, Str97, Ste97].

Recently, paradigms which generalize the execution time model have emerged. Execution time is modeled as a set of discrete values in [Kuo97], as an interval in [Sun96], and as a time-invariant probability distribution in [Leh96, Str97, Tia95, Atl98]. These approaches assume that the execution characteristics (set, interval or distribution) are known a priori.

Others have taken a hybrid approach; for example, in [Hab90] a priori worst case execution times are used to perform scheduling, and a hardware monitor is used to measure a posteriori task execution times for improving hardware utilization via dynamic adaptation. [Liu91, Str97] view jobs as consisting of mandatory and optional portions, with one of these having characteristics that can not be known a priori. In [Liu91] the mandatory portion has an a priori known execution time, while the optional portion has an unknown execution time. In [Str97], the optional portion is used for handling timing violations of the mandatory portion and thus has an a priori known execution time. In [Bra98, Wel98] resource requirements are observed a posteriori, allowing applications which have not been characterized a priori to be accommodated. Also, for those applications with a priori characterizations, the observations are used to refine the a priori estimates. These characterizations are then used to drive dynamic resource allocation algorithms.

Engineering approaches for stochastic and dynamic real-time systems have the potential to extend the applicability of real-time computing research into new domains of use. Thus, we propose to focus on advancing the modeling and analysis techniques for such systems.

Abe98 L. Abeni and G. Buttazzo, "Integrating multimedia applications in hard real-time systems," in Proceedings of the 19th IEEE Real-Time Systems Symposium, 3-13, IEEE Computer Society Press, 1998.

Atl98 A. Atlas and A. Bestavros, "Statistical rate monotonic scheduling," in Proceedings of the 19th IEEE Real-Time Systems Symposium, 123-132, IEEE Computer Society Press, 1998.

Bak91 T.P. Baker, "Stack-based scheduling of realtime processes," Journal of Real-time Systems, 3(1), March 1991, 67-99.

Bra98 S. Brandt, G. Nutt, T. Berk and J. Mankovich, "A dynamic quality of service middleware agent for mediating application resource usage," in Proceedings of the 19th IEEE Real-Time Sys. Symposium, 307-317, IEEE Computer Society Press, 1998.

Hab90 D. Haban and K.G. Shin, "Applications of real-time monitoring for scheduling tasks with random execution times ," IEEE Transactions on Software Engineering, 16(12), December 1990, 1374-1389.

Jah95 F. Jahanian, "Run-time monitoring of real-time systems," in Advances in Real-time Systems, Prentice-Hall, 1995, 435-460, edited by S.H. Son.

Kuo97 T.E. Kuo and A. K. Mok, "Incremental reconfiguration and load adjustment in adaptive real-time systems," IEEE Transactions on Computers, 46(12), December 1997, 1313-1324.

Leh96 J. Lehoczky, "Real-time queuing theory," in Proceedings of the 17th IEEE Real-Time Systems Symposium, 186-195, IEEE Computer Society Press, 1996.

Liu73 C.L. Liu and J.W. Layland, "Scheduling algorithms for multiprogramming in a hard-real-time environment," Journal of the ACM, 20, 1973, 46-61.

Liu91 J.W.S. Liu, K.J. Lin, W.K. Shih, A.C. Yu, J.Y. Chung and W. Zhao, "Algorithms for scheduling imprecise computations," IEEE Computer, 24(5), May 1991, 129-139.

Ram89 K. Ramamritham, J.A. Stankovic and W. Zhao, "Distributed scheduling of tasks with deadlines and resource requirements," IEEE Transactions on Computers, 38(8), August 1989, 110-123.

Sha91 L. Sha, M. H. Klein, and J.B. Goodenough, "Rate monotonic analysis for real-time systems," in Scheduling and Resource Management, Kluwer, 1991, 129-156, edited by A. M. van Tilborg and G. M. Koob.

Ste97 D.B. Stewart and P.K. Khosla, "Mechanisms for detecting and handling timing errors," Communications of the ACM, 40(1), January 1997,87-93.

Str97 H. Streich and M. Gergeleit, "On the design of a dynamic distributed real-time environment," in Proceedings of the 5th Inter-

national Workshop on Parallel and Distributed Real-Time Systems, 251-256, IEEE Computer Society Press, 1997.

Sun96 J. Sun and J.W.S. Liu, "Bounding completion times of jobs with arbitrary release times and variable execution times," in Proceedings of the 17th IEEE Real-Time Systems Symposium, 2-11, IEEE Computer Society Press, 1996.

Tia95 T.S. Tia, Z. Deng, M. Shankar, M. Storch, J. Sun, L.C. Wu and J.W.S. Liu, "Probabilistic performance guarantee for real-time tasks with varying computation times," in Proceedings of the 1st IEEE Real-Time Technology and Applications Symposium, 164-173, IEEE Computer Society Press, 1995.

Wel98 L. R. Welch, B. Ravindran, B. Shirazi and C. Bruggeman, "Specification and analysis of dynamic, distributed real-time systems," in Proceedings of the 19th IEEE Real-Time Systems Symposium, 72-81, IEEE Computer Society Press, 1998.

Xu90 J. Xu and D.L. Parnas, "Scheduling processes with release times, deadlines, precedence and exclusion relations," IEEE Transactions on Software Engineering, 16(3), March 1990, 360-369.

19 Knowledge Management: An Interdisciplinary Approach

Seminar No. **00281** Report No. **281** Date **09.07.–14.07.2000**
Organizers: M. Jarke, D. E. O'Leary, R. Studer

Motivation text:
Knowledge Management (KM) can be a critical success factor for enterprises. Lean management structures rely on decision making on all organizational levels of an enterprise so that knowledge must be available and managed at all levels of an organization. Downsizing companies must capture the knowledge of the employees that they lose or their knowledge is forever lost. Short product development schedules require the tight integration and management of knowledge of different departments in an enterprise.

KM is inherently an interdisciplinary issue involving aspects of human resource management, enterprise organization and supporting and

enabling information technology (IT) methods and techniques. KM can result in an improvement of the enterprise competitiveness only if these technologies are integrated and fully leveraged.

The proposed Dagstuhl seminar aims at bringing together people from different disciplines, i.e. Information Systems, Management Sciences and Computer Science, to address these issues. A majority of persons will be invited from the Computer Science area, especially from areas like Business Process Modeling, Intelligent Information Systems, Knowledge Engineering, Data Mining, Case-based Reasoning, Information Retrieval, Computer Supported Cooperative Work, and Reference Models for Enterprises.

Among others, the following topics will be discussed at this seminar:

- What are knowledge-intensive business processes that may be supported and improved by a KM approach?

  This aspect includes an analysis of the information flow between different subtasks of a process as well as an analysis of the required kind of knowledge to perform these subtasks.

- What kind of knowledge has to be stored in an Organizational Memory?

  This aspect addresses the diversity of knowledge being stored in an Organizational Memory, like e.g. Best Practices, project experiences, competence descriptions of employees, or formal knowledge to solve a specific task. Relevant in this context are also approaches to structure knowledge, e.g. by ontologies, and to provide meta-level descriptions of the contents of the Organizational Memory. On the other hand, one also has to consider what basic technologies are most appropriate for implementing Organizational Memories.

- How is the knowledge made available to the employees?

  This includes pull and push techniques as well as intelligent filtering techniques. Human factors are also relevant in this context.

- What are the best and most economical approaches for generating and keeping the contents of the Organizational Memory up-to-date?

  An important issue for the acceptance of a KM system is to provide mechanisms for getting rid of outdated information, for systematically updating the contents and for evaluating the quality of the

stored knowledge. Knowledge Engineering as well as data and text mining approaches provide relevant methods for handling some of the involved problems.

- What kind of knowledge is provided by reference models and how can these reference models be adapted to the specific enterprise environment?

  Reference models come with a large collection of different types of knowledge, among others domain concepts or standardized business processes. A careful analysis of the provided knowledge would also provide a basis for customizing reference models as part of an organizational learning process.

- How can the cooperation and coordination between different employees or different departments be enhanced by supporting IT techniques?

  This aspect includes the problem of how to make available individual expertise to a group of people and of how to decide which knowledge should be shared among people. These problems are especially relevant in geographically distributed enterprises that do not easily allow the creation of people networks. The usage of CSCW as well as of intra and internet technologies for connecting employees to the Organizational Memory are of importance in this context.

- What enterprise organization and culture is needed to make KM feasible?

  Issues to be considered in this context are e.g. organizational structures which support knowledge sharing, incentive systems as part of a human resource management strategy, or empowerment of employees on all management levels.

- What are approaches to evaluate the (economic) value of a KM strategy?

  Building up a KM system requires a lot of resources. A critical aspect from a management point of view is the question of how to determine the economic value of a KM approach, e.g. with respect to enhanced product quality, shorter product development times or improved customer satisfaction.

The goal of this seminar is to bring together people from universities and industry in order improve the exchange of ideas and experiences between scientists and practical persons working in a variety of areas which are relevant for Knowledge Management.

20 Logics for Emerging Applications of Databases

Seminar No. **00291** Report No. **282** Date **16.07.–21.07.2000**
Organizers: J. Chomicki, R. van der Meyden, G. Saake

Motivation text:
The designers and users of present-day information systems deal with more and more complex applications that have to meet stringent quality requirements. Logic, in the form of many different logical formalisms, is a suitable tool to address some of the emerging problems, particularly those that benefit from a declarative formulation. The advantages of logic include formal, well-understood syntax and semantics, and thoroughly studied computational properties.

Emerging applications requiring database management have motivated new developments in logical frameworks for their support. Examples of these new developments are:

- logical frameworks for various aspects of electronic commerce, e.g. negotiation and communication, EDI, electronic contracts
- logics for security, for example in distributed data stores
- workflow logics as descriptive frameworks for specifying, querying, animating and modifying workflows
- logical frameworks for handling data of varying quality in data warehouses ("soft" integrity constraints, ..)
- logics for information integration, database interoperation, and virtual enterprises
- logical frameworks for the specification of multimedia applications

- logics for temporal, spatial, and spatio-temporal database applications
- logical frameworks in scientific data handling
- logical specification of agents and policies
- logical aspects of knowledge discovery in databases and OLAP.

In addition to addressing the above new developments, the seminar will focus on the critical evaluation of existing logical frameworks for databases and information systems in the light of past experience and future potential. In particular, the following are candidates for such a review: nonmonotonic logics, logics of time and action, deontic logic, deductive and constraint databases. Leading researchers from emerging database applications areas are also being invited to comment on the role of logic in their areas, hopefully generating lively discussions.

In 1995, the seminar 9529 about "The role of logics in information systems" took place in Dagstuhl. The seminar was successful in bringing together researchers from different areas connecting formal logic with database problems. The seminar discussions lead to the book "Logics in databases and information systems" published by Kluwer in 1998. Five years after this workshop, new application areas will have emerged leading to new challenges for combining logical frameworks with database technology. The Dagstuhl seminar 00291 is dedicated to logics for emerging applications of databases, continuing the interdisciplinary work started at the first seminar.

21 Intelligent Data Analysis

Seminar No. **00331** Report No. **283** Date **13.08.–18.08.2000**
Organizers: Michael R. Berthold, Rudolf Kruse, Xiaohui Liu, Helena Szczerbicka

Introduction

For the last decade or so, the size of machine-readable data sets has increased dramatically and the problem of "data explosion" has become apparent. On the other hand, recent developments in computing have

provided the basic infrastructure for fast access to vast amounts of online data and many of the advanced computational methods for extracting information from large quantities of data are beginning to mature. These developments have created a new range of problems and challenges for the analysts, as well as new opportunities for intelligent systems in data analysis and have led to the emergence of the field of Intelligent Data Analysis (IDA), a combination of diverse disciplines including Artificial Intelligence and Statistics in particular. These fields often complement each other: many statistical methods, particularly those for large data sets, rely on computation, but brute computing power is no substitute for statistical knowledge.

Although many interesting systems and applications have been developed in the field, much more needs to be done. For instance, most of the data collected so far have not been analyzed, and there are few tools around which allow the effective analysis of "big data". Different analysis strategies may be applied to the same problem and it is often difficult to judge which is the most appropriate; and the division of the work between the computer and the analyst most effectively is by and large still work of art. Many different application areas offer vast potential for improvement, examples include such diverse fields as the automatic monitoring of patients in medicine (which requires an understanding of the underlying decision process), optimization of industrial processes, and also the extraction of expert knowledge from observations of their behavior.

To discuss these and many other related topics, the First International Symposium in Intelligent Data Analysis (IDA-95) was held in Baden-Baden, followed by the second IDA symposium (IDA-97) in London and the third (IDA-99) held in Amsterdam. All three symposia were successful in attracting a large number of people with different backgrounds including AI, pattern recognition and statistics as well as people from diverse application fields. However, these large symposiums tend to lack of close interaction and discussion among people, which are relatively easy with a limited number of participants. Therefore the goal of this Dagstuhl seminar was to bring together a small number of experts from the various disciplines to discuss important issues in Intelligent Data Analysis, review current progress in the field, and identify those challenging and fruitful areas for further research.

Scope

The goal of this Dagstuhl Seminar was to focus on some key issues in intelligent data analysis that are directly relevant to the above aspects, both from the application and theoretical side:

- Strategies: Data analysis in a problem-solving context is typically an iterative process involving problem formulation, model building, and interpretation of the results. The question of how data analysis may be carried out effectively should lead us to having a close look not only at those individual components in the data analysis process, but also at the process as a whole, asking what would constitute a sensible analysis strategy. This strategy would describe the steps, decisions and actions which are taken during the process of analyzing data to build a model or answer a question. The work reported so far in this area has very much been restricted to relatively small problems. It is important to further our understanding of strategic issues in the context of large data-analysis tasks.

- Integration: In addition to careful thinking at every stage of an analysis process and intelligent application of relevant domain expertise regarding both data and subject matters, Intelligent Data Analysis requires critical assessment and selection of relevant analysis approaches. This often means a sensible integration of various techniques stemming from different disciplines, given that certain techniques from one field could improve a method from another one.

- Data Quality: Data are now viewed as a key organizational resource and the use of high-quality data for decision making has received increasing attention. It is commonly accepted that one of the most difficult and costly tasks in modern data analysis is trying to obtain clean and reliable data. There are so many things that can be wrong with real-world data: they can be noisy, incomplete and inconsistent, and it is not always easy to handle these problems. Since the use of "wrong" kind of data or very "low-quality" data often leads to useless analysis results, research on data quality has attracted a significant amount of attention from different communities including information systems, management, computing, and statistics. Much progress has been made, but further work

is urgently needed to come up with practical and effective methods for managing different kinds of data quality problems in large databases.

- Scalability: One of the key issues involved in large-scale data analysis is "scalability", e.g. if a method works well for a task involving a dozen variables, is it still going to perform well for one with over 100 or 1000 variables? The "curse of dimensionality" is still with us, and there has been a lot of intensive research on the development of efficient, approximate or heuristic algorithms that are able to scale well. For example, feature selection is one of the principal approaches to the scaling problem and a considerable amount of progress has been made in the field. However, many challenging issues still remain: How to develop a scalable algorithm when the data involved are highly-structured? And what would be the most effective way of sampling the most representative cases from a huge database for model-building? Currently, technical reports of analyzing "big data" are still sketchy. Analysis of big, opportunistic data (data collected for an unrelated purpose) is beset with many statistical pitfalls. We need to accumulate much more practical experience in analyzing large, complex real-world data sets in order to obtain a deep understanding of the IDA process.

Structure

Due to the interdisciplinary nature of the audience the real challenge of this meeting was the initiation of interactions across different disciplines. In order to provide the required background, one tutorial-style introduction was scheduled for each day, aiming to familiarize researchers with concepts from the various fields. These five surveys focused on some of the most important aspects of IDA:

- Generation, Representation, and Manipulation of Perceptions
- Association Rules
- Combining Classifiers
- Handling Missing Data
- Developing Mining Tools

The remainder of each day was used for presentations from the other attendees. On Friday we were fortunate to have Nicholas Deliyanakis from the European Commisson join the seminar and present the new directions in funding under the 5^{th} Framework Programme.

Results

Results of this seminar are a better, deeper understanding of issues and methodologies in intelligent data analysis as well as insights into relevant real world problems. We hope this will lead to new interactions between disciplines, also initiated through insights into applications and the quite often very different emphasis they put on results and handling of the various techniques. Hopefully the proposed seminar has started new collaborations between people with different backgrounds, and has provided new insights regarding what is, or what is not, an intelligent way of analyzing data.

22 High Performance Computing and Java

Seminar No. **00341** Report No. **284** Date **20.–25.08.2000**
Organizers: Susan Flynn-Hummel, Vladimir Getov, François Irigoin, Christian Lengauer

The object-oriented programming language Java is being viewed as the modern alternative for C++ and has rapidly captured people's attention, largely because of its features for interactive and Internet programming. One advantage of Java over C++ is that it includes mechanisms for parallelism and coordination, which makes it a natural language for distributed computing.

Java is still commonly perceived as execution-inefficient. It is not being realized widely enough that this inefficiency is a property of the language *implementation*, not of the language *per se*. Initial implementations interpreted relatively unoptimized bytecode via a relatively unsophisticated Java Virtual Machine (JVM). Recent developments in compilation technology –for instance, increased static analysis and just-in-time (JIT) compilation– and extensions of the JVM with instruction-level optimizations have done away with many of the initial sources of Java's

execution inefficiency. In certain applications, Java is these days competitive with C or C++, but a lot safer to execute and easier to program.

In the face of these developments, interest in Java has grown also in the field of high performance computing. Obviously, high performance applications place extreme demands on execution efficiency.

The first question which is immediately being asked when Java is being proposed as a vehicle for high performance computing is: why?

The most frequent answer is: to get access to the incredible amount of resources that goes into program development in Java (this includes class libraries, compilation technology, programmer education, etc.). Without any other answer, one would mimic high performance programming in Fortran, C or C++ when programming in Java, and one would expect the Java compiler to produce essentially the target code which a Fortran, C or C++ compiler would produce. This approach is actually being pursued in some research and is legitimate as an immediate aid to a new generation of programmers, who make Java their preferred choice.

An answer that is much more ambitious to implement but that does Java more justice is: to develop a new technology in high performance computing that allows us to do things that cannot easily be done with Fortran, C or C++. For example, one distinct advantage of Java over Fortran, C, and C++ is its support of secure, portable and dynamic target code. This makes it a promising vehicle for irregular applications, such as those with thread-based parallelism, or for run-time compilation and optimization – if Java can deliver the expected execution efficiency.

So, the next question is: can Java deliver? The immediate answer seems to be: in its present form to a limited extent. Many obstacles to high performance are due to current implementations of the language. However, some are part of its design and others are imposed by the unusually restrictive semantics of Java.

How can Java be made more suitable for high performance computing? Should high performance applications be adapted to the present limitations of the language, or should the language be extended (moderately) to become better suited? If the former, how can we tune Java implementations to exploit the performance potential of the language to the fullest? If the latter, what form should these extensions take? Is it permitted to add/modify language features or can/should one stick to class library extensions?

These and similar questions were explored by 35 researchers at Schloß Dagstuhl. The abstracts in the report summarize the presentations. Three full papers of participants in the October 2001 issue of the *Communications of the ACM* provide an overview of the current state of the art in getting high performance from Java.

The outcome of the seminar has been: Java has a lot more potential for high-performance computing than is commonly believed. Yes, if implemented naïvely, there are problems with scientific programming – no, they are largely not inherent but can be overcome. Yes, exploiting the full potential of Java's parallelism requires efficient virtual machines and high-performance networking software – which is in the process of being developed.

We are grateful to the IBM Thomas J. Watson Research Center for financial support for the seminar.

23 Experimental Algorithmics

Seminar No. **00371** Report No. **285** Date **10.09–15.09.2000**
Organizers: Rudolf Fleischer, Bernard Moret, Erik Meineche Schmidt

In recent years, many areas of theoretical computer science have experienced a shift to more applied research. Clear evidence of this fact is the surge of experimental papers in areas which used to be completely satisfied with a thorough theoretical analysis of the problems and algorithms. One reason for this is that people have learned that a purely theoretical analysis of an algorithm is often insufficient for practical purposes because

- Many practical problems belong to "difficult" problem classes (NP-complete problems or worse), so asymptotic worst-case bounds do not give a satisfactory answer to the question whether an algorithm is "useful" or not.
- More often than not, practical algorithms build on various heuristics where no tight bounds even exist.

So people have started to implement and test their algorithms besides of doing the usual theoretical analysis. Unfortunately, it is often not clear what these experiments actually tell us. Is an algorithm good just because it seems to behave well on random instances or on some benchmark test set? There is no sound basis for experimental algorithmics which could give us answers to questions like

- What are relevant experiments?
- What can be learned from experiments?
- What is a good benchmark test set?

and finally

- What is a good experimental paper?

The aim of this seminar was to bring together two groups, theoreticians and practitioners, to discuss these problems and start establishing a methodology of experimental algorithmics. In all, 45 researchers with affiliations in Austria, Canada, Denmark, France, Germany, Great Britain, Greece, Hong Kong, Italy, the Netherlands, Spain, and the USA participated in the meeting. Three invited keynote speakers, Jon Bentley, David S. Johnson, and Kurt Mehlhorn, gave one-hour position talks. The remaining 26 presentations given by participants of the meeting covered a wide range of topics in experimental algorithmics. The abstracts of most of these presentations are contained in the seminar report. Two evenings were reserved for discussions on specific topics, a summary of the outcome of the discussions is included at the end of the report.

24 Code Optimisation: Trends, Challenges, and Perspectives

Seminar No. **00381** Report No. **286** Date **17.09.–22.09.2000**
Organizers: Carole Dulong, Rajiv Gupta, Robert Kennedy, Jens Knoop, Jim Pierce

From September 17 to 22, 2000, the Dagstuhl Seminar 00381 on "Code Optimisation: Trends, Challenges, and Perspectives" took place in Schloß

Dagstuhl covering research being conducted on a broad range of topics related to code optimization. A major goal of this seminar was to bring together researchers and practitioners from industry and academia working in the field of optimizing compilation, and to create a setting for interaction among the attendees leading to exchange of ideas and future collaborations. This goal was clearly met. Attendees from three continents, 24 from Europe, 5 from Asia, and 11 from North America, 12 of which were from industry and 28 from academia, represented a broad crossection of the community. This was also reflected by the broad range of topics covered at the meeting including:

- Classical code optimizations (PRE, inlining etc.)
- Instruction scheduling and register allocation
- Mobile code - representations and optimizations
- Profiling and profile guided optimization
- Memory optimizations
- Code generation/optimization for DSP/Network/Embedded processors
- Optimizations for Java
- Analysis techniques (data flow, symbolic etc.)

The scientific programme of this seminar consisted of 32 contributed talks, plenary discussion sessions, and informal meetings in the evenings. Predominating topics in the plenary discussion sessions were the impact of compiler optimization on performance, the relationship between code optimization and code verification, and future challenges for code optimization. The questions below summarize some of the key issues discussed here:

- Compiler Optimizations for Performance:
  - The impact of many optimizations is small.
  - How do the advances in performance from hardware techniques compare with those achieved through compiler techniques?

  - Have we not paid sufficient attention to optimizations with big payoff (e.g., memory optimizations)?

- Code Optimization and Code Verification:

  - Systems being deployed must function correctly or consequences can be disastrous. Do the goals of optimization and verification conflict?

An immediate outcome of this workshop is an *International Workshop on "Compiler Optimization Meets Compiler Verification (COCV 2002)"* going to be organized as a satellite event of the *5th European Joint Conferences on Theory and Practice of Software* (*ETAPS 2002*), Grenoble, France, April 6 - 14, 2002. This workshop will take place on April 13, 2002, in Grenoble, France. More information on this event can be found on the home page of COCV 2002 at

`http://sunshine.cs.uni-dortmund.de/~knoop/cocv02.html`

Further information including contact information of participants, slides of presentations, as well as brief position statements of future challenges in code optimization can be found at the Seminar's home page:

`http://sunshine.cs.uni-dortmund.de/~knoop/`
`dagstuhl_00381/dag_00381.html`

Motivation of the Seminar

The last decades have been witness to continuous, rapid, and far reaching progress in code optimisation. Currently, optimisation faces new challenges caused by the increasing importance of advanced programming paradigms like the object-oriented, (data-) parallel and distributed ones, the emerging dissemination of new innovative processor architectures, and the explosive proliferation of new application scenarios like Web-computing in the retinue of the meanwhile ubiquitious Internet.

New paradigms, new architectures, and new application scenarios apparently demand new compilation and optimisation techniques, but also offer new potentials for optimisation on both machine-dependent and machine-independent levels.

In the light of this situation the aim of the seminar is to bring together researchers and practitioners from both industry and academia working on any phase of optimising compilation to exchange views, share experiences, identify potentials and current and future challenges, and thus, to bridge gaps and stimulate synergies between theory and practice and between diverse areas of optimisation such as machine-dependent and machine-independent ones.

Central issues which should be discussed in the seminar are:

- *Paradigm and software/hardware boundaries*:
  Do we require new techniques to reasonably accommodate the specialities of new paradigms, architectures, or Web-driven application scenarios? Will unifying approaches that transcend paradigms be superior or even indispensible because of the economic demands for reusability, portability, and automatic generability? Similarly, the boundaries between hardware and software optimisations are changing and redefined as *e.g.*, by IA64. Does the boundary lie in the right place? What are the missing architecture hooks for the compiler to really be as good as the hardware? Is that even possible?

- *Optimisation of running time vs. memory use*:
  Will there be a renascence of storage-saving optimisations and a shift away from emphasis on running time due to the growing importance of embedded systems and the distribution of executables across the Internet?

- *Static vs. dynamic and profile-guided optimisation*:
  Very wide architectures are very sensitive to profile-guided optimisations. The profile data set, however, is most commonly not known at compile time. What are practical ways of gathering profile information without slowing down applications, and practical ways of using this information for dynamic optimisations? Concerning Internet-based applications, must approaches for just-in-time computation be complemented by approaches for just-in-time optimisation? What are the key issues here?

- *Formal methods*:
  What is the role of formal methods in code optimisation regarding the requirements of reliability, validation or even verifiability

of correctness and (formal) optimality of an optimisation? What should be its role? How can the benefits possibly offered by formal methods best be combined with those of empirical evaluations?

- *Mastering complexity*:
  The increased complexity of compiler optimisation can lead to validation nightmares and can increase compiler team size to a counterproductive level. This phenomenon proves to be a key problem in practice. How can it be mastered? In particular, how do we avoid the problem faced by growing software and hardware teams? Can formal methods improve on this situation? What could be their impact?

- *Experimental evaluations*:
  Do we need a common, publicly available compiler testbed for experimental evaluations and comparisons of competing approaches? What would be the key requirements? Is it indispensable for reasonably pushing synergies between theory and practice?

- *Synergies*:
  What and how can people from different communities working on code optimisation, e.g. on a machine-dependent and machine-independent level, learn from each other?

At the threshold of a new millenium, and in face of the rapid change of paradigms on both the software and hardware sides, it seems to be worthwile to take stock of the state of the art, to reflect on recent trends, and to identify current and future challenges and perspectives on code optimisation.

25 Algorithms and Complexity for Continuous Problems

Seminar No. **00391** Report No. **287** Date **24.09.–29.09.2000**
Organizers: S. Heinrich, S. Pereverzev, J. Traub, G. Wasilkowski

The seminar was attended by 36 scientists from 12 countries. It was devoted to the complexity of a range of important problems and efficient

algorithms for solving them. Among those problems were numerical integration and approximation, solving integral and operator equations, solving partial differential and stochastic differential equations, optimization. Since the complexity of problems depends on the dimensionality of the underlying class of functions, over 20 talks (out of the total of 32) discussed multivariate problems and the dependence of the complexity on the number d of variables.

Integration of functions of very many variables is of great practical interest in a wide range of applications including statistical mechanics, chemistry, physics and mathematical finance. Although it has been studied for a long time, the classical numerical results provide either very negative results of *exponential* dependence on d or insufficient information of the dependence on d. Since in the problems mentioned above, the value of d could easily be in thousands, millions, or sometimes equal to infinity, the dependence on the dimension is of great importance. Sometimes, it is even more important than the dependence on the error demand. Not surprisingly, 14 talks presented new results on high-dimensional integration and we will concentrate on some of them here. There was a number of results dealing with construction and/or properties of low discrepancy points as well as of (Quasi-) Monte Carlo methods. The model of quantum computation and its power for numerical integration was presented. Applications to specific finance problems were discussed, as well as complexity and optimal methods for integrating singular functions or functions over unbounded domains. Monte Carlo and/or Quasi-Monte Carlo methods were also discussed in the context of other than integration problems including differential equations and computer vision. In particular, an optimal randomized method was proposed to solve integral equations with singular kernel.

In addition to presentations, one afternoon was devoted to the *Open Problems Session* where a series of new important research problems has been proposed.

26 Semi-Formal and Formal Specification Techniques for Software Systems

Seminar No. **00411** Report No. **288** Date **08.10.–13.10.2000**
Organizers: Hartmut Ehrig, Gregor Engels, Fernando Orejas, Martin Wirsing

During the last 20 years several different formal and semi-formal specification techniques have been successfully developed and used. Applications comprise the specification of simple programs, data types and processes as well as of complex hardware and software systems. The variety of specification techniques ranges from formal set theoretical, algebraic and logic approaches for specifying sequential systems and from Petri-nets, process algebras, automata, graph grammars for specifying concurrent and distributed behaviors to semi-formal software engineering methods such as UML which has become the de facto software engineering standard for developing large and complex systems.

Formal and semi-formal approaches have their advantages and disadvantages: the informal diagrammatic methods are easier to understand and to apply but they can be ambiguous. Due to the different nature of the employed diagrams and descriptions it is often difficult to get a comprehensive view of all functional and dynamic properties. On the other hand, the formal approaches are more difficult to learn and require mathematical training. But they provide mathematical rigor for analysis and prototyping of designs. Verification is possible only with formal techniques.

The talks and discussions of a previous seminar held in Dagstuhl in 1998 on these topics have shown that many researchers and research groups are putting more and more effort in closing this gap by integrating semi-formal and formal specification techniques. Their studies and experiences show the added value of combining semi-formal and formal techniques and at the same time open a whole range of new problems and questions which cannot be asked when studying formalisms in isolation. The talks and discussions of the previous seminar have shown particular interest and first results in studying the relationship between UML as de facto standard for diagrammatic modeling languages and CASL as the new standard algebraic specification language.

The goal of this seminar was to study possibilities and solutions for

integrating and validating different formal and semi-formal specification techniques. More than 50 participants attended the seminar and more than 30 talks were presented. Most of the talks in this seminar did not concentrate on a single technique in isolation but analyzed, compared or integrated at least two techniques such as data type and process specifications, sequence charts and state charts, sequence charts and data manipulation, B and state charts, formal models and safety analysis, CSP and Object-Z, Petri nets and algebraic module concepts or tile systems and logic programming. In almost a third of the talks UML and its integration with different other specification techniques, such as ASM, graph transformations, CASL or new kinds of behavioral diagrams, was studied; particular emphasis was given to the analysis and foundation of the Object Constraint Language OCL and its integration in the software development process. One other major topic was the use of transformations of graphs or Petri nets for different applications such as the integration of heterogeneous software specifications, visual specification of diagrammatic modeling techniques, semantics of UML and OCL or dynamic meta modeling.

In addition to the talks several discussion sessions were organized on the applicability of formal methods in practice and on the integration with object-orientation. In order to give the many young participants of the seminar the possibility of expressing their opinions and expectations we divided the speakers in two groups: young versus experienced researchers. This lead to stimulating, controverse and inspiring discussions which continued in the evenings.

27 Modelling of Sensor-Based Intelligent Robot Systems

Seminar No. **00421** Report No. **289** Date **15.10.–20.10.2000**
Organizers: H. Bunke, H.I. Christensen, G. Hager, R. Klein

During the week of October 16th to 20th, 2000, the International Seminar on "Modelling of Sensor-Based Intelligent Robot Systems" was held at Schloss Dagstuhl. This Seminar was the fourth of a series started in 1994.

Intelligent Robot Systems are very complex soft- and hardware sys-

tems. These robots include facilities for action selection based on sensory input and prior knowledge. Intelligent algorithms are interface between sensors and actions. The research area includes significant theoretical and practical problems, especially high-level planning (using AI techniques), geometric methods, real-time software systems and methods for interpretation and fusion of sensory information.

A significant problem is modelling of dynamically changing environments, and robust methods for sensory perception. The gap between theory and practice is a wide one and a lot of research must be done for example for building a human-like robot.

Results

Researchers from more than ten different countries met in an ideal setting and discussed in depth the following issues:

- Mapping and navigation
- Visual servoing
- Planning
- Learning
- Exploration and coordination
- Vision
- Applications
- System architectures and implementation

Perspectives

It became obvious that the ultimate flexible and intelligent robot is a fantasy. The trend is towards specialised robots that are designed for a particular application. The manufacturing of specialised robots must however be based on standardised components to make the final product economically tractable. Overall trends in end-user products and the corresponding needs for basic research were intensively discussed.

An abstract selected by the editor of Dagstuhl News:

Use of Robots in Computer Aided Surgery

Richard Bächler, PhD
M.E. Müller Institute for Biomechanics, University of Bern

Computer aided surgery (CAS, also termed "Image Guided Surgery") has brought the computer not only into the operating room, but into the surgical site itself. By providing advanced planning and navigation aids to the surgeon, the accuracy and safety of surgical procedures can be improved significantly. This is especially valuable in surgeries where the anatomy deviates from the "normal" anatomy, as anatomical landmarks and known features might be missing, e.g., due to trauma or tumors. The main components of a CAS system are the therapeutical object (i.e., the patient respectively the anatomy operated on), the virtual object (i.e., an image of patient's anatomy), and the navigator that provides the link between the therapeutical and virtual objects.

In order to improve the surgeon's tactile abilities, a huge variety of instruments have been devised. The most recent addition to the conventional instrument sets has been the introduction of robots into the operating room. By taking advantage of a robots mechanical features (sub-millimetric positioning accuracy, no tremor, repeatability of movements) the accuracy of the execution may be improved significantly. In the context of a CAS system, the surgical robot is often used as the navigator. Currently, different fields of application for surgical robots can be identified:

- Active robots perform cutting tasks thus replacing the surgeon. The first use of a robot in the operating room was the RoboDoc (developed at IBM), a surgical robot designed for total hip arthroplasty surgery. The RoboDoc system is capable of milling the cavity for the femoral implant in THA surgery with high precision. It performs the preparation for primary as well as for revision THA. Major drawbacks of the system are the use of fiducial pins for registration purposes, the larger and thus more invasive approach to the surgical site, and the prolonged surgery time. Current clinical studies are aimed at determining if the advantages outweigh the disadvantages. Other active robotic systems have been designed for total knee arthroplasty and maxillo-facial interventions.

- Another field of application is the use of robots as "intelligent" tool holders. A pilot application has been developed by the german

space agency DLR that holds the laparascopic camera in abdominal surgery. By tracking color labels on the surgeon's instruments, the robot is able to follow the movements of the surgeon. A study showed that the robotic assistant performed his task with less soft tissue contact than a human assistant (thus reducing the need to retract the camera and wipe the lens) and was able to guide the camera according to the surgeon's needs. This resulted in a shorter surgery time and reduced the stress in the operating room significantly, as the surgeon did not need to tell his assistant where to point the camera. Other systems that fall into this category have been developed, e.g., to hold a trochar along a planned trajectory to allow precise biopsies.

- Dexterity enhancement systems aim at providing the surgeon with an advanced instrument to carry out micro-surgical interventions with an accuracy and stability not possible with the human hand. One of the current research projects is the "steady hand" system developed a Johns Hopkins University.

- Telepresence surgery requires an effector on the patient's side of the system as well as a haptic feedback system on the surgeon's side. The effector is usually a robotic system adapted to the requirements of the planned surgery adding sensors for haptic feedback. The haptic feedback component may also be provided by using a robotic system. In this case, the system can then be used for simulation of surgery. This is currently one of the main research areas for computer aided surgery, as a controlled training and simulation environment has a huge potential for the education of young surgeons, as well as for the planning of new surgical techniques or complex surgeries.

In conclusion, surgical robots have successfully entered the operating room, but the first generation systems still have a number of disadvantages that must be improved before surgical robots will really have a wide-spread use. Current requirements are less invasive approaches, "intelligent" and maybe even semi-autonomous behaviour, easy to use user interfaces. Advancements in these areas will make the surgical robot a standard "instrument" in the operating room of the 21st century.

28 Self-Stabilization

Seminar No. **00431** Report No. **290** Date **22.10.–27.10.2000**
Organizers: A. Arora, J. Beauqier, S. Dolev, T. Herman, W. P. de Roever

Distributed system research is at an interesting juncture. Many basic problems have been identified and intensively studied, often using traditional models of distributed computing and networks. However the nature of applications of distributed systems is changing, and new technological developments are challenging conventional system design. These trends encourage new approaches to problems of distributed control, fault recovery, and adaptive behavior.

Self-stabilization is now acknowledged as an important theme in distributed computing, with papers regularly appearing in prominent conferences. The importance derives from the fact that self-stabilization makes initialization independent from desired convergence in system computations. As a consequence of this fact, protocols and algorithms that are self-stabilizing can automatically adapt to dynamic environments and can recover from faults that put the system in arbitrary states. Many examples of algorithms use only local information to stabilize, and thus scale well to large distributed systems.

This report contains abstracts of presentations given at the second Dagstuhl Seminar on Self-Stabilization. Specialists in the area of stabilization research see new opportunities to apply stabilization to other areas of computer science and engineering, and this was a theme of the seminar. Results presented at the seminar included new directions: relating control theory ideas to self-stabilization; application to higher-level systems research and middleware reliability; application to data structures; group communication that is self-stabilizing; and stabilization in the model of mobile agent computing. In addition to the presentations described by the abstracts, there were several informal talks and open sessions with lively discussions of future directions. Some traditional issues provoked conversations on many topics: model conversions; atomicity and fairness implementations; impossibility results; heuristics and frameworks for proving the convergence of self-stabilizing algorithms; and experiments with new technical definitions of self-stabilization that could benefit more applications.

Two abstracts selected by the editor of Dagstuhl News:

On the Benefits of Modeling Self-Stabilizing Algorithms as Variable Structure Feedback Systems

Oliver Theel
Darmstadt University of Technology, Darmstadt, Germany

The self-stabilizing property of certain distributed algorithms exhibits interesting analogies to stable feedback systems used in various engineering domains, like electrical or mechanical engineering. Informally, a feedback system is stable, if after a certain finite period of time, the system reaches and remains in a pre-defined state. Contrary to the self-stabilization research domain, which is a rather new area of research in computer science, control theory in the engineering domain has a century-old background and offers a broad theoretical foundation with powerful criteria for reasoning about the stability of feedback systems. In this talk, we show how to model a distributed algorithm implemented an a set of processes executing guarded commands as an instance of a discrete-time variable structure system model. Based on this modeling, we prove convergence (and closure) of the sample algorithm by means of Ljapunov's "Second Method." The talk closes with remarks on how Ljapunov Theory can furthermore be used to derive upper and lower bounds for the convergence speed as well as how to automatically identify variant functions for a certain subclass of self-stabilizing algorithms.

Defining Redundancy: In Search of the Holy Grail

Felix Gärtner
TU Darmstadt, Darmstadt, Germany

Redundancy is a key concept in fault-tolerance, however it still lacks a formal definition. In this work we explore the notions of redundancy which are inherent in the fault-tolerance theory of Arora and Kulkarni. In this theory, every fault-tolerance mechanism can be constructed from two simple components: detectors and correctors. Detectors are necessary and sufficient to maintain safety specifications, while correctors are necessary and sufficient to eventually satisfy safety specifications. We investigate the assumptions of this theory and discover that fault-tolerance components contain either non-reachable states or non-reachable transitions. This gives rise to definitions of redundancy in space and redundancy in time. This is very much unfinished work, not all theorems have

been proven yet, and the definitions are somewhat orthogonal to the understanding of the terms in the literature. Because of this and several other reasons, this work might never be finished. Thus, it resembles – at least to the authors – a search for the holy grail.

29 Effective Implementation of Object-Oriented Programming Languages

Seminar No. **00451** Report No. **291** Date **05.11.–10.11.2000**
Organizers: Uwe Aßmann, Laurie Hendren, Barbara Ryder

Motivation

Object-oriented languages are becoming increasingly popular for the development of software systems of all kinds, ranging from small web-based applications to large server-side applications. Modern object-oriented languages such as Java contain features such as exception handling, dynamic binding, extensive control of visibility, and threads. Although these features add to a language's expressive power and provide many benefits from a software engineering point of view, they also make it more difficult to implement a language efficiently.

The purpose of this seminar is to bring together experts in the areas of object-oriented language design, analysis and translation with experts in run-time systems, and to investigate problems and techniques related to effective implementations of object-oriented languages. It is becoming increasingly apparent that all of these areas can contribute to the ultimate goal of obtaining efficient object-oriented implementations. For example, by choosing the right language features, the job of the compiler and run-time system can be made easier. Moreover, static compiler analyses can benefit from profiling information as a means for obtaining feedback on predictable program behavior, and for selecting profitable transformations.

Program

The seminar was organized around the general themes of garbage collection, call graph construction and JVMs for Java. There were invited

tutorials on garbage collection and call graph construction algorithms as well as invited presentations by three Java compiler groups: IBM's Jalapeno compiler, Sun's Hotspot compiler and Microsoft's Marmot compiler. There were 24 additional 30 minute presentations by seminar participants which discussed topics in optimization, analysis, profiling, language design, object persistence and industrial applications. Each presentation session finished with at least a 15 minute question session; often discussions continued into the coffee break or lunch. One night was reserved for a demo session in which participants showed their software tools to other attendees. Many attendees were graduate students; all of them presented their research. Although the bad weather forced cancellation of the traditional Wednesday hike, participants did have a wine tasting dinner in Trier and an IBM-sponsored banquet as part of the seminar.

Goals

An important seminar goal was to gather industrial compiler groups and academic researchers together for substantive research discussions; this goal was achieved as exchanges over topics such as on-stack replacement, utility of SSA-like program representations, effectiveness of different garbage collection methodologies took place both in the seminar sessions and especially during evening informal gatherings. Another goal was to engender more communication between colleagues in the functional programming community and those in the object-oriented community; we were less successful in this effort as several of our functional programming colleagues declined our invitation. Finally, conversations at the seminar made it clear that many researchers, industrial and academic, feel that compiler infrastructure is a critical need of the community to enable better research; colleagues in attendance were sympathetic to providing research infrastructure for joint research between academic and industrial research folks.

Research Areas Uncovered

Many research questions focussed on Java and possible extensions to the programming language. It is clear that genericity and assertions are the awaited features that the community wants in Java, for example. Much discussion centered around how to provide those features of

object-oriented languages that provide flexibility and dynamicism, while maintaining optimized execution performance. Dynamic class loading in Java was cited as a particularly difficult feature to accommodate.

The next big research accomplishment predicted by seminar participants is the development and use of practical interprocedural analyses and the optimizations they facilitate. For example, there was much discussion of how to perform effective "inlining"; no consensus was reached, but it was agreed that more empirical investigation is needed. In addition, participants agreed that thread management is an area needing work.

30 Logic, Algebra, and Formal Verification of Concurrent Systems

Seminar No. **00481** Report No. **292** Date **26.11.–01.12.2000**
Organizers: V. Diekert, M. Droste, A. Muscholl, D. Peled

The seminar brought together researchers developping formal methods and tools for the analysis of concurrent systems with those working on algebraic and logical theories that can potentially enhance the capabilities of such tools.

The interaction of different approaches and methodologies has resulted in new verification techniques and has improved existing ones. For instance, automata on infinite words are simple mathematical models that have been used as the theoretical basis for state-based automatic verification of temporal logic specifications. A recent example of using theoretical models for improving the efficiency of algorithmical analysis is the ITU standard of message sequence charts. The theoretical study of this partial-order-based notation resulted in algorithms that are used to detect errors in preliminary designs of communication softwares.

A basic concern in formal methods is to design and analyze appropriate logical and algebraic formalisms used for describing properties of systems and for checking these properties efficiently for interesting classes of models. In logic as well as in algebra, formalisms with partial-order semantics offer special challenges. For instance, both the expressivity and the algorithmic properties of temporal logic frameworks depend to a large extent upon the kind of concurrency expressed in the model.

Even for the rather simple partial-order model of Mazurkiewicz traces, the existing temporal logics still do not offer a good trade-off between expressivity and complexity, in spite of the considerable progress that has been achieved over the past few years. It is not by coincidence that some of the progress was based on applying algebraic methods, which often offer in a natural way the right way to formalize properties related to concurrency.

The seminar was attended by 35 researchers and 30 talks were presented. They presented ongoing work on generalized models of concurrent systems, expressivity of temporal logics in discrete and timed partial-order models and the verification of infinite-state systems. In a special evening session Yuri Matiyasevich presented the main results and open problems related to Hilbert's 10^{th} problem, a century after its presentation to the 2^{nd} International Congress of Mathematicians.

The discussions during the seminar were very stimulating and brought an intense exchange of ideas. They showed that promising solutions for achieving a good understanding and application of concurrent behavior can result from combining several existing techniques and specification formalisms, with algebra as a guiding formalism.

As a result of the seminar, we think that a natural challenge is the question whether there is a partial-order temporal logics for Mazurkiewicz traces (or other meaningful partial-order based models) which is expressively complete but which has an elementary complexity.

31 Bioinformatics

Seminar No. **00491** Report No. **293** Date **03.12.–08.12.2000**
Organizers: Douglas Brutlag, Thomas Lengauer, Martin Vingron

After two Seminars in the years 1992 and 1995 this was the third Dagstuhl seminar with a broad scope ranging over a variety of fields in Computational Biology.

In the five years since the previous Dagstuhl seminar Computational Biology has experienced dramatic growth and a significant shift of focus. In addition to the classical grand challenge problems such as gene identification and protein folding, computational biology has been confronted

with a host of application-oriented problems. These problems originate from new experimental data being generated with efforts in genomics and proteomics and ask for unveiling the biological secrets that are hidden in these data. Furthermore, the expected publication of the human genome sequence left its distinct mark on the seminar.

This Dagstuhl seminar intended to focus specifically on these topical issues of the field. Specific topics included

support for large scale sequencing (shotgun sequencing the human genome);
annotating biological sequences (coding and non-coding regions);
comparative genomics;
structural genomics;
functional genomics;
analysis and interpretation of expression data;
modeling of cellular processes and pathways;
medical applications, genetics, genotyping;
proteomics.

Besides the talks and informal discussions there were two evening jam sessions on expression patterns and gene identification in the light of the imminent publication of the human genome sequence. There also was a session on new modes of teaching in computational biology.

As probably a premier for Dagstuhl seminars, a crew from German public television visited the group on Tuesday, Dec 5 and aired a short segment on the seminar over national television (Channel: 3sat, Program: Nano) on the following day. Interview partners in the program were Doug Brutlag, Tom Lengauer, and Knut Reinert. An internet version of the segment can be found at
http://www.3sat.de/nano/astuecke/13391/index.html.

32 Security through Analysis and Verification

Seminar No. **00501** Report No. **294** Date **10.12.–15.12.2000**
Organizers: Pierpaolo Degano, Roberto Gorrieri, Chris Hankin, Flemming Nielson, Hanne Riis Nielson

Security has increasing relevance and importance for real life applications such as Internet transactions, electronic commerce, electronic voting and smart cards thereby stressing the need for appropriate checking measures. Techniques from program verification and program logics have already proven their worth for the machine-assisted validation of secure communication protocols. This seminar emanates from recent applications of program analysis which allow fully automatic validation of systems against certain types of attack. It took place from 10. December 2000 to 15. December 2000 and comprised a number of talks, informal discussions as well as a discussion in plenum.

The nature of security. Security has many facets: confidentiality, integrity, authentication, watermarking, prevention of denial of service, auditing etc. It emerged quite clearly during the seminar that there is no clear consensus about the precise definition of many of these concepts. There was agreement that to the extent possible one should try to reconcile and clarify the differences, possibly in the form of defining a formal language or logic for security requirements. During the discussions it became clear that confidentiality, as in certified mail, was a hard concept to capture precisely, despite the fact that consideration of information flow (both forward and backward) accounts for many ingredients of integrity and secrecy.

Formal approaches to security. The main techniques represented at the seminar involve flow-based program analysis, model checking, type systems, simulation techniques, process algebras, cryptographic proofs and dynamic enforcement of security policies. Several presentations aimed at spanning a range of these and a particularly promising direction was the integration of cryptographic reasoning into the more classical approaches to formal methods.

One approach to the security of a system centered around the idea that the harder to make an intrusion the more secure the system. Several presentations expanded on the ability to intercept, manipulate and forge messages, including polynomial-time bounded attackers, Dolev-Yao saboteurs and other forms of characteristic intruders.

As pointed out by the security community there is a need to address a variety of more realistic applications and protocols. Indeed a number of techniques have only been applied to the simplest protocols. By con-

trast, applications towards multicast protocols and group-key management seem to challenge state-of-the-art in the most advanced approaches.

There also is a need to identify the most appropriate specification formalisms and programming calculi for capturing both qualitative and quantitative, possibilistic and probabilistic aspects of secure systems and protocols.

There was no agreement about whether the approach taken should aim at studying protocols and procedures at the abstract level or should eventually aim at studying the fielded software system. However, there was consensus that special care is required when applying well-defined mathematical theories and formal validations to actual systems.

Conclusion. Thanks to the pleasant surroundings the different communities taking part in the meeting were very successful in establishing an amiable social atmosphere thereby facilitating the necessary cross-fertilization for the field to progress. Given the predictable growth in electronic commerce and mobile computing the problems discussed are increasingly important and can only benefit from the interaction of a variety of approaches. It is our prediction that the field will only be slightly perturbed by future developments in quantum cryptography and the emerging computational models.