

## Contents

---

Preface: Professor Zhenjiang Hu	02
Main Chapter of Seminars No.1-No.100	05
Introduction of Shonan Meetings and Committee	107
Impression of Participants	111
Memory of Shonan Meetings	119
Introduction of Shonan Village Center	123

---



NII  
SHONAN  
MEETING

1000<sup>th</sup>  
*Commemoration*

REFLECTIONS  
ON THE OCCASION  
OF THE **100**<sup>TH</sup>  
SHONAN MEETING

October 30th 2017 marked the 100th event of the Shonan Meetings, the first international Dagstuhl-style meetings in Asia for top-class academics and engineers to discuss openly problems and research visions in informatics. By November 2nd 2017, 100 meetings had been successfully held with 2447 participants from 56 countries and regions. This success would not be possible without the considerable effort of the meeting organizers, who have carefully designed the program, and of the participants, who have actively contributed to intensive discussion and research communication. In particular, it owes much to the Shonan Village Center for taking care of all on-site local arrangements and to the local government of Kanagawa Prefecture for generously providing space in the annex for holding the Shonan Meetings. We, the members of the organizing committee, are proud of the Shonan Meetings, and we would like to take this opportunity to express our deepest gratitude.

While the first Shonan Meeting was held in February 2011, the feasibility study at NII started as early as 2008. Following the idea of the Dagstuhl Seminars, we planned our Shonan Meetings to have four features: each meeting is small with about 25-35 participants, all participants are invitation-based, it is Asia-based where at least one of the meeting organizers is from Asia, and it provides the best meeting environment with Japanese hospitality. We had a dream of making the Shonan Meetings the “Mount Liang (梁山泊) in Informatics,” an academic hub where heroes in informatics gather and take on the important challenges. One big concern at the time was whether researchers from America or Europe would be willing to take such a long trip to Japan to attend the meetings. This concern quickly disappeared after a successful trial, the GRACE Meeting on Bidirectional Transformations held in Shonan Village Center in December 2008. The GRACE Meeting attracted about 30 internationally active researchers; 95% of them were very satisfied with the meeting and wished to have such a style of meetings in Japan in the future.

This was indeed very encouraging!

The Shonan Meetings did not start as smoothly as we hoped. Just after the first, successful Shonan Meeting in February 2011, we had a big earthquake and tsunami in Japan, which forced us to stop several Shonan Meetings that were supposed to follow the first one. It was difficult to decide when these meetings would be resumed. With the great effort of the organizers and the participants, the Shonan Meetings resumed in just half a year, which was a miracle. Since then, the Shonan Meetings have gained a reputation as a premier venue for intensive research collaboration, and they have attracted more and more proposal submissions. Last year, we accepted about 30 excellent meeting proposals.

While the NII Shonan Meetings are invitation-based, the results from the meetings are published widely in a timely manner as technical reports. Some of them are even published as books in Springer's NII Communication Series. The books in the series aim to provide a vision of future trends inspired by discussions at the meetings and the latest research directions based on surveys and discussions. In addition, we invite internationally outstanding scientists to organize the NII Shonan Schools (NSS), which aim to provide a series of advanced lectures for about 30 participants who are mainly PhD students and young researchers and provide an opportunity for the participants to build up, enhance and widen their international collaboration network.

Moshe Vardi, former Editor-in-Chief of CACM, raised the question of "Where have all the workshops gone?" in an editor's letter in 2011, calling for a revival of the "old-style" workshops where researchers could informally share their latest work in progress and freely interact with each other. The Shonan Meetings have witnessed the importance of such workshops, and it is no wonder that some other countries such as Korea are now planning similar workshop centers. We are glad that the Shonan Meetings can provide a top-notch international collaboration environment to researchers in informatics, and we believe that the Shonan Meetings will continue to make a big impact on the world.



Zhenjiang Hu  
Professor

NII Shonan Meeting Academic Chair

教授 胡 振江



MAIN CHAPTER  
OF  
SEMINARS  
NO.001-NO.100

# Graph Algorithm and Combinatorial Optimization

# 001



Shonan Village Center, February 13–18, 2011

Seminar No.001

Organizers

Satoru Iwata, RIMS, Kyoto University, Japan

Ken-ichi Kawarabayashi, National Institute of Informatics, Japan

Structures that can be represented as graphs are ubiquitous, and many practical problems can be represented by graphs. The link structure of a web in Internet could be represented by a directed graph: the vertices are the web pages available at the website and a directed edge from page A to page B exists if and only if A contains a link to B. A graph structure can be extended by assigning a weight to each edge of the graph.

Networks have also many uses in the practical side of graph theory, network analysis (for example, to model and analyze traffic networks). Within network analysis, the definition of the term “network” varies, and may often refer to a simple graph.

A similar approach can be taken to problems in travel, biology, computer chip design, and many other fields. Development of algorithms to handle graphs is therefore of major interest in computer science.

These applications of graphs often gives rise to optimization. Basic optimization problems on graphs, including the shortest path, maximum flow, minimum spanning tree problems allow efficient exact algorithms. The algorithmic developments of these problems have led to the theory of combinatorial optimization, combined with polyhedral combinatorics, matroids and submodular functions.

On the other hand, most practical optimization problem on graphs such as the traveling salesman, stable set, maximum cut problems are NP-hard. Approximation algorithms for these NP-hard combinatorial optimization problems have been investigated extensively for a couple of decades. Design of approximation algorithms often requires deep insights from structural graph theory and polyhedral combinatorics.

The purpose of this workshop is to bring experts in graph algorithm and combinatorial optimization to share ideas, and to stimulate joint projects.

# Agda Implementors Meeting

Shonan Village Center, September 8–14, 2011

Seminar No.008

## Organizers

Peter Dybjer, Chalmers University of Technology, Sweden  
Yoshiki Kinoshita, National Institute of Advanced Industrial Science and Technology (AIST), Japan  
Shin-Cheng Mu, Academia Sinica, Taiwan



Agda is a dependently typed functional programming language. It has inductive families, i.e., data types which depend on values, such as the type of vectors of a given length. It also has parametrised modules, mixfix operators,

Unicode characters, and an interactive Emacs interface which can assist the programmer in writing the program.

Agda is a proof assistant. It is an interactive system for writing and checking proofs. Agda is based on intuitionistic type theory, a foundational system for constructive mathematics developed by the Swedish logician Per Martin-Löf. It has many similarities with other proof assistants based on dependent types, such as Coq, Epigram, Matita and NuPRL.

The Agda Intensive Meetings (called Agda Implementors Meetings in the earlier days) have been held twice a year in principle since 2004. The record of recent meetings can be found in the following URL:

<http://wiki.portal.chalmers.se/agda/pmwiki.php?n=Main.AgdaMeetings>

The AIM meetings are really “Agda Users and Implementors Meetings,” that is, meetings where Agda users and implementors get together and exchange ideas, write Agda code, while simultaneously implementors work on modifications of the Agda system.

The AIM meeting usually consists of talks and code sprint sessions. Talks are invited by the organiser, and all participants are expected to propose his/her own project for a code sprint session.

At the start of AIM, the participants get together and are expected to participate in projects they wish to take a part in. Then it is discussed which project to start. A short meeting is held at the end of every day so that each project reports its progress. A longer meeting is held at the end of the whole AIM to summarize the result and future of the project.

Everyone who is seriously interested in Agda, not only implementors but those who are willing to write code in Agda or documentation about Agda is welcome to the meeting. They are expected to make a constructive contribution to the code sprint.



# 003 Dependently Typed Programming



Shonan Village Center, September 14–17, 2011

Seminar No.007

## Organizers

Shin-Cheng Mu, Academia Sinica, Taiwan  
Conor McBride, University of Strathclyde, United Kingdom  
Stephanie Weirich, University of Pennsylvania, USA

A thriving trend in computing science is to seek formal, reliable means of ensuring that programs possess crucial correctness properties.

One path towards this goal is the design of high-level programming languages which enforces that program, by their very own structure, behave correctly. In the past decade, the use of type systems to guarantee that data and program components be used in appropriate ways has enjoyed wide success and is still a main focus both from researchers and developers.

With more expressive type systems, the programmer is allowed to specify more fine-grained properties the program is supposed to hold. Dependent type systems, allowing types to refer to (hence depend on) data, are particularly powerful since they reflect the reality that what defines “appropriate” program behaviours is often dynamic. With the type system, programmers are free to communicate the design of software to computers and negotiate their place in the spectrum of precision from basic memory safety to total correctness.

Dependent types have their origin in the constructive foundations of mathematics, and have played an import role in theorem provers. Advanced proof assistants, such as Coq, has been used to formalise mathematics (e.g. the proof of the four colour theorem) and verify compilers. Only recently have they begun to be integrated into programming languages. The results so far have turned out to be fruitful. A number of dependently typed programming language and systems have been proposed, including Cayenne, DML, Epigram,  $\Omega$  mega, ATS, Agda, Guru, etc, showing a maturity beyond the proof-of-concept stage. Papers on dependent types submitted to and published in major conferences have significantly increased in number. Some more “mainstream” functional programming languages, such as Haskell, also start to adopt features such as GADT and type family that are strongly influenced by dependent types.

Dependently typed programming, however, is yet to be considered a practical tool and an efficient device to ensure program correctness and to reduce development costs. Many issues remain to be resolved, including but not limited to:

- design of a small but expressive core language on which to built up a verifiable metatheory;
- metaprogramming, reflectivity, and the possibility of representing dependently typed terms in a dependently typed language;
- representing variable-binding and supporting both structural inspection and functional usage;
- interpreting representations to automate problem solving;
- separating and redesigning a language of “proof” as distinct from a language for “programming”;
- integrating extensional reasoning about functions with computation in dependent type systems, while at the same time concealing the internal structure of proofs;
- modelling interaction and communication in distributed, concurrent systems using dependent types.

This workshop aims to provide a venue for people actively working in this field so that these issues could be discussed.

# Knowledge-leveraged Computational Thinking through Natural Language Processing and Statistical Logic

004

Shonan Village Center, September 19–21, 2011

Seminar No.012

## Organizers

Akihiro Yamamoto, Kyoto University, Japan

Kentaro Inui, Tohoku University, Japan

Sadao Kurohashi, Kyoto University, Japan



A long-standing problem in Natural Language Processing has been a lack of large-scale knowledge for computers. The emergence of the Web and the rapid increase of information on the Web drastically changed the environment of NLP. The Web is not only a marvelous target for NLP, but also a valuable resource from which knowledge could be extracted for computers, making research and development activities on large-scale text processing and large-scale knowledge acquisition much more popular.

However, beyond the success of large-scale NLP and knowledge acquisition, we are starting to face a new problem: how to manage and use the automatically acquired knowledge. We are still not confident that automatically acquired large-scale knowledge resources will indeed solve NLP problems in real world applications. How to incorporate the acquired knowledge into existing NLP frameworks and how to manage them are yet unsolved issues.

Based on this background, the workshop introduces the new research field of Computational Thinking, where computers themselves directly make use of the large-scale knowledge on the Web by combining natural language processing and statistical logical inference. In natural language processing, we need robust structuring of text, making advances in discourse analysis, zero pronoun resolution, and syntactic and semantic parsing. In addition, to handle the uncertainty and ambiguity inherent in language and in automatically acquired knowledge, we also have to construct a theory and system of inductive knowledge derivation, allowing us to carry out statistical inductive logic computation on a massive scale.

Based on these advances, we can develop a high-dimension statistical induction framework that finds links in the useful knowledge represented in language, and by doing so, construct a knowledge analysis engine that goes beyond mere search, effectively harnessing the entire Web knowledge space.

The aim of the workshop is to bring researchers and practitioners together in order to discuss Knowledge-leveraged Computational Thinking. Possible topics of the workshop include, but are not limited to:

- Construction of large corpora and sharable large resources
- Knowledge-leveraged robust structuring of text
- Bootstrapping knowledge acquisition from very large corpora
- Framework of very-large-scale statistical logic
- Inductive knowledge derivation and distillation
- Knowledge-based information access, analysis, and organization

# Workshop on Automated Techniques for Higher-Order Program Verification

# 005



Shonan Village Center, September 23–27, 2011

Seminar No.005

## Organizers

David Van Horn, Northeastern University, USA  
Naoki Kobayashi, Tohoku University, Japan  
Luke Ong, University of Oxford, United Kingdom

With the increasing importance of software reliability, program verification has been an important and hot research topic. Recently, we have seen some good progress in automated techniques for verification of higher-order programs.

Studies of game semantics have yielded compositional model checkers and automated program equivalence checkers for Algol-like programming languages, and studies of higher-order recursion schemes and pushdown automata have yielded model checkers for higher-order functional programs. Classical control flow analysis has been recently revisited to yield more precise and/or efficient methods than Shivers' k-CFA.

The aim of the workshop is to bring together researchers on automated techniques for higher-order program verification and analyses, and provide them with an opportunity to exchange new research results, and discuss further extensions.

The workshop also aims for cross-fertilization of different techniques for higher-order program verification, such as game semantics, type theories, higher-order grammars and pushdown systems, control flow analyses, and abstract interpretation.

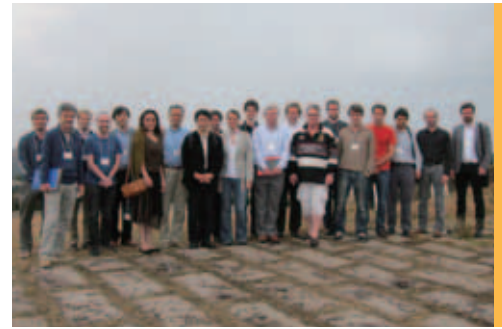
# Hybrid Quantum Devices

Shonan Village Center, November 6–9, 2011

Seminar No.006

## Organizers

Gerard Milburn, The University of Queensland, Australia  
Jörg Schmiedmayer, Technische Universität Wien, Austria  
Yasuhiro Tokura, NTT, Japan  
Kae Nemoto, National Institute of Informatics, Japan



In recent years techniques for quantum control have dramatically improved and resulted in a number of demonstrations in different physical systems at the few qubit level. Based on these new developments, many proposals to implement quantum devices have been mushrooming both in number and variety. At this stage of quantum information device research, it is essential to make sure the fundamental quantum control has its future as a practical device for the future technologies. The implementation of quantum devices, however, is usually not carefully designed as a fundamental device for larger scale quantum information processing and could potentially limit the scalability of the system. Recent theoretical analysis shows that hybrid and distributed quantum devices exhibit better scalability and so are promising for scalable architectures in large-scale quantum information systems. In this workshop, we focus on such hybrid and distributed devices bringing such state of the art quantum information systems together. In particular, devices from optics to microwave will be extensively investigated. We will focus on current problems, future directions and what can be achieved in the next few years.

# International Symposium on Symbolic Systems Biology

# 007



Shonan Village Center, November 14–17, 2011

Seminar No.010

## Organizers

Katsuhisa Horimoto, National Institute of Advanced Industrial Science and Technology (AIST), Japan  
Katsumi Inoue, National Institute of Informatics, Japan  
Oliver Ray, University of Bristol, United Kingdom

*Symbolic Systems Biology* is a growing area of research involving the application of formal logic-based methods to systems biology and bioinformatics. With biological data being acquired at ever increasing rates, purely numerical techniques must be combined with symbolic approaches in order to help formalize expert knowledge and integrate information across different levels of biological abstraction. Recently, a number of symbolic approaches have been developed and usefully applied to a variety of biological problems. Such methods include formal logics (e.g., propositional/first-order/modal frameworks), computational logics (e.g., constraint/logic/answer-set programs), graphical models (e.g., Boolean/Bayesian/Petri nets), synthetic inference (abduction/induction), formal methods (e.g., model checking/pi-calculus/hybrid logic), qualitative reasoning, action languages, and statistical relational learning.

But, currently, there is an urgent need for a methodical comparison of these existing symbolic systems biology approaches and the types of biological problems to which they have been applied. This is necessary to understand the strengths and weaknesses of each approach and the characteristic features of their respective applications. Such a study would help to facilitate research in symbolic systems biology by providing a roadmap of which systems are best suited to which problems and allowing more effective exploitation and re-use of algorithms and data.

At the same time, there is also a need for more collaboration between numerical and symbolic biologists. This is necessary to better understand the advantages and drawbacks of the quantitative and qualitative approaches and progress towards a synergistic integration of the two. Ideally this should be done with advice from experimental biologists who are better informed of hot applications and emerging methods of data acquisition. This will help to provide a real context in which symbolic systems biology can be more usefully developed.

# Large-scale Distributed Computation

Shonan Village Center, January 12–15, 2012

Seminar No.011

## Organizers

Graham Cormode, AT&T Labs-Research, USA

S. Muthukrishnan, Rutgers University, USA

Ke Yi, Hong Kong University of Science and Technology, Hong Kong



As the amount of data produced by large scale systems such as environmental monitoring, scientific experiments and communication networks grows rapidly, new approaches are needed to effectively process and analyze such data. There are many promising directions in the area of large-scale distributed computation, that is, where multiple computing entities work together over partitions of the (huge) data to perform complex computations. Two important paradigms in this realm are distributed continuous monitoring, which continually maintains an accurate estimate of a complex query, and MapReduce, a primarily batch approach to large cluster computation.

### Distributed Continuous Monitoring

In many settings the new data is observed locally—at a router in a network, at a sensor in a larger sensor network. The volume of observations is too large to move to a central location and process together; instead, it is necessary to perform distributed computation over the data. Since the new observations are continually arriving, we must produce a continual answer to complex monitoring queries, all while ensuring that the communication cost necessary to maintain the result, and the computational cost of the tracking, are minimized to meet the data throughput demands. In recent years, there have been several advances in this field.

However, there remain many challenging questions to address in this area.

### Cluster Computing

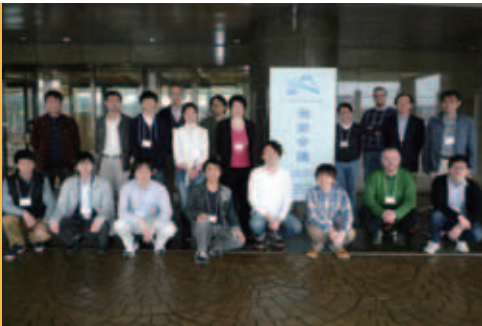
As data sizes increase while the power of individual computing cores begin to saturate, there has been a move to adopt cluster computing: harnessing multiple computing nodes to work together to solve huge data processing tasks in parallel. The best known example of this is the MapReduce paradigm, and its open source Hadoop implementation. Computationally speaking, the approach is for each compute node to process data which is stored local to it in a distributed file system, and Map this data by deriving a new data set indexed by a key value. The system then collects all tuples with the same key together at a second node, which proceeds to reduce these tuples to obtain the (partial) output. This paradigm has proved highly successful for many large scale computations, such as search engine log-analysis, building large machine learning models and data analysis such as clustering.

The aim of this workshop is to bring together researchers active in the areas of distributed data processing, to address these fundamental issues. We hope to encourage greater cooperation and interaction between currently separate communities, ultimately leading to new advances in these important developing areas.

(Excerpt from the submitted proposal)

# Hybrid Systems: Theory and Practice, Seriously

# 009



Shonan Village Center, April 23–26, 2012

Seminar No.023

## Organizers

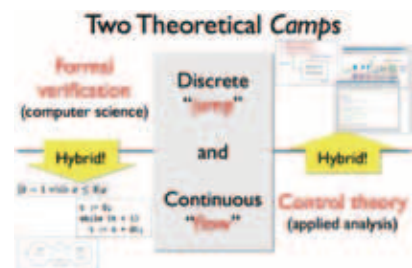
Ichiro Hasuo, The University of Tokyo, Japan  
Takuro Kutsuna, Toyota Central R&D Labs, Japan  
Toshimitsu Ushio, Osaka University, Japan

Hybrid systems—those which exhibit both continuous *flow* and discrete *jump* dynamics—are everywhere in the modern world, with cars, airplanes and all others controlled by computers.

Their failure can therefore have an immense impact on human lives and infrastructures, posing the problem of their quality assurance—getting hybrid systems right—as a pressing one.

### Need of Hybrid Research Community for Hybrid Systems

The name *hybrid system* itself manifests a research challenge: due to the heterogeneity of its dynamics as well as the diversity of its applications, no research effort is comprehensive if it stays within the realm of a single, already established, research discipline. Currently there are two theoretical “camps” aiming at hybrid applications:



The proposed SHONAN Meeting hopes to serve as a meeting point of these three camps—two theoretical ones that have been developed rather separately (up to now, to our regret), and the community of practitioners with whom theoreticians have not had much contact.

### Wanna Be Serious? Then Be Informal!

We all know, however, that such interdisciplinary attempts are very likely to fail. Piet Hut, an astrophysicist and Professor of Interdisciplinary Studies at Institute of Advanced Study, Princeton, US, suggested one way to cope with the difficult task. It goes as follows.

Why is it that two people from different disciplines do not understand each other?

Thus a question: how do we communicate tacit knowledge?

Piet’s answer is clear enough: by being informal! In addition to formal occasions like talks/lectures, it is essential to mingle together over meals, drinks, excursions, etc. That is his experience and also what we all witnessed throughout *Adventures of Categories*, a project for interdisciplinary collaboration for which Piet Hut and Ichiro Hasuo (one of the organizers for this proposal) were organizers.

### Shonan Meeting as an Optimal Venue

For such “serious” interdisciplinary collaboration, the scheme of NII Shonan Meeting offers an optimal setup with intensive interaction lasting four days. We will feature much more time for free-style discussion than workshops would do usually; specifically alteration between 1-hour talk(s) and 1-hour discussion, with which we have had a success in the aforementioned *Adventures of Categories* project.

(Excerpt from the submitted proposal)

# Engineering Autonomic Systems (EASy)

Shonan Village Center, May 15–17, 2012

Seminar No.004

## Organizers

Arosha Bandara, Open University, United Kingdom  
Shinichi Honiden, National Institute of Informatics, Japan  
Yijun Yu, Open University, United Kingdom



Increasing complexity of software systems poses a number of challenges for software engineers, IT service managers and end-users. In 1993, Kephart and Chess published their vision of autonomic computing, which aimed to address some of the challenges of software complexity. The essence of this vision was to create systems that would be able to adapt to their operating environment in a manner analogous to the autonomic nervous system, allowing human administrators and users to concentrate on setting the longer-term high-level goals for the system rather than the operational minutiae required to keep it running on a day to day basis. The autonomic paradigm brought together existing research in the areas of control systems, adaptive networking and context-aware computing which focussed on the challenges of developing the underlying technical frameworks that could enable autonomic operation. At the same time new areas of research, such as Business Driven IT Management, have developed to investigate ways in which users can configure autonomic systems in ways that meet the higher-level objectives of organisations.

As a result of these research efforts, several approaches have been proposed to address different aspects of the autonomic computing challenge—from policy-based systems and biologically-inspired computing at the architecture level to techniques for analyzing autonomic systems and mapping business requirements into specifications for autonomic behaviour. However, numerous issues remain, including the following:

- How do we engineer software for autonomic systems?
- What are the usability issues of autonomic systems? How do we ensure effective interaction with complex software systems that have autonomic components?
- How will deployed autonomic systems evolve? How can we ensure their evolution happens in a systematic way?
- Autonomic systems can be just one component of much larger (and more complex), software intensive socio-technical systems—what are the theoretical frameworks that help us understand these systems and ensure they meet the stakeholders' requirements?
- How do we prevent failures in these complex socio-technical systems?

The purpose of this workshop is to bring together leading researchers from a diverse range of disciplines to discuss the latest research in the area of self-adaptive (autonomic) systems, and explore new ideas, positions, opinions, problems and solutions that could advance the state of the art in this area.

(Excerpt from the submitted proposal)



# Bridging the Theory of Staged Programming Languages and the Practice of High-Performance Computing

# 011



Shonan Village Center, May 19–22, 2012

Seminar No.019

## Organizers

Oleg Kiselyov, Monterey, CA, USA

Chung-chieh Shan, Cornell University, USA

Yukiyoshi Kameyama, University of Tsukuba, Japan

We propose a discussion-heavy workshop to bridge the theory of programming languages (PL) with the practice of high-performance computing (HPC). The topic of the discussion will be code generation, or staging, a form of meta-programming. Both the PL and HPC communities have come to realize the importance of code generation: whereas PL theorists widely regard staging as the leading approach to making modular software and expressive languages run fast, HPC practitioners widely regard staging as the leading approach to making high-performance software reusable and maintainable. Thanks to this confluence of theory and practice, we have the rare chance to bring together PL researchers and the potential consumers of their work.

A particular area of current interest shared by PL and HPC researchers is how to use domain-specific languages (DSL) to capture and automate patterns and techniques of code generation, transformation, and optimization that recur in an application domain.

Alas, the communication between PL researchers working on staging and HPC practitioners could be better. On one hand, HPC practitioners often do not know what PL research offers. Staged programming can detect and prevent these problems early during development, thus relieving users from scouring reams of obscure generated code to debug compiler errors, or waiting for expensive trial computations to produce dubious results. On the other hand, PL researchers often do not know how much HPC practitioners who write code generators value this or that theoretical advance or pragmatic benefit—in other words, how the HPC wish list is ranked by importance.

This workshop aims to solicit and discuss real-world applications of assured code generation in HPC that would drive PL research in meta-programming. Specifically, we would like to determine

- how viable assured (MetaOCaml-like) meta-programming is for real-world applications;
- how strong the demand is for static assurances on the generated code: well-formedness, well-typedness, numeric stability, absence of buffer overflows or dimension mismatch errors, etc.;
- how important portability is, whether to move to a different target language or a different hardware platform;
- which difficulties are “just” engineering (e.g., maintaining a viable, mature meta-programming system), which difficulties are of finding a good syntax, and which difficulties are foundational (e.g., code generation with binders and effects).

In short, we ask how program generation can or should help HPC.

We anticipate the workshop participants to consist of three groups of people: PL theorists, HPC researchers, and PL-HPC intermediaries (that is, people who are working with HPC professionals, translating insights from PL theory to HPC practice). The workshop would benefit PL and staging theorists by informing them what HPC practice actually needs. HPC practitioners may also benefit, for example by learning new ways to understand or organize the code generators they are already writing and using. The goal of the workshop is to collectively produce a roadmap for future research and development that will enhance the state-of-the-art in engineering for autonomic systems. The organisers hope that such a roadmap will guide future researchers to make new breakthroughs in that will address the challenges of a world which will increasingly depend on complex software systems. We plan for the workshop to have lots of time for discussion, and we emphasize presentations that elicit questions rather than merely provide answers.

(Excerpt from the submitted proposal)

# 012 Parallel Methods for Constraint Solving and Combinatorial Optimization

Shonan Village Center, May. 28–31, 2012

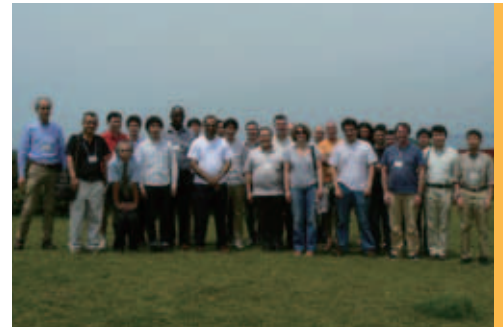
Seminar No.014

## Organizers

Philippe Codognet, JFLI-CNRS, UPMC, The University of Tokyo, Japan

Kazunori Ueda, Waseda University, Japan

Hiroshi Hosobe, National Institute of Informatics, Japan



In the last decade, with the development of multi-core workstations, the availability of GPGPU-enhanced systems and the access to Grid platforms and supercomputers worldwide, Parallel Programming reached mainstream programming and appeared as a key issue in order to use in an efficient manner the computing power at hand.

With the move towards Exascale computing during this decade, this trend will develop all the more.

Search methods and combinatorial optimization techniques are not isolated from this phenomenon, as bigger computing power means the ability to attack more complex combinatorial problems.

In the last years some experiments have been done to extend to parallel execution search methods such as Constraint Programming or SAT solving (Boolean satisfiability), and combinatorial optimization methods such as Local Search, Meta-heuristics and Branch & Bound. However these works have mostly been done for shared memory multi-core systems (i.e. with a few cores) or for small PC clusters (a few machines).

The next challenge is to devise efficient techniques and algorithms for massively parallel computers with tens or hundreds of thousands of cores in the form of heterogeneous hybrid systems based on both multi-core processors and GPUs.

We would like to provide a cross-community forum for researchers working on search methods (Constraint Solving, Artificial Intelligence, Logic Programming, SAT solving, etc.), combinatorial optimization methods (metaheuristics, local search, tabu search, evolutionary algorithms, ant colony optimization, particle swarm optimization, memetic algorithms, and other types of algorithms) and High Performance Computing (Grids, large PC clusters, massively parallel computers, GPGPUs) in order to tackle the challenge of efficient implementations on all kinds of parallel hardware: multi-core, GPU-based or heterogeneous massively parallel systems.

Topics that will be addressed include:

- parallelization of existing search algorithms and new parallel methods;
- constraint solving and SAT solving on parallel hardware;
- heuristic search algorithms and combinatorial optimization on parallel hardware;
- programming paradigms, languages and implementation issues for Grids, large PC clusters, massively parallel computers, and GPUs;
- heterogeneous massively parallel systems
- adaptive strategies and learning for parallel search and optimization;
- applications and benchmarking;
- theoretical studies and complexity.

This meeting is designed to be a forum for researchers willing to tackle those issues, in order to exchange ideas, theoretical frameworks, design of algorithms and methods, implementation issues, experimental results and further boost this growing area through cross-fertilization.

# A Meeting of IT-Enable Services (ITeS)

# 013



Shonan Village Center, July 30–August 2, 2012

Seminar No.021

#### Organizers

Yu-Hui Tao, National University of Kaohsiung, Taiwan  
Nagul Cooharajanone, Chulalongkorn University, Thailand  
Shiro Uesugi, Matsuyama University, Japan  
Hitoshi Okada, National Institute of Informatics, Japan

This meeting is designed to serve as a forum to discuss IT-enable Services (ITeS). The Internet is certainly one of the core infrastructures in the global economy. As multinational corporations operate globally, mobility of goods, human resources and information resources are expanding in a tremendous scale. Penetration of broadband lets the movement accelerated not only in the scale but also in the scope. Use of broadband lets services in digital form to become ITeS. The scale and scope of ITeS are expanding.

Expected topics for discussions also include but not limited to information management, business administration with the help of ICT, innovative use of the Internet for providing services, and so on.

We invite not only social science researchers but also engineers, all who are interested in open up the frontier of collaborative development of application of Internet—the application in ITeS.

<Topics are the followings, but it is not limited>

IT-enabled Service Platform Management  
Business model and Applications of ITeS  
Effective use of Ubiquitous network architecture  
Contents Provision and IT-enabled Services  
Protection of Privacy and Security of ITeS  
Business Ethics and Public Policy of ITeS

# 014 Whole-Session Evaluation of Interactive Information Retrieval Systems

Shonan Village Center, October 9–12, 2012

Seminar No.020

## Organizers

Nicholas Belkin, Rutgers University, USA  
Susan Dumais, Microsoft Research, USA  
Noriko Kando, National Institute of Informatics, Japan  
Mark Sanderson, RMIT University, Australia



Information retrieval (IR) has a long and proud history of evaluation of IR system performance, from the beginnings of such research in the early 1960s at the Cranfield Institute of Technology, UK, through the most recent (the 20th) 2011 Text REtrieval Conference (TREC), held at the National Institute of Science and Technology, Gaithersburg, MD, USA. The very strong emphasis on substantive and rigorous evaluation has been a hallmark of this area of computer and information science, and has led, over the years, to ever more substantive models of IR in general, and of IR techniques, and to substantial increase in performance of IR systems.

However, it is the case that the specific evaluation paradigm that has been almost universally applied in IR research is, in some respects, quite limited, and may no longer be applicable to the evaluation of performance of contemporary and future IR systems, nor to the development of better theories of IR, and more effective IR systems.

Thus, an effective IR system, which we will now call an Interactive IR (IIR) system to distinguish it from the type of non-interactive system whose performance the traditional paradigm evaluates, is one which effectively supports the searcher throughout the search session, and indeed with respect to the search session as a whole.

Although the problem we have identified is reasonably clear—that for further progress in theoretical understanding and increased effectiveness of IIR systems, evaluation of performance of IIR systems should be in terms of how well they support searchers with respect to whole search sessions—how actually to evaluate such interactions is not at all clear.

We believe that a quite new evaluation paradigm will be necessary in order to achieve this goal, based on new and more reasoned understandings of IIR itself, with new measures of performance, and new methods for applying such measures. Thus, the goal of the meeting we propose is to bring together researchers in IIR theories, techniques and methods; in evaluation of IR systems in general; and in particular in session-based evaluation of IIR systems, in order to form a framework within which a new paradigm of evaluation of information retrieval systems can be developed.

There are many issues which arise as problematic when considering session-based evaluation of IIR systems. Here, we mention a few such issues as examples of topics for discussion at the proposed meeting.

A fundamental problem in such an evaluation paradigm is how to identify the goal of a search session, in order evaluate the performance of the IIR system in helping the searcher to achieve that goal.

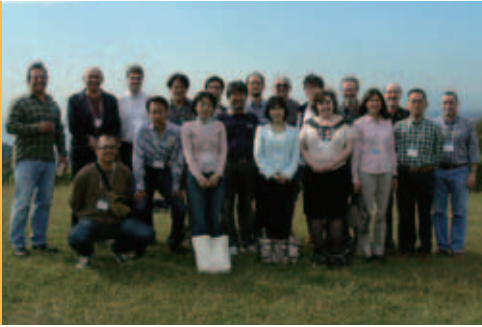
Beyond these fundamental issues associated with understanding the nature of the IIR situation, there are quite substantive methodological problems associated with the evaluation of whole search sessions.

We propose to address these, and related issues in the meeting in a relatively structured way.

(Excerpt from the submitted proposal)

# Grid and Cloud Security: A Confluence

# 015



Shonan Village Center, October 15–18, 2012

Seminar No.015

## Organizers

Barton P. Miller, University of Wisconsin, USA  
Yoshio Tanaka, National Institute of Advanced Industrial Science and Technology (AIST),  
Japan  
Elisa Heymann, Autònoma de Barcelona, Spain

## Topics of the Seminar

The threats are clear: crime gangs, rogue nation-states and terrorists are using the Internet to raise money, steal information, promote their causes, and disrupt the infrastructure of those that they oppose. These organizations are motivated, well-funded, and highly skilled.

The cost to society of not addressing these threats is high: lost revenue and funds, lost data, violated privacy, disrupted network services, and even the potential to disrupt physical services. Often the extent of damaged is not known until long after an attack is successfully executed.

The target environment is rich: online services and databases that hold valuable personal, financial, commercial, and scientific data, and that provide critical services to each of the areas. The cost of disruption can be measured in terms of monetary loss, damage to business, loss of privacy, and severe delays in scientific progress.

In the past several years, Grids and Clouds have emerged as particularly productive environments for leveraging the technology of the Internet.

In each of these two environments, unrelated users are accessing common computational, network, and storage resources, and organizations are extending their trust boundaries well beyond the traditional physical limitations of their own facilities.

The software that comprises the services provided by each of these environments is complex and multifaceted, and not well understood by the typical user. As a result, the threats that seem obvious to a user may have an actual risk that is quite minimal, while there are significant and emerging threats of which users are completely unaware.

Our goal for this Seminar is to take advantage of Shonan's unique environment to bring together a diverse community of researchers, practitioners, and developers across several dimensions:

1. Grid and Cloud security;
2. Industry, government, and academia;
3. Theoretical and practical interests; and
4. Scientific and business communities.

## Goals of the Seminar

A Shonan meeting offers the unique opportunities to harvest the benefits of both a meeting of Grid and Cloud experts as well as a meeting of practitioners and theoreticians. The week-long format offers an opportunity for these communities to familiarize themselves with each other and establish a basis for collaboration.

IT security management is typically divided into the areas of confidentiality, integrity, and availability. Having the chance to gather experts from all these areas guarantees a comprehensive overview on the subject. The meeting would consist of representative background presentations to set the context for discussions, sessions for identifying ways of identifying work that can be leveraged across areas, sessions for developing joint research and development agendas going forward, and a capped off by a session that focuses joint problem-solving of a target issue selected during the week.

(Excerpt from the submitted proposal)

# The Future of Multimedia Analysis and Mining

# 016

Shonan Village Center, November 3–6, 2012

Seminar No.002

## Organizers

Shin'ichi Satoh, National Institute of Informatics, Japan

Nozha Boujema, INRIA, France

Alexander G. Hauptmann, Carnegie Mellon University, USA



Recent explosive growth of the amount of accessible multimedia information requires far more intelligent access to multimedia data. Multimedia analysis and mining play a key role to address this problem. For instance, multimedia analysis enables semantic access to multimedia information at any description level and for any applications or needs, even though the original multimedia data may not have any prior semantic annotation. Multimedia mining helps to provide high-level semantic and structural information to expose key information within a large-scale multimedia database. However, the development of such technologies is often severely limited due to the famous “Semantic Gap” in multimedia content analysis. This is well-known as a supremely difficult issue that is very hard to overcome.

On the other hand, researchers in this field now have access to far more computational resources thanks to recent developments in GPU use, multi-core technologies or the availability of cloud computing, as well as far more data resources thanks to the explosive growth of available multimedia data especially via Web. Several research projects have already begun to take advantage of these points independently.

In this meeting, we aim to discuss recent research trends and their impact on multimedia research. Then we consolidate key research challenges and explore promising new research directions, hopefully toward “bridging the semantic gap”.

Important topics include the following:

- Large scale issues (advances and benchmarking)
- Content Mining, search and retrieval
- “Deep understanding” vs “multimedia filtering”
- Social networks, collaborative tagging
- Profiling and recommendation
- Content vs context

# Quantitative Methods in Security and Safety Critical Applications

# 017



Shonan Village Center, November 9–12, 2012

Seminar No.017

## Organizers

Jin Song Dong, National University of Singapore, Singapore  
Annabelle McIver, Macquarie University, Australia

### The history and motivation for quantitative formal methods

Quantitative Formal Methods deal with systems whose behaviour of interest is more than the traditional Boolean “correct” or “incorrect” judgment. That includes timing (whether discrete, continuous or hybrid), as well as probabilistic aspects of success or failure including cost and reward, and quantified information flow.

The major challenge for researchers is to develop quantitative techniques that are both supple and relevant: the former is important because theories that amplify our reasoning powers are key to understanding system behaviour; the latter is important because our ultimate goal is to improve the practice of developing, deploying and certifying actual running software in the field.

### Quantitative modelling

The success of these case studies suggests that it is possible to apply quantitative modelling techniques to novel emerging technologies. The following provide topical examples.

1. High-Speed trains.
2. Quantitative security systems.
3. Wireless communications for mesh networks.
4. Smart electricity grid.

### Modelling and verification challenges

One of the major challenges in this area is that generic verification systems are difficult to apply.

Moreover some of the semantic issues are yet to be resolved; from the practical verification perspective there is a great deal of untapped potential both in recent theoretical and developing tool capabilities. Researchers have a better understanding of how to apply theoretical structures including categorical features, metric spaces and domain spaces which show how to take advantage of algorithmic techniques which similarly have been developing, but without necessarily with a particular semantics in mind. It is now within reach of researchers to bring together those advances to apply them to modern systems which rely on critical, quantified analysis.

### Meeting goals

The broad aims of this proposed meeting are to explore effective modelling and analysis methods with which to tackle the above challenging problems. Through a focused study of a selection of these case studies, the goal is to bring together existing and emerging techniques in semantics and algorithms, possibly combining them in new ways. In Asia there are a number of strong research teams already working in the area of quantitative semantics and verification tools. Our proposed invitees include many of those researchers as well as a number of key researchers from the United States and Europe.

### Scope, topics and proposed meeting format

The focus of this proposed meeting is to share and disseminate new (mathematical) semantic, modelling techniques and novel tool developments which are needed for modern system analysis.

(Excerpt from the submitted proposal)

# NII Shonan Configurable Computing Workshop

Shonan Village Center, November 12–15, 2012

Seminar No.025

## Organizers

Peter Athanas, Virginia Tech, USA

Brad Hutchings, Brigham Young University, USA

Kentaro Sano, Tohoku University, Japan



Configurable computing is an emerging technology proving to be capable of providing high computational performance on a diversity of applications, including 1-D and 2-D signal processing, simulation acceleration, computer graphics, and high-performance computing. High performance is achieved by rapidly reconfiguring the functionality and interconnectivity of the computing resources to match the computational requirements of specific applications. Rapid reconfiguration provides the illusion of having a much larger (virtual) hardware platform. With this approach, specific application properties, such as parallelism, execution profile, and data resolution can be exploited by creating custom operators, pipelines, and interconnection pathways.

In recent years a rapidly growing interest in using reconfigurable computing architectures for realizing and developing application-specific computer systems has been observed. The advances in reconfigurable technologies, in algorithms for implementation approaches and in automatic mapping methods of algorithms into hardware and processor spaces form a new computing paradigm of computing and programming, e.g. “Computing in Space AND in Time.” This requires different and novel approaches in engineering for developing reconfigurable systems and implementing complex algorithms, including theory, architecture structures, algorithms, design systems and industrial applications that demonstrate the benefits of this promising way of computing. The fast pace of development is not leaving industry enough time to develop the necessary theoretical foundation that underpins CAD tools, OS, designs, architectures and circuit technologies. Traditional hardware and software design processes and the tools to support them are not adequate for the design of run time reconfigurable systems.

Additional topics in this area of research include productivity and observability.

## Interdisciplinary Seminar

Researchers and practicing engineers alike need to be able to operate in interdisciplinary environments. Not only have boundaries largely disappeared between Computer Science and Electrical Engineering, for example, but we also see that the boundaries between computing and many other fields such as computational biology, chemistry, etc., are continuing to soften. Actual surveys in the EU Project MORPHEUS describe the increased request of the industry for multidisciplinary skilled engineers. The summary of the feedback can be reported like this that software engineers needs to understand new hardware paradigms (reconfigurable computing) and hardware engineers need to understand that software defines the product at the end. Supporting this trend, new workshops like the International Workshop on Reconfigurable Computing Education (RC-Education, <http://www.fpl.uni-kl.de/RCeducation08/>) and the AETHER-MORPHEUS Workshop-Autumn School (AMWAS, <http://www.alari.ch/AMWAS08/>) were established in order to bring together people from different research areas. But there is more activity required to bridge these gaps. Apart from the example described above, interdisciplinary skills needs to be introduced also for physicists, chip designers and hardware architecture specialists in order to master the challenges coming up with future complex electronic systems. This seminar will bring together researchers from industry and academics with an excellent reputation and the required wide base of disciplines which targets all areas of interest for future reconfigurable architectures.

(Excerpt from the submitted proposal)



# Social Issues in Computational Transportation Science

# 019

Shonan Village Center, December 17–20, 2012

Seminar No.022

## Organizers

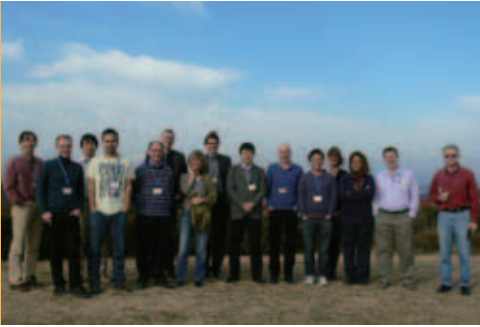
Ouri Wolfson, The University of Illinois, USA

Monika Sester, Institut für Kartographie und Geoinformatik, Leibniz Universität Hannover, Germany

Stephan Winter, The University of Melbourne, Australia

Glenn Geers, NICTA, Australia

Masaaki Tanizaki, Hitachi, Ltd., Kyoto University, Japan



## Overview

This application proposes a meeting bringing together researchers working in areas contributing to Computational Transportation Science (CTS). In the proposed seminar we plan to focus on the *social computing* aspect of CTS.

## Topics

In the seminar that is being proposed, we plan to focus on the direction of social computing. The researchers and practitioners from industry will review the development in this direction, discuss issues and solutions, and plan an edited book on CTS from the social computing perspective.

Furthermore, closer links to researchers and practitioners from ITS will be sought. In the following we provide more in-depth discussion of social computing in CTS.

From a state where every individual is acting autonomously in isolation, or with minimal (visual) interaction with their environment, it is quite a paradigm shift to think of transportation as an interconnected, communicating and cooperating complex system. Such a paradigm shift brings up research questions in multiple dimensions including, but not limited to:

- 1) *Managing Competition and Collaboration among Travelers*
- 2) *Crowd Sourcing*
- 3) *Behavioral Economics and Persuasive Technologies*

## Objectives and Expected Results

This NII Shonan Meeting has three objectives:

1. Discovering Social Computing in CTS
2. Growing a community in CTS
3. Seeking interaction and collaboration with scientists and practitioners from ITS

First, and more visibly, it will go into depth in one of the core CTS research areas: Social computing. Informal presentations of ongoing work in this area will inspire and help us to better shape and understand the bigger picture on social computing in CTS. We aim to develop ideas for a number of joint publications that cover this field and can be used to introduce in this field. Ideally we find a joint publication platform for an edited piece that may form the point of reference in the future. Secondly, by cross-disciplinary discussions we hope to shape a community. We will develop strategic ideas to nurture CTS such that we might have a vibrant annual conference with hundreds of participants in the future.

Such a seminar is needed for shaping new cooperation, and providing space and time for coordination and inspiration, and NII Shonan's infrastructure provides just the needed environment.

(Excerpt from the submitted proposal)

# Multi-Activity in Interaction: A Multimodal Perspective on the Complexity of Human Action

Shonan Village Center, February 18–20, 2013

Seminar No.030

## Organizers

Lorenza Mondada, Basel University, Switzerland

Mayumi Bono, National Institute of Informatics, Japan

Aug Nishizaka, Meiji-Gakuin University, Japan



The workshop intends to bring together researchers in ethnomethodology, conversation analysis, workplace studies and invite them to contribute to the study of multi-activity.

Multi-activity is a pervasive feature of contemporary work spaces: people are often engaged in more than one activity at a time, manage concurrent courses of action, overhear other conversations and phone calls while working, and pay attention at different events happening at the same time. These features characterize very different professional settings; they are particularly salient in computer supported work environments. People use different screens at the same time, engaging in different activities such as writing a report, writing emails, skyping, checking the news; they often phone while reading or writing at their PC; they engage in other activities while they continue to work on their computer, etc. These forms of multi-activity are spread over many types of professional settings—just to give a few examples, in call centers, call takers speak at the phone with customers and use their computer to record and search for information, as well as for dispatching the service asked for; in doctor-patient consultations, the physician both attends the patient and uses the computer to access the patient's file and write his report; in surgical theatres, the surgeon operates on a patient, discusses with colleagues about the last technologies or the insurance policies, while looking at a screen the endoscopic image supporting the procedure; in control rooms, professionals are constantly monitoring various screens while taking decisions and coordinating action at distance.

The workshop will produce significant analytical results contributing to our understanding of how complex activities are organized together in situ and in real time. This understanding is crucial for many fields; more specifically, in the field of technologies and informatics, these findings can significantly contribute to a) a better knowledge of users' practices and b) an integration of this knowledge into user-oriented design of interfaces. Concerning the first point, we know that computer users are most often engaged in more than one task while using their computer. They might be involved in a side conversation or activity; they might also be involved in another activity on the screen (like chatting, having a skype conversation, doing emails and navigating on the internet). This multi-activity affects the way in which they use technologies, computers, and softwares. Concerning the second point, and building upon the first, this knowledge of multi-activities might contribute to design specific interfaces, programs, and the ergonomics of the screen in such a way that it supports these dynamics. Currently, all tasks supported by the technologies are conceived as if an individual user would be using them in an exclusive way—which is not the case: users use them in interaction, in complex ecologies of action and within a multiplicity of other parallel actions. Thus, the workshop will improve our knowledge of these very mundane but very configuring practices, producing innovative insights for user-based technology design.

(Excerpt from the submitted proposal)

# Parameterized Complexity and the Understanding, Design, and Analysis of Heuristics

# 021



Shonan Village Center, May 6–11, 2013

Seminar No.018

## Organizers

Gregory Gutin, University of London, United Kingdom

Kazuo Iwama, Kyoto University, Japan

Dimitrios M. Thilikos, National and Kapodistrian University of Athens, Greece

## Short Summary

In the parameterized/multivariate framework, NP-hardness is just the beginning: a result about the null-parameterization. What follows is a rich dialogue between theory and practice, with the goals of:

- explaining the effectiveness of established heuristics

and

- designing better heuristics in mathematically disciplined ways

The vast majority of practical problems are NP-hard and thus we cannot hope to obtain algorithms that compute exact or optimal solutions efficiently, at least according to the one-dimensional ( $n$ , the input size) worst-case framework that essentially contrasts worst-case running times that are polynomial in  $n$ , versus worst-case running times that are exponential in  $n$ .

Recent discussions of the future of theoretical computer science have pointed to the importance of heuristic algorithms for NP-hard problems. In 2009, in his plenary talk at the FOCS Theory Day, Richard Karp posed the question, why do heuristics work so well in practice?, as one of the most important challenges in the theory of computing. Applied computing communities have been aware of this issue for decades, and have lamented the relative mission-failure of theoretical computer science. Theoretical computer science, has reciprocally lamented its lack of support, both by core funding agencies, and in Computer Science curriculums. There is a central dysfunction between theory and practice to which Karp points.

In real life, typical inputs to a computational problem are the outputs of other resource-constrained processes, both natural and artificial. While datasets may be large, significant (although often multifaceted) structure is almost always present and worth taking into account in algorithm design and complexity analysis.

Parameterized algorithms, both directly and indirectly, have made major contributions to practical computing in such areas as Computational Biology and (increasingly) Artificial Intelligence. The historical record is muddy. Since computing began, applications-determined implementors have of course attended to relevant secondary measurements that allowed their implementations to succeed on real-world data-sets (but without any attention to theoretical frameworks). Meanwhile, onedimensionally framed and trained theorists have crept towards a multivariate outlook, under the rhetoric of exact algorithmics.

(Excerpt from the submitted proposal)

# The Java Modeling Language (JML)

Shonan Village Center, May 13–16, 2013

Seminar No.016

## Organizers

Gary T. Leavens, University of Central Florida, USA

Jooyong Yi, National University of Singapore, Singapore

Peter H. Schmitt, Karlsruhe Institute of Technology, Germany



Program verification has been a topic of research interest far into the history of computing science. Today, it is still a key research focus, see e.g., Hoares Verified Compiler Grand Challenge and the Verified Software Initiative, whose flagship activities are the series of VSTTE workshops (Verified Software: Theory, Tools, and Experiments) and the launch of a series of verification competitions. A main facet in this effort is the ability to formally express properties that must be verified. Building on a long line of work in formal methods for reasoning about behavioral specifications of programs, several recent languages balance the desire for completeness and the pragmatics of checkability. In the context of the object-oriented programming paradigm, the Java Modeling Language (JML) is the most widely-adopted specification language in the Java formal methods research community.

The Java Modeling Language (JML) is a formal, behavioral specification language for Java. It describes detailed designs of Java classes and interfaces using pre- and postconditions, invariants, and several more advanced features. JML is used as a common language for many research projects and tools, including a runtime assertion checker (jmlc), tools to help unit testing (jmlunit), an extended static checker (ESC/Java), and several formal verification tools (e.g., LOOP, JACK, KRAKATOA, Jive, and KeY). JML is seeing some use in industry, particularly for financial applications on Java Smart cards and for verifying some security properties of a computer-based voting system.

Since JML is widely understood in the formal methods research community, it provides a shared notation for communicating and comparing many advances, both theoretical and practical, and it serves as a launching pad for research on advanced specification language features and tools. Researchers are using JML to study or express results for a wide variety of problems; these problems include verification logics, side effects (including frame axioms and modifies clauses), invariants, behavioral subtyping, null pointer dereferences, interfacing with theorem provers, information hiding, specifying call sequences in frameworks, multithreading, compilation, resource usage, and security. In addition to the tools mentioned above, JML is also used to express, compare, or study tools for checking specifications, unit testing, and specification inference. JML is used to state research problems for formal specification languages and for general discussions of specification language design. JML has also inspired at least three other similar specification languages, Spec#, BML, and Pipa, and has influenced the design and tools for Eiffel. Representatives of these communities are included in the invitation list. JML tools are used in the implementation of at least two other specification languages: ConGu and Circus. At present, there are at least 19 research groups around the world that are cooperating on JML-related research. These groups, and others, have published over 200 papers directly related to JML (see [www.jmlspecs.org/papers.shtml](http://www.jmlspecs.org/papers.shtml)).

(Excerpt from the submitted proposal)

# 023 Dimensionality and Scalability



Shonan Village Center, May 20–23, 2013

Seminar No.024

## Organizers

Michael E. Houle, National Institute of Informatics, Japan

Vincent Oria, New Jersey Institute of Technology, USA

Arthur Zimek, Ludwig-Maximilians-Universität München, Germany

## Background

For many fundamental operations in the areas of search and retrieval, data mining, machine learning, multimedia, recommendation systems, and bioinformatics, the efficiency and effectiveness of implementations depends crucially on the interplay between measures of data similarity and the features by which data objects are represented.

When the number of features (the data *dimensionality*) is high, similarity values tend to concentrate strongly about their means, a phenomenon commonly referred to as the curse of dimensionality. As the dimensionality increases, the discriminative ability of similarity measures diminishes to the point where methods that depend on them lose their effectiveness.

One fundamental task, arising in applications of multimedia, data mining and machine learning, and other disciplines, is that of content-based similarity search. For such applications, features are often sought so as to provide the best possible coverage across a range of anticipated queries. However, for any given query, only a relatively small number of features may turn out to be relevant. When the dimensionality is high, the errors introduced into similarity measurements by the many irrelevant feature attributes can completely overwhelm the contributions of the relevant features.

To support operations in such areas, feature selection and other dimensional reduction techniques from machine learning are often considered in an attempt to improve the discriminability of similarity measures, and the scalability of methods that depend upon them. Yet even here, the complexity of searching through combinations of features can be prohibitive.

## Dimensionality and data modeling

Many researchers and practitioners from different areas who are specifically working on problems involving the selection of features tend to be aware of the difficulties involved with high-dimensional data settings. Researchers in other areas are generally aware that the performance of their solutions depends on the dimensionality of their data sets, but are often not clear as to why.

## Objectives

The goal of this meeting is to bring together researchers and students active in the areas of databases, data mining, pattern recognition, machine learning, statistics, multimedia, bioinformatics, visualization, and algorithmics who are currently searching for effective and scalable solutions to problems affected by the curse of dimensionality.

Participants will not be expected to give presentations of fully completed research. Instead, emphasis will be placed on group input into the development of data characterizations, and the identification of future directions for research on dimensionality and scalability.

(Excerpt from the submitted proposal)

# 024 Privacy by Transparency for Data-Centric Services

Shonan Village Center, August 6–8, 2013

Seminar No.032

## Organizers

Isao Echizen, National Institute of Informatics, Japan  
Günter Müller, University of Freiburg, Germany  
Ryoichi Sasaki, Tokyo Denki University, Japan  
A Min Tjoa, Vienna University of Technology, Austria



The objective of the seminar is to close an existing “expressivity” gap between privacy requirements and policy compliance for data centric services by **transparency mechanisms**. The evolution of privacy and security mechanisms occurs in distinguishable steps following the progress of technology. Ongoing relevant IT-initiatives clearly show that transparency is the next most essential factor, which will play a prominent and major role in this evolution of privacy.

“Access control” is the most widely used metaphor used to model security. As a consequence authentication is the only source for protecting private data with the advantage of relatively easy control. In the early 90ties, data minimization at authentication time was the means to protect privacy. Data minimization generated a set of successful mechanisms, whose most well-known example are digital signatures, Public key Infrastructures, and identity management. In its most extreme case, anonymization totally omits data for authentication, while “Secure Multiparty Computing (SMC)” proposed 1983 by Dolev and Yao can be used to minimize personal data to the agreed limit. The reason for the disappointing acceptance of access control mechanisms for privacy in modern data centric services is characterized by Acquisti’s “privacy paradox.” This privacy paradox explains the gap between awareness and actual actions of users with regard to privacy enforcement. The reasons range from cumbersome and hard-to use mechanisms to lack of trust, but in essence it has been shown, that not the access but the usage of data is of concern. In 2004 Park and Sandhu’s specification of usage control model solved this gap. The mechanisms encompass privacy policy languages such as P3P and its Freiburg variant ExPDT. Sticky policies, secure logging, and data provenance are early examples of Transparency Enhancing Technologies (TET). With the advent of innovative business opportunities of “Big Data” it became obvious that transparent usage control is the most promising approach if privacy is to be controlled.

Transparency mechanisms are a better balance of interests of users and industry by offering “signaling and screening” functions. While “signaling” allows specifying privacy rules or policies under which services are to be conducted, “screening” encompasses all mechanisms to control the enforcement of the signaled rules.

The goal of this NII Shonan Meeting is to set **thematic milestones for the technical implementation of transparency on the one hand, and on the other, trace ways in which technical progress, users and industry could profit from transparency**. Specifically at the technical level, this NII Shonan Meeting (a) determines the current expressivity gaps between the privacy requirements and the policy specification languages, (b) compares existing mechanisms for testing adherence to privacy policies, and (c) identifies ways in which monitoring and audit could be combined into a “continuous auditing”. At the deployment level, this NII Shonan Meeting (a) lists tangible business models for transparency mechanisms (b) sketches guidelines on how to carry over transparency to different scenarios: e.g. Smart-Grids, e-Home, cyber-physical systems, and (c) categorizes the requirements for privacy sensitive and user-friendly design of dashboards. While the European organizers have strong background in (a) Privacy Enhancing Technologies (PET) and (b) in usage control, the Japanese organizers have internationally recognized competence in (c) digital rights, (d) critical infrastructures, and (e) public security, and the participants to be invited are composed of experts contributing to the technical objective as well as to the deployment issues.

(Excerpt from the submitted proposal)

# Discrete Algorithms Meet Machine Learning

# 025



Shonan Village Center, August 10–13, 2013

Seminar No.036

## Organizers

Hal Daumé III, University of Maryland, USA

Kevin Duh, Nara Institute of Science and Technology (NAIST), Japan

Samir Khuller, University of Maryland, USA

Structured prediction is the task of learning a function that maps inputs to structured outputs, and forms the heart of problems in natural language processing. Syntactic analysis, word alignment for machine translation, semantic understanding, action/goal recognition and translation itself are all examples of structured prediction tasks. In all these tasks, the goal at prediction time is to produce a combinatorial structure, typically by exploiting an off-the-shelf or hand-rolled combinatorial optimization algorithm. The fact that a combinatorial optimization algorithm consumes the output of a machine learning system in order to make predictions renders the learning problem difficult: one must reason about statements like “if I change my model in such a way, how will this affect the matching that the Hungarian algorithm (for weighted matchings) would return?”

Our goal is to solve such questions not only for specific problems such as graphs but also to develop a robust framework in the language of modern combinatorial algorithms by generalizing from spanning trees to matroids, or matchings to matroid intersections. All of these frameworks are polynomial-time for making predictions, and our goal is to construct similar polynomial-time learning procedures.

Although studying polynomial time solvable problems are useful for some specific natural language processing problems, we also wish to turn our attention to approximation algorithms for NP-complete prediction tasks (such as natural language generation or machine translation). So far, such problems have been the bane of structured prediction, requiring either the theoretically ungrounded use of approximate prediction within learning, or efficient but lower-quality search-based solutions with heavy pruning. The key idea we wish to pursue is that NP-complete problems are characterized by efficient verification. We propose to push this further to efficient separation at learning time (which appears to work for all polynomial-time algorithms we have looked at thus far). If this is possible, then we can explicitly train models to give correct solutions, even when a polynomial time approximation algorithm is run at test time. Such a result has the potential to change how the natural language processing community thinks about learning in computationally hard problems.

From the algorithmic perspective, for decades, our basic assumption has been that the input data is sacrosanct and essentially all basic optimization algorithms make this assumption. The problems we envision solving, now question this basic assumption, since we have to modify the input data, while minimizing the norm of the perturbation, so as to satisfy certain properties. We hope this meeting will unite the large sub-community in natural language processing community that deals with combinatorial prediction problems with the algorithms community that studies them.

# Engineering Adaptive Software Systems (EASSy)

Shonan Village Center, September 9–12 2013

Seminar No.027

## Organizers

John Mylopoulos, University of Trento, Italy  
Hausi Müller, University of Victoria, Canada  
Yijun Yu, Open University, United Kingdom  
Shinichi Honiden, National Institute of Informatics, Japan



As software-intensive systems continue to invade all aspects of personal, business and social life, they are required to operate in ever more open and dynamic environments where the one constant is uncertainty. Coping with such uncertainty calls for systems that monitor their environment and adapt so that they can continue to fulfill their requirements. The problem of engineering such systems is being addressed in a number of research communities, including Software Engineering, Systems, Ubiquitous Computing, Service-Oriented Computing, Multi-Agent Systems, Robotics and more.

As a result of research efforts within these communities, there have been many proposals on how to engineer such adaptive systems. Some are policy/requirements-based, others biologically-inspired, still others focus on awareness as the key facility that leads to adaptivity. The main objective of the proposed workshop is to bring together some of the authors of these proposals so that they can compare and contrast their respective approaches. In the process, we hope that participants in the workshop will go away with a better understanding of what ideas work, and under what circumstances. Some of the more specific issues to be discussed at the meeting include:

- How do we engineer adaptive software systems? What are the concepts, tools and techniques that can support requirements elicitation, architectural design and implementation of such systems?
- How do we reengineer legacy software systems in order to turn them into adaptive ones?
- Comparative review of adaptation mechanisms in Robotics, Multi-Agent Systems, Software Engineering, Socio-Technical Systems, Ubiquitous Computing, etc.
- Usability issues for adaptive software systems. How do we ensure effective human interaction with complex software systems that have adaptive components?
- Evolution of adaptive software systems. How do deployed adaptive systems evolve? How can we ensure convergence and stability for such systems?
- Evolution and control of systems-of-systems, where each component system has its own requirements and its own adaptation mechanism. How do we ensure convergence, stability and coherence for such systems-of-systems?
- Runtime models: most of the approaches to adaptivity are model-based in the sense that they deploy models of system requirements and the domain to support monitoring, diagnosis and compensation in the case of failure. How are such runtime models different from their design counterparts? How do we reason with runtime models to support adaptation functions, i. e., monitoring, diagnosis and compensation?
- How do we prevent failures for such systems through run-time reasoning mechanisms? Since such mechanisms are inherently intractable, how can we support incremental run-time reasoning that predicts and/or prevents failures?

As a deliverable for the workshop, we envision a follow-up special issue in an international journal, where workshop participants and the community-at-large will be invited to contribute original papers presenting comprehensive approaches to the engineering of adaptive software systems. The organizers hope that such a collection of papers will offer an authoritative account of the state-of-the-art on adaptive software systems and will guide future research in this fast-moving research field.

(Excerpt from the submitted proposal)



# Many-cores and On-chip Interconnects

# 027



Shonan Village Center, September 23–25, 2013

Seminar No.031

#### Organizers

Tomohiro Yoneda, National Institute of Informatics, Japan  
José Flich Cardo, Escuela Técnica Superior de Informática Aplicada, Spain  
Jiang Xu, The Hong Kong University of Science and Technology, Hong Kong  
Michihiro Koibuchi, National Institute of Informatics, Japan

Over the past few decades, a considerable number of studies have been conducted on the improvement and implementation of VLSIs. It has been recognized that the performance improvement of a single processor core is limited due to clock skew, power consumption, heat dissipation, leakage current, instruction level parallelism, and complexity. As a result, the rise of chip multiprocessor (CMP) and multi-processor system-on-a-chip (MPSoC) has rapidly been gaining pace, and they have become accepted as an integral part of the modern processing architecture. In these multiple core architectures, it has been recognized that a simple bus architecture does not scale with the system size as the bandwidth is shared by all the cores attached to it. Thus, the concern with on-chip networks has been growing as a feasible solution to many-core systems. Recently, it is also reported that fully asynchronous on-chip networks for NoCs have many advantages over the corresponding synchronous designs. On the other hand, as semiconductor process technology scales and on-chip networks become large, routers and links that compose on-chip networks should have tolerance against several kinds of faults. For example, even if one link or router goes down, the remaining part of the network should continue to work. Routers and links should adapt to performance degradation caused by effects like PMOS transistor negative bias temperature instability (NBTI), hot carrier degradation (HCI), VDD drop, temperature increase, and so on. Also, transient faults caused by soft-errors or noises should be tolerated. Furthermore, as more and more complicated applications are run on NoCs, the demand for on-chip networks with low-latency/high-throughput is increasing.

In this meeting, we would like to discuss the future direction of many-core and its on-chip interconnect technologies related to the above issues for supporting strong and weak scaling of parallel applications. Our technical interests include on-chip communication technologies, architectures, methods and applications, and asynchronous design for achieving low-power, high reliability, low-latency and high-throughput computing toward high-performance systems that include not only for High Performance Computing (HPC) but also embedded systems.

# 020 Computational Behavior Science

Shonan Village Center, September 26–28, 2013

Seminar No.039

## Organizers

Atsushi Nakazawa, Kyoto University, Japan  
James M Rehg, Georgia Institute of Technology, USA  
Yoichi Sato, The University of Tokyo, Japan  
Akihiro Sugimoto, National Institute of Informatics, Japan



## Objective

It is known that early diagnosis and treatments are very important for Autism Spectrum Disorders (ASD). However, finding the infant patients who have this disease is quite difficult and takes long-term observations by psychiatrists. In recent years, several projects have been started in U.S to solve this issue. These projects aim to develop a way to quickly diagnose the disorder by introducing digital technologies including image/video analysis, wearable sensing devices and speech recognition.

This workshop is held by Japanese researchers related to this field and U.S and international researchers who are mainly joining the NSF Expedition project “Computational Behavior Science (CBS)”, and includes the following topics:

- Digital visual behavioral analysis techniques for infants, in particular, facial expression, eye gaze and motion analyses.
- Behavior analyses using wearable sensing devices and speech analysis.
- Lectures about ASD by psychologists and caseworkers.

## BENEFITS

Studies about ASD have been conducted in the psychologists’ community and Autistic Spectrum Society; however, there are few studies introducing digital analysis techniques. Therefore, this workshop is unique in the sense that it is held by computer scientists and focuses on introducing digital analysis techniques for this task.

Through this workshop, we potentially have the following benefits:

- Develop/share knowledge about the studies related to digital technology for finding an infant’s ASD.
- Create domestic/international interpersonal relationships in this field.
- Create international projects to develop the techniques for this issue, including eye gaze tracking and behavioral analysis based on the RAPID ABC video database developed by NSF CBS group.

# Compact Data Structures for Big Data

# 029



Shonan Village Center, September 27–30, 2013

Seminar No.029

## Organizers

Kunihiko Sadakane, National Institute of Informatics, Japan  
Wing-Kin Sung, National University of Singapore, Singapore

Big Data are structured and unstructured datasets whose size is in the order of billions or trillions. Because of their diversity and size, it is difficult to store, search and analyze them. This meeting therefore focuses on algorithms and data structures for efficient manipulation of Big Data. Especially, the meeting is devoted to compact data structures for managing Big Data.

Typical examples of big data are genomic sequences and gene expression data, Web and SNS data, sensor data in intelligent transport systems, etc. Traditional data structures do not scale to handle such data, and therefore we should design new data structures to handle them.

Although the amount of data explodes, the amount of the underlying information inside the data may not be exploded. It is observed that many big datasets are redundant. In the Web, many webpages were copies of others. In global positioning system (GPS), GPS position data change continuously, which can be compressed using differential encoding. In genomics, although different individuals have different genomes, the individual genomes have highly similarity. Therefore we can compress such data by identifying the similar parts. After the data is compressed, other issues are how to access and search them efficiently. Traditional data structures are not designed to handle compressed data and they may not manipulate Big Data well because the size of the data structures exceeds the limit of memory usage, or searching time increases due to their size. To handle these problems, researchers have worked on developing compact data structures. Such data structures are also called compressed or succinct data structures. They are much smaller than standard data structures, while keeping the same access time to data in theory. However actual performance of such compact data structures for storing Big Data is unknown or unsatisfactory.

The aim of this workshop is to bring together researchers active in the areas of compact data structures to exchange ideas for handling Big Data. We will discuss methods for compressing and storing Big Data. We will also discuss how to design time- and space-efficient data structures for them through discussion and sharing knowledge, we hope to promote collaborations and further improve data structures for Big Data.

# Coinduction for Computation Structures and Programming Languages

# 030

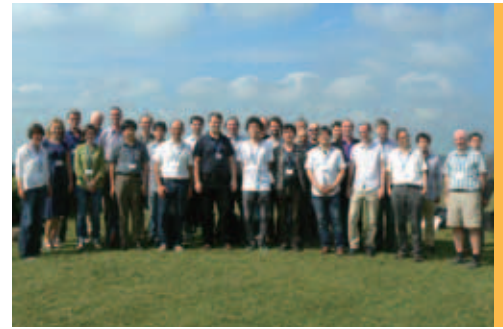
Shonan Village Center, October 7–10, 2013

Seminar No.026

## Organizers

Ichiro Hasuo, The University of Tokyo, Japan

Tarmo Uustalu, Institute of Cybernetics at TUT, Estonia



This meeting has as its aim to study the applications of coinduction (coinductive data/predicates, bisimilarity, corecursion and coinduction) to reasoning about computation and programs, hence in programming language semantics, program logics. This is motivated by the appropriateness of coinduction for analyzing infinity, so also infinity in computation structures, the prime example being infinitely running computations, e.g., of machines or programs, especially in the context of reactive computation.

Induction and coinduction are, by themselves, dual notions. But in typical non-self-dual settings of actual interest they come out as quite different. The most important outcome of the asymmetries is that induction is about finite construction and infinite use, but coinduction is about infinite construction and finite use of data.

Lately there is a growing interest in coinduction in the areas of programming semantics and formal verification of software systems, in part thanks to advances in type-theoretical programming languages and proof assistants. There are many important and practical examples where one needs to reason about finitely observable infinite computations and replacing this by, e.g., inductive reasoning about finite initial fragments thereof is unnatural or inadequate. For instance, one may want to prove that some program transformations do not change the observational behavior of possibly nonterminating transformational programs or of infinitely running reactive programs.

Although coinduction should be very useful, in reality it is not really well understood. In fact, coinduction is surrounded by quite some popular confusions and has become a standard tool only in concurrency theory and coalgebra. The theory of coinduction is on many occasions subtle and challenging. It is often difficult to translate between the vocabularies developed in different disciplines (e.g., coalgebra, proof theory, type theory). The applications are sometimes not supported well by tools. For example, it is by no means clear how to best support coinductive data and predicates in type-theoretical programming languages and proof assistants. Unfortunately, the mechanisms offered by the current systems are weak and/or cumbersome. As a result, the corresponding applications are underdeveloped.

Of interest for the meeting are at least the following topics:

- coinductive computation structures, in particular for possibly infinite behaviors (of machines, programs etc), interaction, concurrency,
- coinductive program semantics, type systems, program logics,
- computability for coinductive data,
- theory of coinductive types, bisimilarity, corecursion/coinduction in type theory, proof theory, category theory,
- support for coinductive types in dependently typed programming languages, type-theoretic proof assistants.

Coinduction is applied by researchers from different communities, with diverse technical backgrounds. There are multiple approaches to coinduction, it is used in different applications. The meeting encourages exchanges between researchers representing different areas and communities: programming languages design, implementation, semantics, functional programming, theory of concurrency, category theory (coalgebra), proof theory, type theory.

# Software Analytics: Principles and Practice

# 031



Shonan Village Center, October 21 –25, 2013

Seminar No.037

## Organizers

Ahmed E. Hassan, Queen's University, BlackBerry, Canada  
Katsuro Inoue, Osaka University, Japan  
Tao Xie, University of Illinois at Urbana-Champaign, USA  
Dongmei Zhang, Microsoft Research Asia, China

A wealth of various data (e.g., source change history, test cases, and bug reports) exists in the practice of software development. Further modern software and services in operation produce rich data (e.g., operation logs, field crashes, and support calls). Hidden in these unexplored data is rich and valuable information about the quality of software and services and the dynamics of software development. Companies (Microsoft, Google, Facebook, Cisco, Yahoo, IBM, RIM, etc.) are increasingly adding analytics as an important role in their organizations, leveraging the wealth of various data produced around their software or services.

Software analytics is concerned with the use of data-driven approaches to obtain insightful and actionable information for completing various tasks around software systems, software users, and software development process. Insightful information is information that conveys meaningful and useful understanding or knowledge. Actionable information is information upon which software practitioners can come up with concrete solutions (better than existing solutions if any) towards completing tasks. Typically such information cannot be easily obtained by direct investigation on the raw data without the aid of analytic technologies.

Especially recently the area of Big Data has emerged as a critical and strategic focus by the society. Big data is everywhere now but it is still under-utilized in the area of software engineering. However, leveraging big data is very relevant in software engineering as software and services get larger and more inter-connected, often being developed by a large number of engineers in distributed fashions and being used by a huge number of users around the world. Software analytics needs to be prepared for the upcoming decade's exciting and yet challenging problem of leveraging big data for software engineering tasks.

Software analytics is an ideal topic for this kind of interaction. It combines challenging research problems with real practical importance for the software industry, and the wider society that it serves. It presents an excellent and wide-ranging set of open research questions to academics concerning, amongst other things, analytic-algorithm design, data analysis, information visualization, scalable computing, software-artifact analysis and mining, social factors, empirical software engineering, measurement, process improvement, and technology transfer and adoption. Software analytics is also of critical practical significance to almost every organization involved in the production and use of software and services. Answers to the currently open research questions in software analytics can have a major impact upon industrial practice, with far-reaching implications for the development of the global economy. This combination of academic challenge and industrial relevance makes software analytics a natural topic for the proposed seminar.

(Excerpt from the submitted proposal)

# 032 Implicit Computational Complexity and Applications: Resource Control, Security, Real-Number Computation

Shonan Village Center, November 4–7, 2013

Seminar No.033

## Organizers

Akitoshi Kawamura, The University of Tokyo, Japan

Jean-Yves Marion, CNRS, The University of Tokyo, France

David Nowak, CNRS, The University of Tokyo, France&Japan



Computational complexity theory aims at classifying computational problems according to their inherent difficulty. The standard way to achieve this classification consists in formalizing a precise execution model (e.g., a Turing machine) and posing explicit bounds on time and memory resources. On the other hand, Implicit Computational Complexity (ICC) aims at studying computational complexity without referring to external measuring conditions or a particular machine model, but only by considering language restrictions or logical/computational principles implying complexity properties. The area of ICC has grown out from several proposals to use logic and formal methods to provide languages for complexity-bounded computation (e.g., polynomial time, logarithmic space computation). ICC methods include, among others, linear logic, typed programming language, second order logic, term ordering. The last decades have seen the development of logical formalisms that characterize functions computable in various complexity classes (polynomial or elementary in time, logarithmic in space).

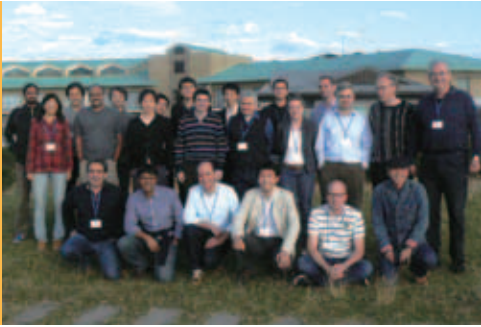
The goal of the proposed meeting is to explore foundational as well as practical interconnections between formal logic and computational complexity, such as it is done in ICC. The main outcome of this meeting will be to trigger new interactions and enrich the various approaches. In particular, and aside from traditional ICC approaches, we would like to focus on computation involving real numbers and topological spaces, thereby providing a deeper understanding of computational complexity in non-discrete realms of mathematics. By bringing together experts in implicit complexity and in complexity in analysis, we will promote new interaction between the two fields. People in those two fields are currently working separately, but there is enough common ground between them to make it worth having those two communities talking and working together. The meeting would also foster discussions about applications, i.e., the design of methods based on ICC and suitable for static verification of program resource consumption and of security.

Very recently ICC methods have been applied to security methods and conversely security methods have been used as a new approach in ICC context. In the context of security proofs, the computational power of adversaries has to be limited so that their potential attacks are feasible. An adversary with unlimited computational power could indeed break most cryptographic schemes (e.g., RSA by efficiently factoring large integers). It is usual to rely on Cobham's thesis identifying feasibility with computability in polynomial time. Hence the particular interest in the class of functions computable in polynomial time and its implicit characterization with a programming language that can be used to construct adversaries. Conversely, type systems to control the information flow, which are traditionally used for certifying security policies like confidentiality or integrity, are related to the notion of data stratification. As a result, type systems for imperative programming languages have been developed to control resource consumption.

The other proposed focus is on computation over the reals. Computable analysis, the study of abilities and limitations of digital computers applied to problems in mathematical analysis, has originally evolved from computability theory, but there is increasing interest in computational complexity with bounded time and space. The goals here are to analyze the computational costs of algorithms for problems involving real numbers and to explore the principles and structures of computational complexity in this context, providing a foundation of validated numerical methods for problems arising in physical sciences and engineering. Broader perspectives in computational complexity, including those from implicit complexity theory, have high potential to help here, as can be already seen, for example, in recent studies of computational power of dynamical systems and analog computers, or in the application of type-two complexity theory to time-bounded computable analysis.  
(Excerpt from the submitted proposal)

# Cognitive Social Robotics: Intelligence Based on Embodied Experience and Social Interaction

# 038



Shonan Village Center, November 11 – 14, 2013

Seminar No.038

## Organizers

Tetsunari Inamura, National Institute of Informatics, Japan

Tamim Asfour, Karlsruhe Institute of Technology, Germany

Sethu Vijayakumar, University of Edinburgh, United Kingdom

Understanding mechanisms underlying intelligence of human beings and animals is one of the key approaches towards developing intelligent robot systems. Since the mechanisms of such real-life intelligent systems are so complex, physical interactions between agents and their environment and the social interactions between agents should be considered. On the other hand, the interactions between robots, autonomous systems, their environments and people at various scales present some of the most sophisticated scientific challenges we must solve to realize the next generation of robots. Comprehension and knowledge in many related fields such as cognitive science, developmental psychology, brain science, evolutionary biology, and robot engineering is also required—making a case for strong interdisciplinary interaction of minds for implementing this approach.

In this decade, an academic field named **cognitive developmental robotics** has been formalised with these aims. This approach focuses on embodied intelligence which is one of the deficient points in GOFAI (Good Old Fashioned Artificial Intelligence). While huge strides are being made in cognitive robotics, the quality and performance of the robots' intelligence is yet to match the versatility of human counterparts.

One of the reasons for this deficiency is that perhaps we do not yet fully exploit the interactions and embodiment, both in the physical and social (and virtual) domain. One of the difficulties is huge cost of collecting embodied and social experience due to limitation of robots' operating time, robustness and limited opportunity of social interaction with humans due to limits in deployment. A promising direction is to accelerate the development of intelligence from large-scale long-term, persistent embodied experience we term 'big social experiences'—addressing many of the issues analogous to the 'Big Data' challenges in Informatics.

This Shonan meeting focuses on synthetic research in cognitive social robots, building on the premise that intelligence develops based on physical embodied interaction between body and environment, social interaction between agents and human, with the aim to address both the following aspects:

- (a) **Exploiting Physical Embodiment and Interaction:** Actuators, Control, Multi-contact Planning, Natural Dynamics and Design
- (b) **Networked Social Intelligence:** Networked interactions, Multi-agent systems, Software for connectivity, Emergent Capabilities.

To foster this, we aim for interdisciplinary discussions with wide viewpoint from various research fields such as cognitive science, developmental psychology, brain science, and not just robotics. The scope of this meeting will include human-robot interaction, machine learning, cognitive science, simulation technology for long-term large scale interaction, learning by demonstration, human biomechanics, multimodal sensory experience (auditory, speech, gestures). One of the target applications as material for the discussion would be acquisition of knowledge and skills from natural interaction between embodied agents and human beings; however, this meeting would like to touch on other future big challenges related to the cognitive social robotics with a broad perspective.

# 034 Intelligent Information Processing — Chances of Crowdsourcing

Shonan Village Center, November 18–21, 2013

Seminar No.034

## Organizers

Wolf-Tilo Balke, Technische Universität Braunschweig, Germany  
Seung-won Hwang, Pohang University of Science and Technology, South Korea  
Takahiro Hara, University of Osaka, Japan  
Christoph Lofi, National Institute of Informatics, Japan



Currently a variety of platforms like Amazon's Mechanical Turk, CrowdFlower, or Samasource are offering frameworks with different degrees of sophistication where (usually relatively simple) cognitive tasks can be dynamically posed to a large and readily available workforce. This ability of cheaply distributing simple jobs via the Web allows for new modes of labor and information processing. In fact, the "knowledge society" has already brought severe changes to business processes in today's economy. This is especially true for the basic question of what and where people work.

Here the ubiquity of sophisticated mobile devices and communication services allow for almost unlimited flexibility and freedom in negotiating and outsourcing short-term work contracts and delivering results. Currently, mobile crowdsourcing by smartphone users is a hot research area. In any case, in the industrialized world there is a clear transition from traditional production of goods or processing of raw materials towards the provisioning of services and the flexibility with respect to the place where such services are actually physically provided has dramatically increased. Still, although services could in principle be offered flexibly from virtually anywhere in the world, typical constraints like the local cost of labor or easy access to an educated workforce, remain valid. Crowd-Sourcing promises to break with these traditional work models, by offering a dynamic global information-processing workforce which is available 24/7 with close to no overhead. This shift paves the way for approaching large-scale information task which were previously infeasible for both algorithmic and traditional human-based approaches.

*The central challenge in the current knowledge society is to efficiently and intelligently deal with an overwhelming amount of information, a daunting task for computer systems and humans alike. To this aim the data management and data mining communities consider a wide variety of operators, algorithms, and workflows.*

For some information-heavy areas like for example customer relationship management, where everyday services like ordering procedures, customer data management, complaint handling, etc. have to be performed, outsourcing the work to specialized workers has become a commonly accepted solution for increasing efficiency. On an educationally higher level, business intelligence services can serve as a good example: extracting relevant information from company data and using it to recognize or design value-adding areas like new products, promising customer segments, or better business processes for a company is a profitable business. Indeed 'infopreneur' is a term coined for the growing number of persons whose primary business is gathering and selling electronic information. However, this current form of out-sourcing information-centric tasks is still quite static (i.e. a fixed team of specialists is contracted for a larger task). In contrast, crowdsourcing as understood in this proposal dynamically assigns small intelligence tasks to workers from a large pool in a demand-driven fashion. The advantages are obvious: if at creation time each process can be effectively broken down to manageable tasks and a viable time plan, it can be fulfilled very efficiently. The main factor is elasticity: peaks and slumps in activity can be dynamically handled and missing expertise or competences can be contracted. Thus, the efficiency of the overall process is hard to beat.

(Excerpt from the submitted proposal)



# Computer Visualization — Concepts and Challenges

# 035



Shonan Village Center, March 10–13, 2014

Seminar No.046

## Organizers

Arie E. Kaufman, Stony Brook University, USA

Issei Fujishiro, Keio University, Japan

Computer visualization is a very active research area of computer science due to its usefulness proven in diversified applications. IEEE Computer Society Visualization and Graphics Technical Committee (VGTC) has served as the sponsor for three major international forums on computer visualization, emphasizing Scientific Visualization (SciVis) and Information Visualization (InfoVis) and more recently including Visual Analytics. The Visualization Conference, held annually in the US, started in 1990, and has continued to grow in its attendance and technical program. In Europe, co-sponsored by Eurographics, the Visualization Symposium, EuroVis, is being held annually since 1999, and still continues its rapid growth. In the Asia-Pacific region, the Pacific Visualization Symposium, PacificVis, had its debut in Kyoto in 2008, and is held annually since then with a particular aim to draw more related researchers in the Asia-Pacific region.

After the six successful meetings, PacificVis 2014 will come back to Japan. As the venue for the symposium, selected was Keio University, Yokohama, which is very close to Hayama. We would like to hold an NII Shonan Meeting on computer visualization immediately after PacificVis2014 (March 4–7), in order to invite senior researchers participating in the symposium for continuing intensive discussions on the future of the research area, as well as to encourage young researchers and Ph.D. candidates from Asian countries for exchanging their latest ideas with these senior participants.

In this meeting, we will place our particular focus on specific cutting-edge sub-topics of SciVis and InfoVis, which include biomedical visualization; uncertainty visualization; mapping visualization; multi-scale visualization; and multi-dimensional/ivariate visualization, and have intensive discussions to expand their concepts and to forecast the promising perspectives of their research and development. In order to focus the discussions and the group, we will not include Visual Analytics in this meeting.

# Towards the Ground Truth: Exact Algorithms for Bioinformatics Research

# 036

Shonan Village Center, March 17–20, 2014

Seminar No.045

## Organizers

Sebastian Böcker, Friedrich Schiller University Jena, Germany  
Gunnar W. Klau, Centrum Wiskunde & Informatica (CWI), The Netherlands  
Hon Wai Leong, National University of Singapore, Singapore



Today, bioinformatics has become an integral and indispensable part of life science research: Success stories include the assembly and deciphering of genomes, understanding the complexity of cellular processes by means of biological networks, recovering the “tree of life”, and deciding on treatment plans for HIV or cancer patients. Applications range from fundamental questions such as the origin of life to multi-billion dollar decisions on novel drug leads and molecular modeling. None of these questions could be approached without massive support from bioinformatics.

Many of the core challenges in bioinformatics can be described as combinatorial optimization problems. Examples are the identification of genes and regulatory structures within genomes; discovering genomic or transcriptomic variations; mining biological networks for, say, protein-protein interactions; or, establishing the evolutionary history of organisms, to name just a few. Unfortunately, a large fraction—and arguably the majority—of these problems are NP-hard: Prominent problems are Multiple Sequence Alignment or Maximum Parsimony in phylogenetics, but there are many more—the query “bioinformatics NP-hard” yields over 10,000 hits in Google Scholar.

It is common practice in bioinformatics to approach these NP-hard problems using heuristics. Although the mathematical model provides only an imperfect approximation to the true goal, namely, to discover nature’s ground truth, finding optimal solutions is indispensable to rigorously evaluate the quality of the model. Heuristics and approximation algorithms are useless for this purpose, for which exact algorithms are needed. Furthermore, good exact algorithms provide deep insight in the structure of the underlying combinatorial problem, which leads to a better understanding of what exactly makes the biological question hard to solve.

In particular, modern measurement techniques such as high-throughput sequencing provide such direct access to the biological ground truth, so that problem modeling can be focused to reverse engineer the biotechnology protocol. While the combinatorial modeling of, for example, assembly problems related to sequencing typically lead to NP-hard problems, the dramatic decrease in sequencing costs also enables multiplexing divide-and-conquer approaches such that inputs to each problem instance become smaller. In these scenarios, exact algorithms for hard problems can be feasible both from the computational and economical perspective.

Recently, there has been much progress on solving combinatorial problems in bioinformatics to provable optimality, despite their hardness. Different techniques have contributed to this progress: in particular, Integer Linear Programming, data reduction and kernelization, and fixed-parameter algorithms. In addition, Algorithm Engineering techniques, which exploit the fact that the structure in realistic problem instances often deviates from the worst case scenario, have contributed to the success of many exact approaches. In contrast, classical exponential-time algorithms such as exhaustive search or higher-dimensional Dynamic Programming have played a negligible role in bioinformatics research.

The aim of this workshop is to bring together researchers active in exact approaches for combinatorial bioinformatics problems. We want to tackle the difficult issues these problems pose, and to exchange ideas and theoretical frameworks that allow the design and implementation of algorithms and methods for their solution. Researchers in this workshop come from different areas of algorithmics, such as kernelization and Integer Linear Programming; assembling their views and ideas will foster the applicability of exact algorithms in bioinformatics. Through discussion and sharing knowledge, we will promote collaborations, contribute to the progress in this growing field, and make the field more visible for other scientists.

# Grid and Cloud Security: A Confluence

# 037



Shonan Village Center, March 24-27, 2014

Seminar No.044

## Organizers

Barton P. Miller, University of Wisconsin, USA  
Yoshio Tanaka, National Institute of Advanced Industrial Science and Technology (AIST), Japan  
Elisa Heymann, Universitat Aut3noma de Barcelona, Spain

The security of Grid and Cloud computing environments is critical to today's cyber-infrastructure. The goal of the second edition of this seminar is to bring together a diverse community of researchers, practitioners, and developers, to leverage knowledge that spans the areas of Grid and Cloud security, industry and government and academia, theoretical and practical interests, and the scientific and business communities. This Shonan meeting will continue the Asia-US-Europe collaboration on security we started with the first edition of this seminar, and strengthen the efforts involving academia, industry, and government to bridge the above mentioned areas. In addition, we will add involvement from Australia and South America. Our first Shonan seminar was a success, and we want to expand on that success, taking advantage unique week-long format to add working group sessions to the seminar.

## Topics of the Seminar

In the past several years, Grids and Clouds have emerged as particularly productive environments for leveraging the technology of the Internet: Computational Grids and Cloud Computing.

In each of these two environments, unrelated users are accessing common computational, network, and storage resources, and organizations are extending their trust boundaries well beyond the traditional physical limitations of their own facilities.

The software that comprises the services provided by each of these environments is complex and multifaceted, and not well understood by the typical user. As a result, the threats that seem obvious to a user may have an actual risk that is quite minimal, while there are significant and emerging threats of which users are completely unaware.

## Goals of the Seminar

The seminar is to answer the following questions among others:

- A meeting of experts from the Grids and Clouds area:
  - Grid computing and its predecessors (distributed computing, metacomputing, . . . ) has long had to address security issues. How can we leverage these lessons and apply them to Clouds?
  - What solutions are coming from Cloud computing that can be applied to Grids?
  - What new challenges come as Grids extend into the Clouds, using such facilities as dynamically provisioning Grids from Cloud resources?
- A meeting of practitioners and theoreticians: Security meetings and projects tend to have either a strong theoretical or practical bent.
  - What recent theoretical results in security can improve our way of securing Grids and Clouds?
  - What are the most current pressing practical problems in Grids and Clouds for which we would like to see new algorithms and techniques?

IT security management is typically divided into the areas of confidentiality, integrity, and availability. Having the chance to gather experts from all these areas guarantees a comprehensive overview on the subject. During the week-long session, we will provide an opportunity to share each other's experiences, leverage each other's knowledge, and develop a joint strategy for securing these critical computational and informational resources as Grids and Clouds do offer.

The meeting would consist of representative background presentations to set the context for discussions, and sessions for working groups with common interests. Working group sessions will allow us to spend the extended time needed understand specific problems in depth, develop approaches to their solution, and form stronger international and interdisciplinary collaborations.

(Excerpt from the submitted proposal)

# Knot Theory: Algorithms, Complexity and Computation

# 038

Shonan Village Center, April 28–May 1, 2014

Seminar No.043

## Organizers

Ryan Budney, University of Victoria, Canada

Benjamin Burton, The University of Queensland, Australia

Kazuhiro Ichihara, Nihon University, Japan



Knot theory is, at its most basic level, concerned with the topology of closed loops in 3-dimensional space. This mathematical subject is inherently algorithmic: old and fundamental questions include how to test whether two knots are topologically equivalent (isotopic), or how to enumerate all distinct knots up to a given complexity.

Great progress has been made on these problems mathematically, from Haken's solution to the unknottedness problem half a century ago, to the Gordon-Luecke theorem which converts the broader knot equivalence problem into an algorithmic problem on triangulated 3-manifolds.

Nevertheless, algorithmic problems on knots remain challenging for those who wish to do real computations. For instance, the best-known algorithms for just the "simple" problem of testing unknottedness are exponential-time, and a full enumeration of prime knots has only been carried out for knots of  $\leq 16$  crossings. Such problems have now attracted the attention of researchers in algorithms and complexity, as well as other branches of computer science. Examples of current questions include:

- *How do algorithmic questions from knot theory fit into the hierarchy of complexity classes?*  
Testing unknottedness is known to be NP, and also co-NP assuming the generalized Riemann hypothesis. However, is it in P? Is it NP-complete? Is it fixed-parameter tractable for some natural parameter?
- *What are the best methods and heuristics for solving knot problems in practical software?*  
There are many invariants, geometric methods and heuristic techniques that are extremely effective in practice but do not guarantee a conclusive solution—can we prove that these work for average or generic inputs? For a conclusive solution, methods from operations research have proven remarkably effective in testing unknottedness, but can these be generalised?
- *How can we efficiently generate and effectively manage databases of knots and their properties?*  
The excellent online sources such as KnotInfo and the Knot Atlas are still relatively small (e.g., the exhaustive KnotInfo database contains only the first 2977 prime knots), and generating new data will require a careful interplay between large-scale combinatorial enumeration and complex topological decision problems. As databases grow in scale we must also address issues of computing complex properties in bulk, and effectively exploring and mining the resulting data.

The purpose of this meeting is to bring together experts in computational and algorithmic knot theory. Speakers will range across the spectrum from pure mathematicians to theoretical computer scientists and practical software developers. In particular, we will draw on expertise from a range of inter-related areas:

- development of mathematical software;
- generation and management of data collections;
- computational complexity of knot problems;
- numerical algorithms for studying the geometries of knots;
- discrete algorithms in computational geometry and integer programming;
- visualisation of knots;
- geometric topology, including the tightly related area of 3-manifold topology.

By combining these areas of expertise, we aim to generate new ideas and spawn new long-term projects that can create a clear theoretical picture of the intrinsic algorithmic difficulty of problems in knot theory, and set a new standard for what software can achieve in this rich and complex problem domain.

(Excerpt from the submitted proposal)

# Deep Learning: Theory, Algorithms, and Applications

# 039



Shonan Village Center, May 19–22, 2014

Seminar No.040

## Organizers

Pierre Baldi, University of California, Irvine, USA  
Kenji Fukumizu, Institute of Statistical Mathematics, Japan  
Tomaso Poggio, Massachusetts Institute of Technology, USA

The ability to learn is essential to the survival and robustness of biological systems. There is also growing evidence that learning is essential to build robust artificial intelligent systems and solve complex problems in most application domains. Indeed, one of the success stories in computer science over the past three decades has been the emergence of machine learning and data mining algorithms as tools for solving large-scale problems in a variety of domains such as text analysis, computer vision, speech recognition, robotics, and bioinformatics. However, we are still far from having a complete understanding of machine learning and its role in AI, and plenty of challenges, both theoretical and practical, remain to be addressed.

Complex problems cannot be solved in one single step and often require multiple processing stages in both natural and artificial systems. For instance, visual recognition in humans is not an instantaneous process and requires activation of a hierarchy of processing stages and pathways. The same is true for all the best performing computer vision systems available today. Thus deep learning architectures, comprising multiple, adaptable, processing layers are important for the understanding and design of intelligent systems and, today, are at the forefront of machine learning research. In the past year alone, deep architectures and deep learning have achieved state-of-the-art performance in many application areas.

It is this recent wave of progress that provides the relevant context for the proposed meeting which will focus on all aspects of deep architectures and deep learning, but with a particular emphasis on understanding fundamental principles because there is still very little theoretical understanding of deep learning, in spite of the recent progress. Thus a major thrust of the meeting will be to foster theoretical analyses of deep learning. In addition to theory, topics to be covered will include also algorithms and applications. While, the primary intellectual focus of the meeting will be on deep learning in artificial systems, presentations and discussions bridging learning in natural and artificial learning systems will also be encouraged.

# 040 Software Contracts for Communication, Monitoring, and Security

Shonan Village Center, May 26–30, 2014

Seminar No.050

## Organizers

Atsushi Igarashi, Kyoto University, Japan

Peter Thiemann, University of Freiburg, Germany

Philip Wadler, University of Edinburgh, United Kingdom



The contract metaphor is the basis of today's software verification. Expressing invariants, pre-, and postconditions with logical formulae as part of an interface has become common fare in the construction of dependable software. Often, general logical formulae are replaced by type qualifiers or other abstractions to soften the learning curve.

A wealth of recent work has considered behavioral contracts like session types that describe the temporal behavior of program components and/or generalize the notion of the observed state. Often this kind of contract describes the communication behavior of a group of components, but there are many other use cases like tpestate reasoning or the description of security policies.

We observed that there are different communities (contracts, sessions, security and typing, specification, and static vs. dynamic verification) that use similar approaches to pursue similar research objectives. Hence, the goal of the seminar is to explore the potential for crossfertilization, to identify commonalities and differences, to get inspiration by discussing the techniques used in the other fields, and to find grounds for future collaboration.

## Session Types, Tpestate, and Monitoring

Numerous variations of session calculi have been developed. Some are based on functional and object-oriented calculi. Others include subtyping, asynchronous communication, or they govern multi-party communications. Recent work has established connections with linear logic as a semantic basis for session types and communication automata, which are widely used in the analysis of network protocols.

A closely related line of work considers behavioral types and tpestate, where the type of an object changes over time to indicate its current internal state. Methods of the object may be enabled or disabled depending on its current type.

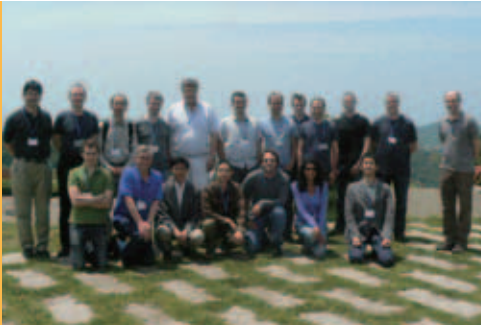
On the other hand, researchers in computer security have considered application monitoring for a long time. A monitor runs alongside the application program and changes state triggered by a stream of events from the running program. If the monitor is used for auditing, then the event stream is condensed to a log and saved for post-mortem analysis. However, the monitor may also be used to enforce a security policy by terminating the program as soon as it attempts to perform an illegal sequence of events.

Recent work indicates that session types and tpestate are converging, but the situation is less clear for the interaction between session typing and security monitoring. In particular, session typing is mostly considered in a static typing context whereas security monitoring usually happens at run time. There may even be room for hybrid approaches that apply a mix of static and dynamic checks, in close analogy to existing work in dynamic typing.

(Excerpt from the submitted proposal)

# Staging and High-Performance Computing: Theory and Practice

# 041



Shonan Village Center, May 27–30, 2014

Seminar No.056

## Organizers

Oleg Kiselyov, Monterey, CA, USA

Jeremy Siek, Indiana University, USA

Yukiyoshi Kameyama, University of Tsukuba, Japan

We propose the follow-up to the meeting “Bridging the theory of staged programming languages and the practice of high-performance computing” (Shonan Seminar 19), which took place in May 2012. In the first meeting, researchers in staged programming languages were learning of the problems in high-performance computing (HPC) from HPC experts. Since then, thanks in part to Shonan Challenges put forward at that meeting, the theory and tools of staged programming languages have progressed to the point where they could already be useful in HPC practice. The proposed follow-up meeting will gauge this readiness and set further milestones. It will be a forum for staged programming researchers to present their progress and for HPC practitioners to evaluate it, fostering the collaboration on real-life applications.

Generative programming, in particular, in the form of staging, is widely recognized in HPC as the leading approach to resolve the conflict between high-performance on one hand, and portability and maintainability on the other hand. In its general form, staging is an implementation technique for efficient domain-specific languages (DSL), letting HPC experts conveniently express their domain-specific knowledge and optimization heuristics. However, the results and tools of the current staging research are little used in the HPC community. Partly this is because HPC practitioners are not aware of the progress in staging; staging researchers are likewise unaware of what HPC practitioners really need. The first Shonan meeting brought together HPC practitioners and programming language (PL) researchers to break this awareness barrier. The meeting aimed to solicit and discuss real-world applications of assured code generation in HPC that would drive PL research in meta-programming.

The first Shonan meeting succeeded in its aim. It developed a set of benchmarks, representative HPC examples, where staging could be of help in producing more maintainable code and letting domain experts perform modifications at a higher-level of abstraction. This set was dubbed ‘Shonan Challenge’.

Shonan Challenge has greatly stimulated research and development of staging, resulting in extensible compilers based on staging (Rompf et al., POPL 2013) and the revival of MetaOCaml (ML 2013). The answers to Shonan Challenges have been presented in the overview paper (Aktemur et al., PEPM 2013), in (Rompf et al., POPL 2013) and in a poster at APLAS 2012. Shonan Challenge problems (specifically, the Hidden-Markov Model benchmark) were discussed at the staging tutorials given in 2013 at the premier PL conferences PLDI, ECOOP, and ICFP/CUFP.

It is time to report the progress in staging back to the HPC practitioners who posed the challenges, to evaluate how well the developed tools meet the needs of the HPC practice already, and what is yet to be done.

As before, we anticipate the workshop participants to consist of three groups of people: PL theorists, HPC researchers, and PL-HPC intermediaries (that is, people who are working with HPC professionals, translating insights from PL theory to HPC practice). To promote the mutual understanding, we plan for the workshop to have lots of time for discussion. We will emphasize tutorial, brainstorming and working-group sessions rather than mere conference-like presentations.

# 042 Water Disaster Management and Big Data

Shonan Village Center, July 7 – 10, 2014

Seminar No.051

## Organizers

Fang-Pang Lin, National Center for High-Performance Computing, National Central University, Taiwan

Philip M. Papadopoulos, University of California, San Diego, USA

Satoshi Sekiguchi, National Institute of Advanced Industrial Science and Technology (AIST), Japan



## Topics and Goals

The water disaster management challenge was recently articulated at PRAGMA 22 (Sapporo) by Kenzo Hiroki (ICHARM, [http://www.icharm.pwri.go.jp/staff/staff\\_hiroki.html](http://www.icharm.pwri.go.jp/staff/staff_hiroki.html), Research Coordinator for Water Disaster Risk Management). Within the PRAGMA community there have been many experiences with flooding and mudslides (Taiwan's Hsiao-Lin Village during Typhoon Morakot (2009); Thailand's flooding in September 2011; flooding in Australia; ..), and some projects to attempt to address aspects of preparedness (Taiwan's flood grid; AIST GEOGrid). Following this background, this meeting will bring together communities from water disaster management and big data and cyberinfrastructure communities to address shared problems, and develop approaches to tackle water disaster problems found in common across these areas. Furthermore, in line with PRAGMA's mission and its more than ten years of practice, the intent is to establish sustainable research collaboration between various communities.

## Topics of the Meeting

Water disasters include flooding from typhoon, seasonal storms and other natural or man-made causes, usually coupling with land, mud-slides and often occur without warning. Water disaster management is a challenge that faces many parts of the world.

The topics will include:

- 1) Water disaster events and mitigation practices
- 2) High frequency sensor networks and fast response
- 3) Remote sensing and aerial imaging
- 4) Flood and land sliding modeling
- 5) Crowd sourcing and data mining for water disaster management
- 6) Tera- and Peta- scale data infrastructure
- 7) Configurable and high availability networking

## Goals of the Workshop

Structure of Workshop:

We anticipate a series of presentations, increasingly technical, to provide examples of the needs of individuals and groups responsible for anticipating, planning for, and responding to natural disasters caused by water. We plan to have individuals from several countries present scenarios where there are technical challenges. We anticipate that there will be a dialog between individuals and groups planning or responding to aspects of disasters and individuals who can understand and ask questions to clarify technical aspects of the challenges. Out of that discussion we hope to classify challenges as long term (research: there are basic questions for which there are no existing solutions), medium term (development: most of the technologies are developed but need to be integrated or tested or modified) and short term (deployment: (affordable) solutions exist).

Goals of Workshop:

The workshop will produce a report for the community. We plan to achieve the following goals in water disaster management and big data:

- a) generate possible research opportunities for funding agencies (such as JSPS or NSF);
- b) create some personnel connections between participants;
- c) develop funding proposals to further ideas from the workshop; and
- d) in the PRAGMA Context establish a working group or new expedition to prototype solutions, in collaboration with groups with the needs.



# NII Shonan School on Coq

# 043



Shonan Village Center, August 25–29, 2014

Seminar No.041

## Organizers

Pierre Castéran, LaBRI, France  
Jacques Garrigue, University of Nagoya, Japan  
David Nowak, CNRS and Lille 1 University, France

The objective of this school is to teach the use of the Coq proof assistant that has received the SIGPLAN Programming Languages Software 2013 Award and the 2013 ACM Software System Award. It will consist of lectures and practices in English by internationally renowned experts in Coq.

### Prerequisites

No previous knowledge of Coq will be necessary to follow the lectures. Students are however requested to attend the school with their own computer with the latest stable version of Coq and one of its user interfaces (CoqIDE—distributed with Coq, or Proof General) installed. They are freely available online:

- Coq and CoqIDE: <http://coq.inria.fr>
- Proof General: <http://proofgeneral.inf.ed.ac.uk>

### What is the Coq proof assistant?

In principle, all mathematics can be formalized in axiomatic set theory, and then checked automatically by a computer. In the last century, more practical foundations for mathematics, based on type theory, have been designed and implemented as proof assistants. One of the most prominent is Coq, developed in France since 1984. They can check proofs and organize them in searchable libraries. They also provide convenient interfaces that help the user make proofs incrementally, and fill in automatically the trivial parts.

### What can be done with Coq?

In the beginning, proof assistants could only be used to formalize toy examples. It is however not anymore the case. Proof assistants have reached maturity and can deal with difficult mathematical results. For example, in December 2004, Gonthier has announced the full formalization of the four colors theorem in the Coq proof assistant (cf. Notices of the AMS 55(11), 2008). A more recent example is the proof of the Feit-Thompson theorem completed in Coq in 2012.

Mathematicians are not the only ones to make mathematical proofs. Computer scientists, for instance, make their own proofs. In a software, a small error can indeed result in serious damages in terms of safety and security. It is thus important to provide formal proofs that a software is correct: extensive testing of software or hardware is not enough because it may still miss errors that would be avoided with the use of a proof assistant. One prominent example is the formalization of a complete Java Card system in the Coq proof assistant and the certification that it meets the highest security requirements for industrial product, i.e., the Common Criteria EAL7 level. This was achieved in 2003 by Trusted Logic, a company that provides secure software for smart cards, terminals and consumer devices. Another example is the compiler CompCert: it is proved in Coq that it will always generate assembly code that has the same semantics than the source code in C. From the mathematician's eye, some of those proofs may not appear difficult but they are nonetheless tricky and error-prone because they involve huge formal objects such as programs or automata, or require to consider hundreds of cases.

# Algorithmic Randomness and Complexity

# 044

Shonan Village Center, September 8–12, 2014

Seminar No.047

## Organizers

Rodney Downey, Victoria University of Wellington, New Zealand  
Kenshi Miyabe, The University of Tokyo, Japan  
André Nies, University of Auckland, New Zealand  
Osamu Watanabe, Tokyo Institute of Technology, Japan



Information and randomness are fundamental notions of computer science. Their practical applications are ever increasing. Classical information theory and probability provide formalizations, but do not allow us to speak of the information content of a specific string or to say that a particular infinite binary sequence is random.

In recent years the people working in computability theory have produced a surge that resulted in rapid progress in our understanding of even the most basic notions, and in the solution of old open questions. This progress has changed and is still changing the landscape and opened up new avenues of research.

Research on the notions of information and randomness has drawn on methods and ideas from computability theory and computational complexity, as well as from core mathematical subjects like measure theory and information theory.

## Specific topics

### Algorithmic randomness and computational analysis

The areas of analysis and randomness are closely connected because the measure of a set (a formal notion corresponding to our intuition of size or volume) is fundamental in both areas. An interesting emerging direction is when both the randomness notion and the function class are defined in terms of feasible computation.

### Generic Computation

In computational complexity theory, Gurevich and Levin independently introduced the idea of average-case complexity. One has a probability measure on the instances of a problem and one averages the time complexity over all instances. For example, an important result of Blass and Gurevich states that the Bounded Product Problem for the modular group,  $\text{PSL}(2, \mathbb{Z})$  is NP-complete but has polynomial time average-case complexity. The apparatus of average case complexity is difficult to apply, and it is natural to seek a more easily applied technique for classifying complexity which is natural.

### Derandomization and complexity hierarchies

Derandomization is the study of how to replace probabilistic algorithms with deterministic algorithms. Earlier work by Allender et al. showed that the techniques of derandomization could be viewed through the lens of resource-bounded Kolmogorov complexity theory, and gave significant applications.

### Resource bounded versions

Classical computational complexity theory comes into play defining resource-bounded versions of Kolmogorov complexity, measure, and dimension. This has led to new characterizations of complexity classes involving efficient reducibility to the set of Kolmogorov random strings. Resource-bounded measure and dimension have been used to gain understanding of properties of complexity classes and their complete sets.

(Excerpt from the submitted proposal)

# Design Methods for Secure Hardware

# 045



Shonan Village Center, September 15–19, 2014

Seminar No.028

## Organizers

Kazuo Sakiyama, University of Electro-Communications, Japan

Patrick Schaumont, Virginia Tech, USA

Ingrid Verbauwhede, K.U. Leuven, Belgium

This meeting will gather researchers working on secure hardware components that support information security such as encryption and decryption, electronic signatures, and authentication. Secure hardware design is characterized by a specialized application domain (cryptography), and it is required for applications that have a reliability requirement under adversary operating conditions. Similar to other application domains such as multimedia, there is a need for fast, compact, low power and/or low energy realizations. In addition however, there is a need for secure realizations—the implementations have to remain reliable and trustworthy under attacks and adversary operating conditions. This makes the application domain quite unique.

The meeting will review the state-of-the-art in secure hardware design, and it will create an open forum for the discussion of important open research questions in this fast-evolving and important field of research. The emphasis of the meeting will be on design methods, and the systematic steps that designers use to construct secure hardware. Several of the questions that we would like to address include the following.

1. What analysis techniques are applicable to the design descriptions of secure hardware? How can risk be quantified and how can we convey the results of this analysis to the designer?
2. When is secure hardware preferable over secure software, and how can we help a designer choose between these two options? What aspects of security can be rendered more efficiently in hardware than in software?
3. How do implementation attacks (fault analysis and side-channel leakage) affect classic hardware design issues such as reliability, design for testability, and verification? What guidelines can be defined for the hardware designer?
4. Can we define correct-by-construction techniques for secure hardware design? What aspects of cryptographic engineering are amendable to compiler techniques?
5. How to define design libraries for secure hardware design? Is the problem similar to the classic case of Intellectual Property Reuse of hardware components? Do the security properties of individual hardware components reflect on the overall design?
6. What runtime techniques are available to evaluate a secure hardware design? How can we detect implementation attacks, and what assumptions do we have to make?
7. What techniques can be used to cost-optimize secure hardware? How do we quantify the security/cost tradeoff?

# 046 Algorithms for Large Scale Graphs

Shonan Village Center, October 14–17, 2014

Seminar No.049

## Organizers

Andrew McGregor, University of Massachusetts, Amherst, USA

Gopal Pandurangan, University of Houston, USA

Sergei Vassilivitskii, Google Inc., USA



## Abstract

Large scale graphs abound in modern data analysis. They are used to model social and communication networks, item-item similarities in recommendation settings, relationships between buyers and sellers in large scale markets and so on. Yet, at the same time, graph algorithms are among the hardest to scale to very large inputs. There are three general directions for scaling such approaches: sparsifying and summarizing the graph, retaining only the crucially relevant information; streaming it, and thus never storing the full input; or processing different parts of it in parallel. The goal of this workshop is to bring together researchers from all three of these areas to share ideas and stimulate further development in this exciting and challenging area.

## Description of the Meeting

The growing interest in “Big Data” has led to an increased interest in designing algorithms and systems for problems with very large inputs. Among these problems, graph problems are among the most challenging: basic algorithmic primitives, such as depth first searches, cannot be performed in the data stream model and existing algorithms can be notoriously hard to parallelize, to the extent that numerous graph-specific parallel systems have been developed. Nevertheless, there have recently been exciting developments on graph problems, both in streaming and parallel models, as well as work on sketching and sparsifying graphs.

## Sketching, Sparsification, and Streaming

One way to deal with the data deluge is to reduce the graph size to something more manageable, while at the same time preserving the key properties of the graph.

## Parallel and Distributed Computation

Computation on modern massive graphs often proceeds in a distributed manner. Whether it is computing PageRank, clustering, or simply computing shortest paths, practitioners employ many different systems to get at their result. Each of these systems makes different tradeoffs depending on the possible priorities: Is it important to reduce communication? Can we reduce the computational power of each node? How can we ensure good load balancing or guarantee real-time responses?

## Conclusions

In a growing number of applications, there is the need to efficiently process large scale graphs. Developing a theory and principled approach to developing such algorithms represents a new challenge that needs the expertise of researchers in data streams, distributed computing, and graph algorithms. The goal of this workshop is to bring together researchers from these diverse communities to brainstorm and stimulate further development in this exciting and challenging area.

(Excerpt from the submitted proposal)

# Computational Intelligence for Software Engineering

# 047



Shonan Village Center, October 20–23, 2014

Seminar No.053

## Organizers

Hong Mei, Shanghai Jiao Tong University, Peking University, China  
Frank Neumann, The University of Adelaide, Australia  
Leandro Minku, The University of Birmingham, United Kingdom

Computational intelligence (CI) techniques have provided many inspirations for improving software engineering, both in terms of the engineering process as well as the software product. The application of CI techniques in software engineering is a well-established research area that has been around for decades. There have been dedicated conferences, workshops, and journal special issues on applications of CI techniques to software engineering.

It is interesting to note that search-based software engineering does not provide merely novel search and optimisation algorithms, such as evolutionary algorithms, to solve existing software engineering problems. It helps to promote rethinking and reformulation of classical software engineering problems in different ways. For example, explicit reformulation of some hard software engineering problems as true multi-objective problems, instead of using the traditional weighted sum approach, has led to both better solutions to the problems as well as richer information that can be provided to software engineers. Such information about trade-off among different objectives, i.e., competing criteria, can be very hard to obtain using classical approaches.

However, most work in search-based software engineering has been focused on increasing the efficiency of solving a software engineering problem, e.g., testing, requirement prioritisation, project scheduling/planning, etc. Much fewer works have been reported in the literature about CI techniques used in constructing and synthesizing actual software. Automatic programming has always been a dream for some people, but somehow not as popular as some other research topics.

The primary aim of this proposed Shonan meeting is to provide an interdisciplinary forum for computational intelligence and software engineering. It is envisioned that there will be 50% of the participants from each of the two areas. The goal of the meeting is to explore and debate this very promising area of CI for software engineering, including search-based software engineering.

The meeting will include presentations by the participants as well as discussion groups for hot topics and future work. The discussion groups play a central role and should reflect on the current state of the art in the different areas of software engineering and computational intelligence, foster interdisciplinary work, and establish new research directions.

(Excerpt from the submitted proposal)

# Science and Practice of Engineering Trustworthy Cyber-Physical Systems (TCPS)

Shonan Village Center, October 27–30, 2014

Seminar No.055

## Organizers

Fuyuki Ishikawa, National Institute of Informatics, Japan

Alexander Romanovsky, Newcastle University, United Kingdom



The society and human activities have been depending more and more on software-intensive systems. Novel system paradigms have been proposed and actively developed, notably Cyber-Physical Systems (CPS). Envisioned systems expands target entities and processes handled by the systems, stepping into more depth of human activities as well as real world entities. There are emerging application areas such as automated driving and smart cities, while existing areas are also evolving with richer features, such as aviation, railways, business process management, navigation systems, etc. Visions for CPS include or extend a lot of variations of system paradigms, Systems of Systems, Ubiquitous Computing, Ambient Intelligence, and so on. Obviously, the increased impact on human activities and real world entities lead to strong demand for trustworthy systems. On the other hand, resulting is unprecedented complexity, caused not only by expanded application features, but also by combined mechanisms for trustworthiness (self-adaptation, resilience, etc.). Construction and provision of trustworthy systems under complexity is absolutely the key challenge in system and software engineering.

The key to tackle the challenge is engineering methods for trustworthy systems. There is no doubt that foundational theories and technical components are essential as building blocks. Building blocks for trustworthy systems spread across verification algorithms, probabilistic analysis, fault models, self-adaptation mechanisms, and so on. The challenge on complexity requires further, elaboration and integration of such blocks into engineering methods. Engineering methods define a systematic and reliable way for set of tasks to model, analyze and verify the system and its trustworthiness nature while mitigating the complexity. Recently, there have been yet more active efforts on engineering methods for trustworthy systems, on the basis of various approaches. Formal methods are one of the promising approaches and have accompany active efforts not only by the academia but also by the industry. Each approach has apparently different, unique features, but essentially relevant to each other, focusing on modeling of the system, modeling of trustworthiness or faults, and their analysis and verification for complex systems, especially Cyber-Physical Systems.

In order to speed up the evolution of engineering methods for emerging complex Cyber-Physical Systems, it is absolutely necessary to promote active discussions beyond specific applications or specific engineering approaches. This proposed Shonan Meeting aims at providing this opportunity by inviting world-leading researchers on engineering methods for trustworthy Cyber-Physical Systems.

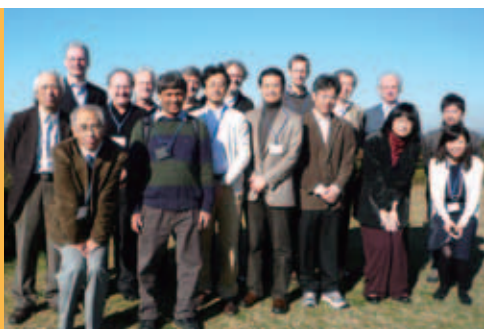
This meeting provides an opportunity for leading researchers to exchange and discuss their latest (possibly immature) ideas as common in the Shonan Meeting.

For these purposes, the meeting will involve two kinds of sessions (possibly with special sessions as described below). One is presentations and targeted discussions, where each researcher presents ideas and positions that then kicks off various directions of discussions. The organizers will ensure that these presentations are structured around specific topics. The other is sessions consisting of intensive follow-up discussions involving mixed groups of attendees. The meeting will use a dedicated method for conducting the intensive discussion, the current plan is to use a variation of the world café method: <http://www.theworldcafe.com/method.html>. The organizers will prepare the discussion topics such as “what are challenges specific to trustworthiness of emerging CPS?”, “what are essential requirements for engineering methods for CPS?” and “what are the future CPS?”. The topics will be extended at the meeting as well.

(Excerpt from the submitted proposal)

# Towards Explanation Production Combining Natural Language Processing and Logical Reasoning

# 049



Shonan Village Center, November 27–30, 2014

Seminar No.057

## Organizers

Randy Goebel, University of Alberta, Canada

Yuji Matsumoto, Nara Institute of Science and Technology (NAIST), Japan

Ken Satoh, National Institute of Informatics, Japan

The recent application of high performance natural language processing (NLP) systems has shifted from general question answering to more targeted application domains (e.g., medical diagnosis, epidemiology, legal reasoning, intellectual property management). Examples include the WATSON project by IBM and the Todai Robot Project by NII. But these recent systems have focused more on engineering problem solutions, and less so on general architectures for building domain models that support more general representation and reasoning within the specific application domains. For example, the WATSON architect uses an associative pattern matching process which supports high performance, but is difficult to follow for humans, and thus prevents easy collaborative problem-solving with humans. Our workshop proposal is focused on the development and integration of modern natural language processing tools to support the development of systems with methods of hypothesis management and explanation. In this way, humans can collaborate with such systems to not only understand the use of representation to create answers, but also support incremental supervised machine learning. The purpose of the proposed meeting is to gather a group of researchers in related areas, such as natural language processing, information extraction, and logical reasoning, in order to formulate approaches to combining these areas to achieve the goal of machine-human collaboration in high-performance natural language-based domain interactive reasoning systems.

A research agenda for such a meeting would include a variety of topics, including language-based information extraction (e.g., open information extraction), reasoning architectures based on abduction and hypothesis management, natural language entailment, and inductive learning.

Specific challenges include the capture and use of legal documents, in order to answer legal questions. In this context, the creation of legal judgements involves not just answers (e.g., guilty, not guilty), but articulation of elaborate explanations supported by interpretation of legal statutes and regulations. To construct such logical explanations, we need to identify information relevant to the questions, ensure that information is transformed to or directly represented in formal representations that support the contraction of explanations. In this phase, we need a NLP analysis of question and information retrieval technique based on the analysis. Subsequent steps include the development of entailment testing methods, to decide whether and how (or how not) a question is entailed by the logical representation of the extracted domain information. The general development of NLP techniques for this kind of subsumption involves ontological manipulation, relation extraction, and logical reasoning.

Because this diversity of requirements, our workshop will need to include researchers from information retrieval, information extraction, question-answer, general natural language processing, and logical representation and reasoning. With this context, the goal is to develop a basis for reliable system which produces an explanation of system's behaviour, and be incrementally constructed and improved to advance their level of performance in a variety of application domains.

Possible topics of the meeting include, but are not limited to:

- Natural Language Processing for Information Retrieval
- Information Retrieval for Complex Logical Formula
- Natural Language Processing for Textual Entailment
- Natural Language Processing for Subsumption Test
- Ontology Research for Subsumption Test
- Combining Logical Reasoning and Natural Language Processing
- Feedback of Logical Reasoning to Solve Disambiguation

# Integration of Formal Methods and Testing for Model-Based Systems Engineering

Shonan Village Center, December 1–4, 2014

Seminar No.048

## Organizers

Tetsuya Tohdo, Denso Corporation, Japan

Werner Damm, Carl von Ossietzky Universität Oldenburg, Germany

Alexander Pretschner, Technical University of Munchen, Germany

Jun Sun, Singapore University of Technology and Design, Singapore



This meeting aims to provide a forum for researchers and practitioners working on Formal Methods (FM) and Testing to identify research issues toward bringing related ideas, techniques, theories and tools in the both research areas to real industrial solutions.

The critical systems manufactured in automobile, railways and avionics need to be designed, developed, operated and maintained with high-level of confidence, and to keep evolving to adapt ever-changing environments, and responding to new demands in a timely manner with high productivity. Observing the industry, so called “Semi-Formal” approaches are widely taken with the use of executable models (the notable example is Simulink widely used in the automotive industry) in conjunction with the use of traditional Testing methods instead of formal techniques. While the adoption of results in Formal Methods research is gaining momentum, there continue to be gaps between practical industry needs and academic achievements. Among others, these gaps reflect concerns with return-on-investment, scalability, practicality, educational challenges, and the seamless embedding of methods and tools into an ever changing landscape of development processes. This current situation motivates us to organize this meeting.

Model-based engineering approach based on the use of modeling descriptions with formal mathematical semantics (as used in FM) in combination with Testing is considered to be one of the best practical solutions over the system-lifecycle to ensure high level of confidence without decreasing productivity. However, the best way of combining FM and Testing techniques is not obvious, and we believe that there are numerous research issues on this approach. The main purpose of this meeting is to recognize and identify research issues which need to be solved in order to develop engineering methods which effectively combines FM and Testing theories, techniques and tools.

This forum has a strong focus on collaboration between academia and industry in order to foster further development of this cutting-edge technology and aims to establish a strong tie between academia and industry especially to Japanese industry. In order to make this happen, we carefully selected invitees from academia who are working on systems engineering, foundational theories between Testing and Formal Methods, and principal researchers conducting research projects on embedded systems, systems engineering and verification, and practitioners who are practicing verification in model-based systems engineering.

Some of the issues we would like to address at this meeting may include but not limited to the followings:

- + Test case generation using formal proofs and model checking
- + Unified theory of Testing and Formal Methods which underpins both technique in a single framework
- + Contractual framework for specification, design and verification
- + Combination of static analysis and dynamic analysis/testing
- + Testing and symbolic verification
- + Approximation and abstract interpretation
- + Refinement calculus and Testing
- + Requirement-based verification
- + Model-based testing
- + Runtime Verification



# Big Graph Drawing: Metrics and Methods

# 051



Shonan Village Center, January 12–15, 2015

Seminar No.054

## Organizers

Takayuki Itoh, Ochanomizu University, Japan  
Karsten Klein, Monash University, Australia  
Giuseppe Liotta, University of Perugia, Italy

## Introduction

Graphs provide a versatile model for data from a large variety of application domains, including for example biology, finance, information security, telecommunication, software engineering, and social sciences. Graph visualization helps scientists and engineers to understand critical issues in these domains. However, the depth of understanding depends on the quality of the drawing.

Good visualization can facilitate efficient visual analysis of the data to detect patterns and trends, and can also be used for the presentation of the underlying information. Important aspects for the development of graph drawing methods are the efficiency and accuracy of the algorithms, and the quality, i.e. readability, of the resulting pictures.

Quality metrics need to be developed to measure and optimize the graph drawing quality for the task at hand. These metrics need to cover both how accurate the drawing represents the data as well as how well the drawing can be perceived by the human user.

## Topics and Aims of the Seminar

The research on big graph visualization touches several research fields, including network visualization, network analysis, graph theory, computational geometry, algorithm engineering, computational complexity, and main application areas like bioinformatics, social sciences, information security, and software engineering. We will bring together experts in the fields to investigate research questions and potential solutions.

The aims of the seminar are as follows:

1. To bring together experts from the graph drawing and information visualization communities from Europe, Asia, the Americas, and Australia that are involved in layout evaluation and large data visualization, and domain experts from application areas that require big graph drawing.
2. To investigate the challenges of today's big data graphs for graph drawing, including layout methods, interaction, and computational aspects.
3. To investigate how the evaluation of drawing approaches can be done for such big graphs, including layout quality and interaction efficiency, and to study and compare the performance of existing metrics and methods for real-world instances.
4. To create new quality metrics and discuss new methods for big graphs, i.e. fast methods that optimize the metrics and are suitable for use in practice, and new interaction concepts.
5. To formulate guidelines for the visualization of large graphs, state the pitfalls, and the problems with current metrics and methods.
6. To define and document corresponding research challenges for future research.

(Excerpt from the submitted proposal)

# Logical Analysis of Descriptions and Their Presentations —A Computational Logic Approach

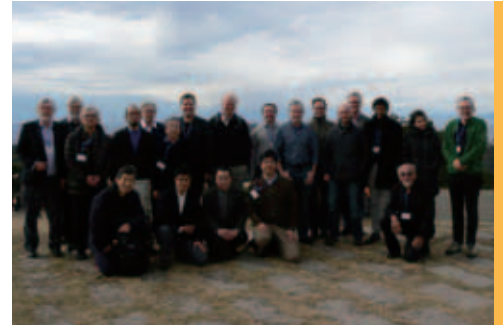
Shonan Village Center, January 26–29, 2015

Seminar No.042

## Organizers

Yoshiki Kinoshita, Kanagawa University, Japan

Bengt Nordström, Chalmers University of Technology, Sweden



## Aim of the workshop

Our aim is to investigate whether logical methods are applicable to analysis of descriptions in, for instance,

- system specification,
- formal mathematics,
- system assurance,
- juris-informatics (law) and
- industrial standardisation.

In these areas, we have descriptions of some objects and relations between them. These descriptions are either formal or informal (in English or some graph language without precise semantics). We want to conduct logical, computational and mathematical studies that clarify the structure of such descriptions. Aspects of this has been studied in discourse representation theory, constructive type theory and informal logic. Results of earlier studies include object role modelling (ORM), state charts, Toulmin diagrams, goal structuring notation (GSN), grammatical frameworks (GF) and proof assistants. Some applications may be found in the above areas.

There has been an emphasis on reasoning aspects when applying mathematical logic in Computer Science and Informatics. In this workshop, we want to investigate the importance of ontological aspects of mathematical logic. These are more important than deductive aspects in, for instance, validation of information systems. We are interested in description languages of varying level of formality, where the formal part makes some mechanical checking and manipulation possible.

The combination of deduction and ontology can also be found in mathematics where the concepts have matured during hundreds of years. Nowadays, a similar combination of ontological and deductive aspects can be found in descriptions of information systems, laws and rules, etc. We are convinced that studies of informal mathematics will be an important contribution to a deeper understanding of these issues.

This is an attempt to put together people from different fields to find a common analysis of Logical Description Languages. We need to agree about the major research issues. The time is not yet ripe to agree about techniques. The invited people are rather few, which is necessary to make progress towards a common understanding.

(Excerpt from the submitted proposal)

# Systems Resilience—Bridging the Gap Between Social and Mathematical

# 053



Shonan Village Center, February 23–26, 2015

Seminar No.060

## Organizers

Hiroshi Maruyama, The Institute of Statistical Mathematics, Japan

Günter Müller, University of Freiburg, Germany

Kazuo Furuta, The University of Tokyo, Japan

The goal of this meeting is to bridge the gap between the “social” and “mathematical” camps of resilience research so that the social aspects of resilience are more appropriately incorporated into the mathematical models and at the same time the mathematical models can provide practical guidance to the design, policy making, and operations of real-world societal systems.

Resilience is said to be an ability of a system to absorb and recover from perturbations. It is considered to be a critical characteristic for a system to survive, especially for social systems like organizations, communities, cities, and our civilization as a whole. Resilience has been studied in many different domains, such as psychology, biology, ecology, engineering, and social sciences, but often their approaches are widely different. We observe that there are at least two seemingly incongruent approaches: social and mathematical.

The social camp, mainly dealing with problems such as socio-ecological resilience and urban resilience, is concerned with resilience as a social norm. Their research approaches are based on case studies, best practices, processes, communication, decision making, consensus building, and other disciplines, and little mathematical models are used except for relatively simple system dynamics to compare different scenarios. Policy makers can learn from these studies to make better decisions in face of possible disruptions. However, these approaches do not guarantee nor give quantitative assurance to how much the resilience strategies can contribute to the survivability of the system. The mathematical camp, on the other hand, is interested in the mechanisms of how systems can collapse and in what conditions resilience strategies work for recovery.

The goal of this meeting is to bring researchers in those two camps and to explore common grounds so that the social and mathematical approaches are integrated to make objective and practical resilience strategies. In order to make our discussions focused, we will first define a layered domains, with the real world at the bottom, Cyber-Physical Systems (CPS) as the second layer, and cyber security at the top. In this process of layered abstraction, we keep the essential aspects of socio-technical systems, including human behaviors, social and economic factors, and technical and systems workings intact so that the microcosm at the top still retains similar (albeit not the same) characteristics of the real world. Then, we will discuss how the case studies, best practices, human factors, etc. are interconnected to selected mathematical models using this narrowed domain. As an outcome of this meeting, we expect to have a joint view on resilience, which is an amalgamation of both social and mathematical approaches.

We shall invite participants from diverse disciplines. Resilience domain in itself is multidisciplinary (relating to several disciplines) and transdisciplinary (using approaches that transcend specialization boundaries). Secondly, these academicians have shown significant interest and contributions to advancing resilience thinking as they have demonstrated in international meetings that our team members also attended. We shall also invite PhD candidates who have demonstrated in our previous meetings critical thinking and research communication skills. Hence, we believe that we can leverage this diversity of participants to achieving the goals we have set for this workshop.

(Excerpt from the submitted proposal)

# 054 Low Level Code Analysis and Applications to Computer Security

Shonan Village Center, March 2-5, 2015

Seminar No.065

## Organizers

Jean-Yves Marion, Lorraine University, France

Akira Mori, National Institute of Advanced Industrial Science and Technology (AIST), Japan

Mizuhito Ogawa, Japan Advanced Institute of Science and Technology (JAIST), Japan



Software security is one of the very major issues of modern societies, since insecure systems can affect people (phishing, unauthorized payment, etc.), organizations (critical information leakage) and even states (cyber attacks). In regard to software security, a crucial problem is to have methods and to devise tools to analyze low-level code for at least three reasons. First, low level codes are usually the one that is effectively run on a computer system. Low-level code is usually obtained by a compilation chain where some parts of the chain maybe unsecure. As a result it is necessary to guarantee low-level code correction ensuring that programs run by the system cannot be attacked. Second, a lot of applications are available on internet as binary codes. Those binary codes are quite obfuscated either to protect the intellectual property or because there are malicious software (malware). Therefore, it is necessary to have tools to analyze binary codes in order to detect malware and prevent an attack.

It may look surprising that such important issues are so poorly addressed by current state-of-the-technology tools, while formal methods in general and automatic program verification in particular have made tremendous progress in the past decade, resulting in many impressive industrial applications, and demonstrating their strength both in bug finding and safety validation. There are two key problems here.

First, security needs are more difficult to characterize (in terms amenable to automatic analysis) than safety needs. Security requirements change depending on the social context, and malicious behavior is not easy to define. For instance, malware often uses windows system calls for malicious behavior, but standard programs also use such capability. To clarify, not only theoretical observation form academia, but also empirical study in industry will be required.

Second, security analyses must often be performed at the level of binary executable. Malicious behavior is often based on very fine low-level details, invisible on the source code (e.g. memory layout, exception handling details, ambiguity in program semantics, bug in compilation), and malware are only available as executable files. This is a key difficulty, as most formal methods have been designed for high-level models or languages, and they rely on hypotheses that no longer hold at the executable code level. For instance, even (static) disassembly is not possible for recent advanced polymorphic virus, like self-decryption. Thus, model generation of binary executables is already a challenging.

By interacting industry and academia, and among different areas in academia, we will start to tackle the two key problems.

Analyzing security needs. Learn concrete examples from industrial experiences, and manual pattern analysis of malware behavior in malware database. They would lead clear formal definitions of target properties.

Binary executable analysis. Exchanging ideas of state-of-the-art research on binary executable analysis, and look for collaboration opportunity.

(Excerpt from the submitted proposal)

# Static Analysis Meets Runtime Verification

# 055



Shonan Village Center, March 16–19, 2015

Seminar No.062

## Organizers

Cyrille Artho, National Institute of Advanced Industrial Science and Technology (AIST), Japan  
Einar Broch Johnsen, University of Oslo, Norway  
Martin Leucker, University of Lübeck, Germany  
Keiko Nakata, FireEye Dresden, Germany

## General background

Our life is increasingly dependent on the correct functioning of software, and the role of software in complex products is expanding fast. The increase in software complexity is paralleled by the amplified risks connected to its failure. However, in most industrial environments, many constraints (time-to-market, cost, lack of skills, etc.) make unrealistic manual verification, validation and certification. We are in need of technologies that can be deployed within industrial constraints and that offer high levels of guarantees for complex systems.

## Static analysis

Static analysis examines software applications without actually executing them. It can perform sophisticated analysis, and does not require test cases or the complete code, making it very useful in industry applications.

However, due to its exhaustive nature, static analysis is difficult to scale for complex data structures without sacrificing the precision. Static analysis has been recognized as a fundamental tool for program verification, bug detection, compiler optimization, program understanding, and software maintenance. On the one hand, new theories are continuously proposed to subsume new computing models of modern software, such as distributed computing and resource-aware computing.

## Runtime verification

Runtime verification is a computing analysis paradigm based on observing a system at runtime to check its expected behavior. However, the quality of runtime analysis depends on test scenarios, which makes it less robust compared to static analysis. Runtime verification complements design time static analysis with lightweight verification techniques that check violation of intended properties online. Specifically, the field has been addressing technical challenges of generating efficient monitors from high level specification and of formally specifying recovery actions at the specification level.

## Goals of the meeting

The goal of this meeting is to bring together the two communities, to combine the robustness of static analysis and the flexibility of runtime verification. Recent years have seen, on one hand, the use of static analysis in the context of runtime verification to reduce the size of runtime models by pruning certain scenarios that are statically analyzed, on the other hand, the use of runtime verification in the context of static analysis to ease verification burden by deferring certain properties to be verified at runtime. These efforts involve both theoretical challenges of unifying different formalisms and engineering challenges of enabling different tools to communicate with each other. Another direction of research is to develop technologies for verifying software under open-world assumptions, where software is made from heterogeneous, possibly third-party, components, whose behavior and interactions cannot be fully controlled or predicted. Exhaustive static analysis is not possible for open-world software, but runtime verification can come to rescue by, for instance, verifying, at runtime, assumptions about the open world that are made during the static analysis. Discussions at the meeting will help identify research needs for combining static analysis and runtime verification. We discuss challenges for the two fields, such as verification of concurrent systems. Participants will be able to discuss technical approaches that have emerged in various related research areas and assess their applicability to the challenges we face.

(Excerpt from the submitted proposal)

# Instruction and Instructed Action: Embodied Reciprocity in Interaction

# 056

Shonan Village Center, March 19–21, 2015

Seminar No.067

## Organizers

Lorenza Mondada, Basel University, Switzerland

Mayumi Bono, National Institute of Informatics, Japan

Aug Nishizaka, Chiba University, Japan



The goal of this workshop is to discuss embodied reciprocity in interaction with a special reference to instruction and instructed action. Instructions can be found everywhere, from ordinary life to the workplace, and are a frequent type of action through which one person shows, demonstrates, but also requests to another to do something, which is realized in instructed action. Although instructions have been widely studied in linguistics and pragmatics, we contextualize them in the tradition of ethnomethodology and conversation analysis. Instead of treating instruction as a unidirectional speech act that determines instructed action, we consider instruction and instructed action as paired actions that are mutually dependent, and that are interactionally achieved (e.g. Garfinkel, 1967). In order for instruction to complete, recipients interpret and produce embodied practical actions (Suchman, 1987). This mutual dependency, or reflexivity between instruction and instructed action, is one of the important themes in this workshop.

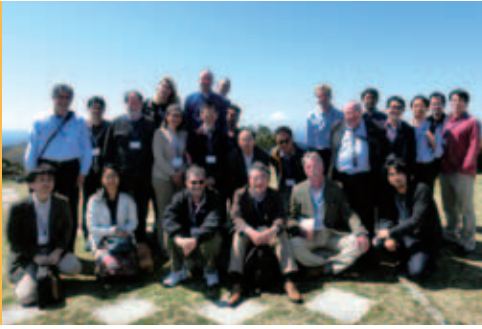
Instruction has been largely studied in educational settings including classrooms (Macbeth, 2011), crochet Lessons (Lindwall & Ekstöm, 2012), trainings in surgery (Koschmann et al., 2011; Sanchez Svensson et al, 2009), and driving lessons (De Stefani & Gazin, forthcoming). While institutional contexts are a useful place to examine how instruction is sequentially constructed and how it is supported by those who provide and receive an instruction, such as a teacher and a student, (for the IRF/IRE sequence, see Sinclair & Coulthard, 1975; Mehan, 1979), we expand the scope of analysis and consider the fact that instruction is prevalent in a wider range of social situations including ordinary interactions between families and friends, workplace, and technologically mediated environments. Specifically focusing on Conversation Analysis, Ethnomethodology and closely related approaches, our aim is to examine the situated and embodied nature of instructions. In a surgical operation room, for example, the surgeon's directives instruct his or her assistant's embodied action, such as handing in an appropriate instrument to the surgeon. Although instructions are highly indexical, they are understood within the praxeological context of the ongoing procedure, in which responsive instructed action reflexively provide for the intelligible character of minimal instructions (Mondada, 2014). In a robot/android theater, the director's instruction of timing a 1-second pause is accomplished in the course of interactions made in multiple rehearsal sessions. Instead of using a watch, actors appropriately arrange and "craft" their bodies in order to perform the pause. (Bono, in preparation). These studies suggest that instruction and instructed action are not simply correlated, but mutually dependent, locally organized actions involving various semiotic resources such as language and the body as well as material objects such as tools and computers.

Studies of instruction and instructed action in various naturally occurring interactions contribute to the field of engineering and informatics in important ways. Instructions characterize much use, problem solving, repairing, and learning of technological tools as observable in helplines for users having problems with their computers, instructions about how to handle photocopy machines (Suchman, 1986), or about how to implement a new software. While the development of technologies such as sensory detecting and machine learning enables engineers to modelize and simulate patterns of human behaviors and actions, it is still not possible to automatically detect and analyze the aforementioned reciprocity that is typically negotiated in interaction and constructed without words. The aim of this workshop is to provide engineers with insights into how to systematically observe such reciprocity and find ways to apply to the fields of computer science.

(Excerpt from the submitted proposal)

# The Future of Human-Robot Spoken Dialogue: From Information Services to Virtual Assistants

# 057



Shonan Village Center, March 26–28, 2015

Seminar No.059

## Organizers

Rafael E. Banchs, Institute for Infocomm Research (I2R), Singapore  
Sakriani Sakti, Nara Institute of Science and Technology (NAIST), Japan  
Etsuo Mizukami, National Institute of Information and Communications Technology (NICT), Japan

Dialogue Systems embrace the ultimate goal of human-robot interaction, in which computational systems communicate with their human interlocutors in the same way humans communicate among them. Although significant progress has been achieved during the last few years and some pioneering commercial systems are finding already their way in the market, current state-of-the-art dialogue systems are very limited in their approach to the human communication phenomenon. Some of these limitations include:

- the lack of ability to properly model world knowledge for reasoning purposes,
- the still low reliability of speech recognition and the limited capacity of dialogue systems to cope with recognition errors in a logical manner,
- the complex and multimodal nature of the pragmatic phenomena and the socio-cultural aspects that affect such interactions, and
- the difficulty of evaluating dialogue quality and the subsequent development of reliable strategies for automatic learning and adaptation.

The main objective of this NII Shonan meeting is to discuss about the most relevant and promising future directions of research in dialogue systems. The discussion is to be centered on how these research directions address the different problems and imitations of current dialogue systems, as well as how they provide the basis for the next generation of intelligent artificial agents.

The assistants will be requested to present work in progress as well as to propose new collaborative efforts for tackling with the main limitations of current dialogue systems mentioned above. As a result of the meeting, a research agenda for main directions and international collaboration for the next three to five years will be defined.

Joint organization of future workshops and shared tasks by the participants in the meeting will be defined with the objective of pushing the state-of-the-art in dialogue systems to a more comprehensive effort for developing a new generation of intelligent virtual assistants.

# Dimensionality and Scalability II: Hands-On Intrinsic Dimensionality

# 058

Shonan Village Center, June 29–July 2, 2015

Seminar No.061

## Organizers

Laurent Amsaleg, CNRS-IRISA, France  
Michael E Houle, National Institute of Informatics, Japan  
Vincent Oria, New Jersey Institute of Technology, USA  
Arthur Zimek, Ludwig-Maximilians-Universität München, Germany



## Background

For many fundamental operations in the areas of search and retrieval, data mining, machine learning, multimedia, recommendation systems, and bioinformatics, the efficiency and effectiveness of implementations depends crucially on the interplay between measures of data similarity and the features by which data objects are represented.

When the number of features (the data dimensionality) is high, similarity values tend to concentrate strongly about their means, a phenomenon commonly referred to as the curse of dimensionality. As the dimensionality increases, the discriminative ability of similarity measures diminishes to the point where methods that depend on them lose their effectiveness. The effects of the curse of dimensionality on search and clustering methods are well-known and well-documented. Domain transformation strategies such as dimensional reduction and feature selection can improve performance to some extent, but the fundamental difficulties associated with high dimensionality nevertheless persist.

Over the past decade or so, new characterizations of data sets have been proposed for assessing the performance of particular methods. Such characterizations include estimations of distribution, estimation of local subspace dimension, and measures of intrinsic dimensionality of data. Although the applications affected by the curse of dimensionality vary widely across research disciplines, the characterizations and models of data that can be applied to analyze the performance of solutions are very general. Across the different disciplines, the data models and data characterizations that have been proposed are quite similar. Unfortunately, researchers from one domain are typically unaware of what researchers from other domains have developed.

We now propose a second NII Shonan Meeting on Dimensionality and Scalability, in order to (i) disseminate more widely the results obtained since 2013; (ii) leverage on the initial collaborations to consolidate research agendas; and (iii) obtain feedback from new researchers concerned with the general problems of dimensionality and scalability.

The detailed objectives of the meeting are:

1. Sharing our latest discoveries, so as to increase the visibility of our contributions among the community that has arisen since the meeting in 2013. With the new model of intrinsic dimensionality of continuous distance distributions, our theoretical perspective is now better established.
2. To foster collaborations among this new community of researchers. The first Shonan workshop fully succeeded in bootstrapping collaborations that resulted in the contributions presented during the one-day seminar in March 2014.
3. The third goal is to bring together new potential collaborators. Since the first Meeting, we as a community have gained a great maturity on the theoretical side as well as on the practical side. It will be obviously beneficial to present these more evolved ideas to those researchers concerned with some aspect of dimensionality and scalability who could not attend the first Meeting.

(Excerpt from the submitted proposal)



# Synchronous Machine Translation

# 059



Shonan Village Center, August 3–6, 2015

Seminar No.068

## Organizers

Jordan Boyd-Graber, University of Colorado Boulder, USA

Hal Daume III, University of Maryland, USA

Yusuke Miyao, National Institute of Informatics, Japan

Synchronous machine translation is the task of automatically translating a foreign language while it is being spoken. When a sentence is being produced one word at a time in a foreign language, the goal is to produce a translation in a target language (e.g., English) simultaneously: with as little delay between the uttering of a foreign language word and the production of its English translation. A requirement for successful synchronous machine translation is to be able to predict what words are going to appear in the input stream, before they have been seen. This is particularly important when translating from verb-final languages like Japanese or German, into verb medial languages like English. Without the ability to predict the yet-to-be-seen verb, the translator has to wait until the entire sentence is uttered before anything beyond the subject can be translated. This is not just an issue of final-verb prediction; any significant reordering involves prediction in order to maintain simultaneity: for instance relative clauses or postpositions in Japanese.

This is an important task to study because effective human interaction with technology requires low latency. If translation systems have their built in latency plus the latency of a sentence, there is no chance for translation systems to be used widely for everyday conversations. Creating systems that can provide low-latency translations will enable better communication across the divide of language and culture.

### Day 1: Existing Work

The first day of the workshop will highlight the existing work in this area conducted by invitees.

Simultaneous translation has been dominated by rule and parse-based approaches. Even projects like Verbmobil use a rule-based incremental decision-making module on top of a statistical translation module. Other recent approaches in speech-based systems focus on waiting until a pause to translate or using word alignments between languages to determine optimal translation units.

The organizers have recently developed a statistical approach that combines prediction of unseen elements and reinforcement learning to decide when to translation.

### Day 2: Evaluation and Data

Statistical machine translation has become a performance-driven enterprise because of the introduction of universal, consistent evaluation mechanisms like BLEU and TER. However, these metrics focus only on translation quality. Effective simultaneous translation must tradeoff between translation quality and latency.

### Day 3: Prediction

One of the primary challenges is that essential sentence elements (such as the verb) in the source language don't appear until well after they are needed in the target language. Thus, a required solution is to attempt to predict final verb. Day 3 will discuss probabilistic, supervised, and rule-based systems for predicting important aspects of source sentences.

### Day 4: Frameworks

The final day of the workshop will discuss frameworks that tie each of these pieces together into a complete system that make tradeoffs given an evaluation framework, predict unseen components, and produce effective simultaneous translations. The final day will also discuss the challenges of incorporating other systems such as speech recognition, speech production, and communicating simultaneous translations through captioning, voice-over, or other techniques.

(Excerpt from the submitted proposal)

# Engineering Adaptive Software Systems (EASSy)



Shonan Village Center, September 7–10, 2015

Seminar No.052

## Organizers

Tetsuo Tamai, Hosei University, Japan

Hausi Muller, University of Victoria, Canada

Bashar Nuseibeh, Open University, United Kingdom



It goes without saying that the modern society is founded on the information infrastructure, which is basically constructed with software, but its sustainability is now in question. Information systems will become obsolete and unusable if they cannot respond to requirements changes demanded by the users and their reliability and safety will be severely degraded if they cannot cope with their environment changes. The key issue here is how to engineer adaptive systems that conquer such sustainability threats. The proposed workshop, the third edition of its series, focuses on this issue of engineering adaptive systems. The technical goals are:

### Adaptation to environment changes

Software systems are deployed or embedded in various environments and those environments frequently change. Software, in spite of its name, often shows stiffness and inflexibility to changes. How to provide adaptability to software is an important and challenging goal.

### Quality requirements

When software obtains the property of adaptability, quality requirements, including those for dependability and usability, may be affected. It is crucial to monitor the current quality requirements satisfaction status and maintain the qualities at the required level. Specifically, topics to be discussed at the workshop include the following.

- How do we engineer adaptive software systems? What are the concepts, tools and techniques that can support requirements elicitation, architectural design and implementation of such systems?
- How do we reengineer legacy software systems in order to turn them into adaptive ones?
- Comparative review of adaptation mechanisms in Robotics, Multi-Agent Systems, Software Engineering, Socio-Technical Systems, Ubiquitous Computing, etc.
- Usability issues for adaptive software systems. How do we ensure effective human interaction with complex software systems that have adaptive components?
- Evolution of adaptive software systems. How do deployed adaptive systems evolve? How can we ensure convergence and stability for such a system, particularly when it is a system-of-systems composed of component systems, each with its own requirements and own adaptation mechanism?
- How do we reason with runtime models to support adaptation functions, i.e., monitoring, diagnosis and compensation? How can we support incremental runtime reasoning that predicts and/or prevents failures?

### Workshop Format:

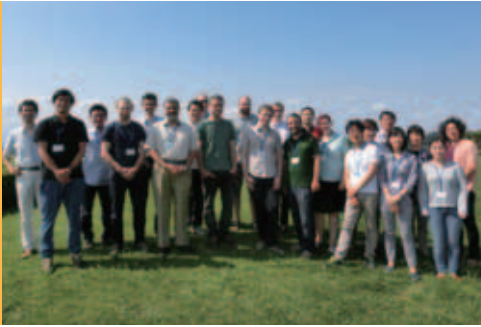
The four-day EASSy workshop will be held at the Shonan Village Center near Tokyo. 20 key participants will be invited to deliver position statements addressing a comprehensive approach to adaptive software system engineering. We envision sufficient time for discussion after each statement, e.g., 1 hour for each presentation slot, including 20-30 minutes for a statement, with 30-40 minutes of follow-up discussion. We plan to invite another 20 more junior participants who will contribute to and benefit from the discussions without having their own presentation slot.

We expect that in many cases these will be junior members of groups represented by a senior researcher in the workshop. As a deliverable for the workshop, we envision a follow-up special issue in an international journal, where workshop participants and the community-at-large will be invited to contribute original papers presenting comprehensive approaches to the engineering of adaptive software systems. The organizers hope that such a collection of papers will offer an authoritative account of the state-of-the-art on adaptive software systems and will guide future research in this fast-moving research field.

(Excerpt from the submitted proposal)

# Challenges for Real-time Human Activity Recognition

# 061



Shonan Village Center, September 14–17, 2015

Seminar No.066

## Organizers

Sozo Inoue, Kyushu Institute of Technology, Japan

Stephan Sigg, Aalto University, Finland

Over the last decade, driven by strong developments in related areas, we have seen remarkable progress in the recognition of human activities or situations. Sensing hardware greatly improved (e.g. size, accuracy and also new sensing modalities and sense-able quantities), enabling an enhanced perception of the world through sensors. Also, machine learning has celebrated tremendous successes (algorithms, toolboxes) and has become a mainstream ability that attracts a huge user base towards activity recognition. Furthermore, rapid development in wireless protocols and near-global coverage of wireless technologies (e.g. UMTS, LTE, Bluetooth, Wlan) enabled the transmission at higher data rates and new usage areas through wireless communication. Finally, novel applications have spread which promote publishing and sharing of all kinds of data (e.g. Facebook, Line, WhatsApp). This leads to novel valuable inputs for machine learning and activity recognition. With these opportunities, new challenges and application areas for activity recognition evolve.

Fueled by the advancing Internet of Things and by opportunistic as well as participatory sensing campaigns, activity recognition is going Big Data. New devices like fitness accessory, smart watches or instrumented glasses promise continuous collection of data for monitored subjects. Sensing systems are developing capabilities to monitor virtually everybody, everywhere and even device-free so that the necessity to equip individuals with sensing hardware or specifically installing system components at any particular physical location diminishes.

In addition, the perception of sensing devices is extended toward sentiment states. Sentiment sensing focuses on people's mental state, intention or emotion, for instance, by interpreting eye-gaze information, body gesture or pose. The perception of sensing systems is thereby extended and directed inwards. Such research questions extend into life sciences where aspects like, for instance, the perception of emotion via gesture and pose continues is considered for a number of years now.

At the same time, the management of privacy, security and trust in such sensing systems, although seminal for widespread acceptance, remains an open issue.

Furthermore, the tremendous amount of information which is continuously acquired by these sensing systems requires new methods and algorithms to deal with such Big Data. Currently ongoing projects allow a glimpse on the challenges related to the handling of such massive amounts of data, but the information available when subjects are continuously monitored by multiple devices at a time will exceed this greatly. This challenge of large scale database construction is also focused by the Human Activity Sensing Consortium (HASC6), where data collection, feature and algorithm development, as well as algorithm and tool standardization for human activity sensing are focused.

The purpose of this Shonan meeting is to bring together researchers from the fields of activity recognition, mobile sensing, mobile security, pervasive computing, information processing and sentiment recognition and management to discuss the technical challenges, possible societal impact, as well as promising industrial applications for emerging applications in activity recognition. The seminar puts a focus on Big Data in activity recognition, sentiment sensing and privacy and trust in continuous sensing platforms. The tackled problems like Big Data, Sentiment Sensing and privacy and trust in activity recognition constitute timely research questions in real-time activity recognition.

(Excerpt from the submitted proposal)

# Semantics and Verification of Object-Oriented Languages

Shonan Village Center, September 21–25, 2015

Seminar No.063

## Organizers

Atsushi Igarashi, Kyoto University, Japan

Nikos Tzevelekos, Queen Mary University of London, United Kingdom



Software applications are increasingly often desired, or even required, to come with guarantees as to their performance, correctness and reliability. Such guarantees have typically been of focal importance to safety-critical applications such as avionics, automotive systems and nuclear reactor controllers. Nowadays, with software being pervasive and having a key role in our daily lifestyles, the focus of the game has greatly broadened. Verified code has become a desideratum of software development in general: it concerns widespread applications written in mainstream languages. The verification problem for such languages poses new challenges as high-level programming features (references, objects, classes, etc.) need to be handled accurately. Moreover, realistic applications can expand to million lines of code, thus demanding accurate but scalable methods. Taking these parameters into account, we herewith propose a seminar devoted to modelling and verifying object-oriented code. The aim of the meeting will be to take stock of the present situation, to foster communication between researchers pursuing diverse approaches and to propose challenges for the future.

The object-oriented paradigm has been embraced by mainstream software companies as well as academic curricula. It has also been gaining ground in verification. Although, traditionally, verification has been conducted for constrained but technically robust programming languages, time has shown that there is a need to loosen the constraints in order to make comprehensive impact. Everyday software development primarily concerns mainstream languages, which provide a range of expressive features, greatly enhancing the development and maintenance of code. Moreover, they also attract a good supply of skillful developers. On the object-oriented front, Java has become widespread and the trend is complemented by the development of C# and Scala. Each of them offers safety compile-time and run-time checks, which are not present in other competitors like C/C++. They emerge as the natural choices to target with modelling and verification techniques, if one is interested in the analysis of safe object-oriented programs. Many efforts have been dedicated in recent years to provide techniques to address the issue. The vision of dependable/verified software has become the driving force behind numerous research projects in the UK and inspired one of the official UK CRC Grand Challenges for the next decade. The formal verification of object-oriented software has also been the subject of international initiatives, such as the European Concerted Research Action IC0701.

Object-oriented code presents a whole variety of interesting technical challenges, stemming from the fact that it involves a unique combination of more general computational concepts, such as recursive types, higher-order references, self-reference, state encapsulation, polymorphism etc. Each of them is a challenge in its own right and over the years dedicated research programmes have been developed to understand the features in isolation. The proposed seminar will provide a forum in which such neighbouring fields can be discussed from the point of view of object-oriented languages. We hope to identify opportunities for transfers of techniques with a view to specializing them to the specific setting of object-oriented code. As well as assessing the more established lines of work in the area, we will also aim to predict new fruitful directions, emerging from the constant evolution of OO-languages, the addition of new features, capabilities and structuring constructs.

(Excerpt from the submitted proposal)

# Validated Numerics Meets Reachability Analysis for CPS Design

# 063



Shonan Village Center, September 28–October 1, 2015

Seminar No.076

## Organizers

Daisuke Ishii, Tokyo Institute of Technology, Japan

Kohei Suenaga, Kyoto University, Japan

Walid Taha, Halmstad University, Sweden

The purpose of this meeting is to bring together researchers in several fields to develop a coordinated roadmap for computational tools for Cyber-Physical Systems; this meeting will bridge the gap between the hybrid systems community and the validated numerics community. Although both have developed methods for rigorous reasoning about dynamical systems, both faced challenges in making their tools available to practicing engineering. The meeting will focus on two key challenges: scalability and usability.

Cyber-Physical Systems (CPSs) consist of computers that are connected tightly to physical environments through sensors and actuators. Examples of CPSs include robots, smart homes, vehicles, medical implants, and sensor networks.

Mathematically, we can view CPSs as hybrid systems that exhibit both continuous and discrete changes. Many CPSs are also subject to real-time and reactive constraints. CPSs are an engineering, multi-disciplinary area that is rapidly gaining wide acceptance and traction in scientific, social, political, and commercial circles. Both the hybrid systems community and the validated numerics community have produced formal tools that are important for rigorous design of CPSs. Reachability analysis of hybrid systems was proposed by the first community as an extension of the model checking method in the context of verification involving continuous quantities. To handle reachable sets of real-valued states, the proposed tools either abstract them into a discrete representation or over-approximate them using numerical objects such as intervals and polytopes. Since the development of interval analysis in 1960's, validated numerics has been used by the second community to produce powerful tools for solving mathematical problems that are formally correct.

Unfortunately, CPS designers encounter some challenges with using both reachability analysis and validated numerics. In particular, formal methods often have a steep learning curve that hampers their adoption in industrial practice. It is therefore important to provide tool support for users that fills in gaps in the background knowledge of logic, algebra, real analysis, etc. However, users may still face difficulties in analyzing their problems with these tools. Because each tool depends on the underlying verification algorithm and the form in which hybrid systems are represented, these aspects restrict the tractable class and size of the problems. When a tool does not work immediately, a difficulty lies in what users can do with the tool and with various restrictions in its implementation such as linearity of arithmetic constraints, and support for the description of large systems. Similarly, validated numerics methods today exist mainly in the form of specialized libraries that are only accessible to experts in this domain. Although integration of the reachability analysis and validated numerics is necessary to push forward the rigorous CPS development, there is still an enormous gap between them. The central motivation for this Shonan meeting is the prospect that tools based on carefully designed, declarative, high-level modeling formalisms that are natural to the hybrid systems domain can help overcome these challenges. Such tools allow us to describe models and verification problems in a straightforward manner that is more easily usable by practitioners, and the interpreters transform the model and extract underlying subproblems to which scalable validated numerical methods can be applied. However, designing such tools requires close interdisciplinary cooperation between experts not only in language design but also in hybrid systems, reachability analysis, validated numerics, and practitioners. It seems particularly important and timely to connect the research communities in a way that can be as widely applicable across as many CPS domains as possible.

(Excerpt from the submitted proposal)

# 064 Mobile App Store Analytics

Shonan Village Center, October 19–22, 2015

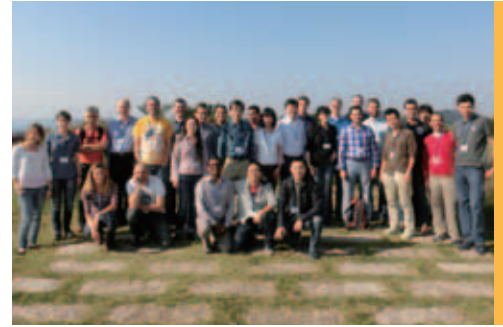
Seminar No.070

## Organizers

Meiyappan Nagappan, Rochester Institute of Technology, USA

Ahmed E. Hassan, Queen's University, Canada

Yasutaka Kamei, Kyushu University, Japan



Today, software engineering research focuses on traditional software systems like the Firefox web browser or Microsoft Windows, which take years to develop and teams of designers, developers and testers. Software engineering is rapidly changing though. Emerging domains, such as mobile devices, are growing rapidly and depend heavily on new software operating systems like Android and the applications that they run, commonly referred to as apps.

However, unlike traditional software, the distribution mechanism for mobile apps are very different; they are released through app markets. The key differentiating factor in an app store is that it is a democratic platform, i.e., both large companies with established products Adobe Reader from Adobe, and Timberman from Digital Melody (a company with 5 employees), can release their apps through the same mechanism for the users to download and install. The data that these mobile app markets contain can be used by software engineering researchers to compile new empirical results that can help mobile app developers.

Additionally, the app markets allow users to post reviews of the apps. This is very different from traditional software. Mobile app developers get continuous feedback from users that can be leveraged to help them. However, today the review system for mobile apps is identical to that of books sold on an e-commerce website such as Amazon. While books are products too, they are very different from mobile apps in that books are not updated every few weeks like most mobile apps. Hence the question arises whether the review systems of books is the best system for mobile apps or not?

Finally, the app stores provide a central location for all the apps, making it easy for researchers to mine the store for meta-data of the apps and the apps themselves. Using the data from the app stores, several companies like App Annie, and Distimo have even built successful businesses selling intelligence gained from observing the evolution of several hundred thousand apps in the app stores.

Although Software Analytics is gaining popularity over the last year, much of software engineering research today is still focused on traditional “shrink wrapped” software, such as Mozilla Firefox, Eclipse or Microsoft Windows. Recently however, researchers have begun to focus on mobile apps and the related software engineering issues. For example, the 2011 Mining Software Engineering Challenge focused on studying the Android mobile platform. Other work focused on issues related to code reuse in mobile apps, monetizing apps through ads, mining mobile app data from the app stores and teaching programming on mobile devices.

Even with all the above papers on mobile apps, there is no central venue to bring all the cross-disciplinary researchers together. There is therefore, a dire need for a community to be built around the line of research with respect to mobile app store analytics.

Hence, with a strong recent body of work like the one's stated above, and the lack of a venue to build a research community around the challenges and opportunities, now is the time for a Shonan meeting on the issue of 'Mobile App Store Analytics'. The proposed seminar would focus on research where the Mobile App Stores, and the data that they have are mined for insights into mobile app development. We intend to bring researchers in multiple disciplines from around the world in one place to discuss the future directions in the area of mobile app store analytics. Each of the researchers we intend to invite, has conducted research and published papers on deriving insights from mining mobile app stores.

(Excerpt from the submitted proposal)

# Logic and Verification Methods in Security and Privacy

# 065



Shonan Village Center, October 26–29, 2015

Seminar No.069

## Organizers

Marco Gaboardi, University of Dundee, United Kingdom

Vivek Nigam, Federal University of Paraiba, Brazil

Tachio Terauchi, Japan Advanced Institute of Science and Technology (JAIST), Japan

Logic and verification methods are standard tools useful for reasoning in a formal way about programs. Tools and techniques based on logic have been used to prove software correct with respect to different requirements: functional correctness, resource consumption, integrity of data, etc. An area where the usefulness of these methods was observed early on is the one of security and privacy. Formal logic methods have been used for formally specifying security and privacy policies and models, and for verifying that programs respect them.

Recently however traditional notions of security and privacy have been extended to take into account new refined aspects of programs. In particular, we assisted at the extension of traditional security properties to notions that are quantitative or probabilistic in nature, and to notions that require complex policies models based on different notions of capabilities or authorizations. Quantitative information flows, differential privacy and distance-bounding protocols are three important examples of such security properties. Quantitative information flows is a natural quantitative extension of the traditional information flow theory where the information that is leaked by a program is measured in terms of notions like min-entropy and leakage. Differential privacy is a strong statistical notion of data privacy requiring that the result of a data analysis is probabilistically almost the same in the case the individual participate in the data or not. Distance-bounding protocols are cryptographic protocols where it is critical to control also the physical distance between the parties involved. These three properties are emblematic examples of the kind of security and privacy requirements that software is required today to ensure.

Verifying and reasoning in a formal way about this kind of security and privacy properties require extensions of traditional techniques that are able at the same time of describing more refined formal models—accounting for the quantities and/or capabilities specific to these properties—and more refined proof techniques designed specifically for these formal models. In order to deal with this supplementary complexity of the security and privacy models, several techniques have enriched traditional logical methods with quantitative information. These techniques have provided new theoretical foundations and practical successful tools.

Part of this success is certainly to ascribe to the advancements in the technology of SMT solvers and interactive proof assistants. These advancements have permitted to reach results that only few years ago seemed impossible to reach. SMT solvers and proof assistants can be nowadays easily integrated in tools for security and privacy analysis, and serve as basic component for the resolution of numeric and/or symbolic constraints relative to quantities and/or capabilities. However, the development of more and more specialized techniques require also a more and more customized use of these tools. For this reason it is important to promote the interaction between the members of the security and privacy community and the community working on verification tools.

The overall goal of the meeting is to foster the discussion between researchers that are working in different areas of security and privacy, logic and verification. The common ground between the different participants will be the use of logic and verification methods for formally reasoning about the different aspects of security and privacy. On the application side, we aim at fostering the discussion around the different tools that are needed to reason about traditional and quantitative notion of security and privacy. On the theoretical side instead, we aim at fostering the discussion around common foundations for the different aspects of security and privacy. A further goal of the meeting is exploring the applicability of the most recent techniques developed in the setting of security and privacy to problem in different research areas.

(Excerpt from the submitted proposal)

# Big Data Visual Analytics

Shonan Village Center, November 8–11, 2015

Seminar No.064

## Organizers

Seok-Hee Hong, The University of Sydney, Australia

Koji Koyamada, Kyoto University, Japan

Kwan-Liu Ma, University of California, Davis, USA



## Summary

High-throughput technologies have produced Big Data in many application domains in Science and Engineering including Biomedical Engineering, Genomics, Software Systems, Computer Networks, Finance, e-commerce, Cyber intelligence, and Homeland Security. The ability to analyze such Big Data for knowledge discovery and decision-making is critical to scientific advancement, business success, clinical treatments, cyber and national security, and disaster management.

Visual Analytics is the science of analytical reasoning supported by interactive visual techniques, which requires interdisciplinary science integrating techniques from visualization and computer graphics, statistics and mathematics, data management and knowledge representation, data analysis and machine learning, cognitive and perceptual sciences.

The main goal of the workshop is to promote Visual Analytics research in Asia-Pacific region, and form a research community to collaboratively solve complex problems arising in a variety of application domains. In particular, special emphasis on the “Big Data” will be addressed.

## Aims and Objectives

This meeting aims to bring world-renowned researchers on Visual Analytics and collaboratively develop innovative scalable Visual Analytics solutions to solve the scalability and complexity issues for analyzing Big Data arising from various application domains including Systems Biology, Social Networks, Finance, Business intelligence, and Security.

## Our specific objectives are

- We will identify research opportunities in Big Data Visual Analytics, focusing on the Asia-Pacific context.
- We will form a broader research community with cross-disciplinary collaboration, including computer science, information systems, statistics, biology and sociology, with a focus on Visual Analytics of Big Data.
- We will foster greater exchange between visualization researchers and practitioners, and to draw more researchers in the Asia-Pacific region to enter this rapidly growing area of research.
- We will assist emerging researchers to find linkages to international researchers, industrial contacts, and competitive research grants and fundings.

## Significance and Innovation

Big Data Analytics is the biggest and fundamental challenge in IT research due to Scalability and Complexity. Innovative scalable techniques for Big Data Visual Analytics will be the key enabler for researchers and end users in many application domains and other disciplines.

## Expected Outcomes

- Innovative techniques and solutions for Big Data Visual Analytics, which will be used by domain experts and end users in various application.
- Joint publications at the top conferences and journals in Visualization, jointly authored by researchers in the area of Visual Analytics.
- Joint funding applications for long-term research collaboration for continuation of the research collaboration beyond 2015.

(Excerpt from the submitted proposal)



# Putting Heterogeneous High-Performance Computing at the Fingertips of Domain Experts

# 0067



Shonan Village Center, November 17–20, 2015

Seminar No.075

## Organizers

Wim Vanderbauwhede, University of Glasgow, United Kingdom

Sven-Bodo Scholz, Heriot-Watt University, United Kingdom

Tetsuya Takemi, Kyoto University, Japan

## Background

High-performance computing (HPC) is critically important in many scientific fields that use numerical models which are computationally intensive both in terms of the speed of computation and in the memory usage.

Writing code that will run with good performance on modern computer systems is becoming increasingly hard because of the advent of many-core systems and an increasingly broad range of accelerators such as Graphics Processing Units (GPUs) and Field- Programmable Gate Arrays (FPGAs).

HPC clusters built from such systems pose an even greater challenge. Not only has the information area that we are embarking into led to new increased data processing demands, sometimes referred to as “big data”, but we also see heterogeneous HPC systems becoming mainstream facilities. As these systems become more affordable they become more widely available.

The compilers and analysis tools required to achieve high performance on modern systems are highly complex in themselves, and aimed at the HPC experts rather than the domain experts. However, many domain experts do not have the means to employ HPC experts, nor can they afford the time to acquire the necessary expertise themselves. As a result, the gap between the performance of code written by many domain experts and the capability of modern computer systems is growing steadily. Even for the HPC experts it is becoming increasingly difficult to achieve optimal performance, due to the huge complexity of heterogeneous many-core systems.

To safeguard the progress of the domain experts and, with it, the progress of scientific research relying on numerical computations, addressing this performance gap is crucial.

## Bridging the gap

The current work flow for scientific computing is typically as follows: domain experts typically write single-threaded code, usually in Fortran or C/C++. If they have in their team parallel programming expertise, or the means to afford support from HPC experts, the code will be parallelised for clusters through MPI and for multicore processors via OpenMP. Porting the code to GPUs requires manual rewriting of parts of the code in CUDA or OpenCL. For FPGAs, the situation is even more complex: here, the code needs to be re-implemented in a hardware design language.

At the same time, there are many efforts in the computing science community to create languages and compilation approaches that can target heterogeneous systems without manual rewrites as well as languages with explicit parallelism support. However, this research assumes knowledge about programming paradigms, architectures, cost models etc. which come natural to computing scientists and HPC programmers but not to domain experts.

We therefore want to bring these communities together to exchange views, so that the computing science research will benefit the domain experts much more directly.

## Aim of the meeting

The aim of this Shonan meeting is to bring together researchers from the disciplines involved, in particular:

- domain experts such as geophysicists, meteorologists etc.,
- High-Performance Computing experts,
- computing scientists with expertise in programming languages and compilers for heterogeneous many-core systems,
- specifically FPGA and GPU experts

To have a discussion on the challenges each community faces and on ways to bridge the gap between the domain experts and today’s and tomorrow’s clusters of heterogeneous many-core systems.

We want to address questions relating performance to code analysis, refactoring, compilation and run-time adaptation, as well as user experience design.

(Excerpt from the submitted proposal)

# Analytics on Complex Networks: Scalable Solutions for Empirical Questions

# 068

Shonan Village Center, February 8–11, 2016

Seminar No.072

## Organizers

George Fletcher, Eindhoven University of Technology, The Netherlands

Taro Takaguchi, National Institute of Informatics, Japan

Yuichi Yoshida, National Institute of Informatics, Japan



Real-world networks, such as web graphs, social networks, and biological networks, have remarkable topological features in common: the scale-free property, that is, the degree distribution obeys a power-law function, and the small-world property, that is, the average distance between two vertices is small. These networks are collectively referred to as complex networks and have been intensively studied in the last decade.

In their empirical study of complex networks, researchers across the fields of data engineering, theoretical computer science, and network science are developing solutions for analytics. Due to rapid growth of the sizes of real-world networks, however, we are facing challenges in scaling these tools, as we will describe below. We desire to solve these issues by bridging the knowledge in the three fields with the single key notion: complex networks.

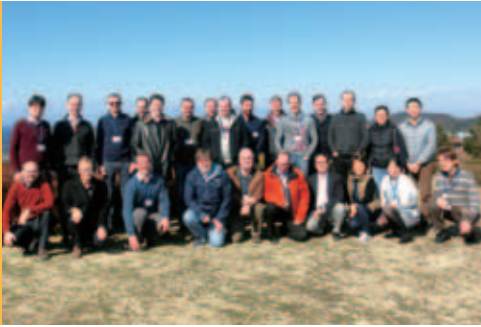
Researchers in the engineering of data-intensive systems, such as database and data mining systems, have traditionally focused on the study of scalable solutions for analytics on big data collections. With the increasing availability and broad interest in massive complex networks, researchers in data engineering have hence focused considerable effort in recent years on scalable mining and querying over such data collections. However, very little work has been done on coordinating the results of these investigations with the foundational work in this area in the theoretical computer science and network science communities. Stronger ties between these communities would help deepen the understanding and development of theoretically grounded analytics solutions for the empirical investigations of interest in the broader scientific community.

From the perspective of theoretical computer science, the important point is that we need algorithms for real-world networks but not for all the possible networks. The majority of theoretical studies on graph problems pay attention to worst-case time complexity or average-case time complexity, which may have little to do with real-world networks. On the other hand, practical algorithms on graph problems proposed in the fields of databases, data mining, and machine learning have shown their efficacy on real-world networks, but these algorithms lack theoretical groundings to understand why they work well in practice. By applying theoretical analysis frameworks to these algorithms, we want to answer foundational questions such as “what is the right time complexity of this particular problem when the given graph is a complex network?” and “why does this particular heuristic make the algorithm faster or more accurate when the given graph is a complex network?”

Network science, a new subfield of statistical physics seeking universality hidden behind real-world networks, can move on to a new stage with aids provided by data engineering and theoretical computer science. For example, a large number of models describing growth processes of complex networks and information spreading on them have been studied in network science. However, their computational properties are largely unexplored, which prevents us from discussing their properties with rigorous theoretical groundings. As another example, various kinds of vertex centralities, which measure the importance of vertices, have been proposed and applied to network data. However, the majority of these centralities are computationally hard and efficient (approximation) algorithms should be studied to apply them to real-world networks at a large scale. In theoretical computer science, strong analytical tools for graphs such as graph minor theory and spectral graph theory have been developed and these theories may shed new light on novel characteristics of complex networks.

As we have seen, complex networks are studied in diverse fields from different perspectives. The main goal of this workshop is bringing researchers in these fields together to take the first steps towards bridging works on scalable analytics over complex networks.

# Immersive Analytics: A New Multidisciplinary Initiative to Explore Future Interaction Technologies for Data Analytics



Shonan Village Center, February 15–18, 2016

Seminar No.074

## Organizers

Takayuki Itoh, Ochanomizu University, Japan  
Kim Marriott, Monash University, Australia  
Falk Schreiber, Monash University, Australia  
Uwe Wössner, University of Stuttgart, Germany

## Introduction

There is a tidal wave of big and complex data across many domains. Several research directions are concerned with the development of methods to support the analysis of such data, including machine learning, information and data visualisation, data analytics and human computer interaction. However, there is still a major gap to fill: how can we immerse analysts in the data in order to provide more natural ways for data exploration, data analysis and collaboration?

In recent years we have seen significant advances in the development of new technologies for human-computer interfaces. Immersive environments like a CAVE make use of ultra-high resolution technology, and combine 2D and 3D visualisations to allow users to immerse themselves into computer generated scenes. Driven by the entertainment industry, tools and methods like voice- and gesture-based control as well as 3D VR and AR visualisation are becoming more and more part of everyday life.

In response to these new technologies, we propose a new facet of data analytics research that we call "Immersive Analytics". This new research direction explores how these new display and interaction technologies can be used to create more engaging experiences and seamless work-flows for data analysis applications. The environments we envision that will be developed by immersive analytics researchers should be usable by experts and analysts to help in the detailed analysis of complex data sets, but will also be accessible to decision makers, that is, the managers who spend more time working face-to-face than in front of a desktop computer, and to the everyday public to help them in tasks like balancing nutrition, social and ethical concerns and price when buying groceries.

The ability to move inside the data with new display devices, to interact with the data representation in various direct and embodied ways allows the user to create a richer experience that may lead to better understanding and finally deeper insight. For that to happen, the visualisation and interaction concepts need to allow a faithful representation of the data and provide intuitive and efficient orientation and navigation in the data space. We expect that to create productive tools many gaps will need to be filled.

Figure: "Immersing people in their data". In this concept, analysts work in a purpose-built, collaborative data analytics room. Interaction technologies like pen, touch and gesture control allow them to interact directly with their data and collaborate in a more egalitarian way than keyboards and mice, which tether individual users to a desktop. Immersive Analytics will systematically research how many kinds of emerging interaction technologies can be harnessed to engage and enable people to work together to better understand data.



## Topics and Aims of the Meeting

Immersive analytics is inherently multi-disciplinary and involves researchers from human computer interaction, data sciences, visual analytics, augmented reality, data and information visualisation. Immersive environments for the analysis of complex data sets can be used on several levels including fundamental research, e.g., in application areas like biology, information security, or social sciences. We will bring together experts in these fields to define the new research direction by clarifying the major research challenges, potential approaches and identifying potential applications. Based on these we will compile a survey and manifesto defining the topic. We would ideally aim for this to be published as a book or special journal issue.

The overarching goal of immersive analytics research is to understand how (and whether) new interface and display technologies can be used to create a more immersive kind of data analysis and exploration. The kinds of devices and environments include augmented and virtual reality displays, large high-resolution 2D and 3D displays, haptic and audio feedback, and gesture and touch controlled interaction. These potentially require very different interaction and visualisation models and techniques to those used in standard visual analytics.

(Excerpt from the submitted proposal)

# Mining & Modeling Unstructured Data in Software —Challenges for the Future

Shonan Village Center, March 7–10, 2016

Seminar No.084

## Organizers

Sonia Haiduc, Florida State University, USA  
Takashi Kobayashi, Tokyo Institute of Technology, Japan  
Michele Lanza, University of Lugano, Switzerland  
Andrian Marcus, University of Texas at Dallas, USA



To analyze, comprehend, and reverse engineer software projects and their software development processes, developers rely on various sources of information. Bug reports, execution logs, mailing lists, code review reports, change logs, requirements documents, and the actual source code contain implicit developer knowledge about the project and past development efforts. Most of this knowledge is captured as unstructured information, that is, natural language text used to exchange information among people.

Researchers in the Information Retrieval, Data Mining, and Natural Language Processing fields have experimented with various techniques and ad-hoc approaches to enable the mining of unstructured data from software artifacts. However, these techniques were not designed to work with the complexities and peculiarities of unstructured software engineering data, and thus are not readily applicable to the software engineering research domain.

The challenges for both researchers and practitioners are to determine the appropriate set of techniques to tackle the problem at hand and to understand how to use them effectively.

The meeting thus aims to address the following topics (but is not limited to):

- 1) Applications of unstructured data mining techniques to support software maintenance, software reverse engineering tasks (e.g., feature location, traceability), and for enhancing software quality;
- 2) Novel sources of unstructured data, such as mobile app stores, phone records, screenshots, interviews, or wiki pages;
- 3) Usage of NLP, IR, and ML techniques for mining unstructured data;
- 4) Classification and dissemination of techniques for extracting unstructured data;
- 5) Identification of open research challenges and proposed solutions;
- 6) Approaches for handling imperfect data, such as summarization approaches;
- 7) Novel extractors for unstructured data and performance evaluation with respect to existing techniques;
- 8) Linking of unstructured and structured data for richer information;
- 9) Negative results (“what did not work”) when mining unstructured data, and experience reports;
- 10) Large-Scale mining of Unstructured Data in Big Data environments;

We aim to facilitate in-depth discussions of techniques for mining unstructured data, their similarities and differences, applications in modern Data Mining, as well as potential pitfalls and problems. The intended outcomes of this meeting are to:

- 1) Facilitate knowledge-exchange in the field of mining unstructured software data and practical applications of techniques through presentations of short (2-page) paper submissions.
- 2) Establish connections between the various research communities that mine unstructured data, resulting in cross-fertilization of techniques and methodologies.
- 3) Put techniques and methodologies for mining unstructured data in a common framework, enabling researchers and practitioners to find the appropriate tools that meet their particular data mining needs.
- 4) Identify open problems and challenges for mining unstructured data, providing the basis for a roadmap of future research opportunities in mining unstructured data research.
- 5) Educate on, discuss, and advance the state-of-the-art in mining unstructured data.

(Excerpt from the submitted proposal)

# Higher-Order Model Checking

# 071



Shonan Village Center, March 14–17, 2016

Seminar No.078

## Organizers

Naoki Kobayashi, The University of Tokyo, Japan

Luke Ong, University of Oxford, United Kingdom

Igor Walukiewicz, CNRS, LaBRI, Bordeaux University, France

Finite state model checking has been widely studied and successfully applied to system verification. The main theme of this meeting, higher-order model checking, is a generalization of finite state model checking, obtained by replacing finite state models with more expressive models called recursion schemes. Higher-order model checking has found applications in analysis of object-oriented and concurrent programs with recursion and higher-order procedures.

Recursion schemes are a kind of simply-typed grammar for generating possibly infinite ranked trees. A recursion scheme is a finite system of equations, defining a finite set of higher-order functions by mutual recursion. The order of a recursion scheme is given by the highest type-theoretic order of the functions defined by it. From a programming language perspective, recursion schemes may be viewed as programs (i.e. closed, ground-type terms) of the simply-typed lambda calculus with recursion, constructed from a set of uninterpreted function symbols. Higher-order model checking is the model checking of trees generated by recursion schemes. The higher-order model checking problem asks, given a recursion scheme  $G$  and a correctness property  $\phi$ , whether the tree generated by  $G$  satisfies  $\phi$ .

The last fifteen years or so have seen a growth of interest in higher-order model checking from both the theory and practice communities. One recent trend, initiated by Salvati and Walukiewicz, is the development of effective denotational semantics for an approach to higher-order model checking by evaluation.

Recursion schemes are a very expressive family of generators of trees: the trees that are generated at orders 0, 1 and 2 are respectively the regular trees, algebraic trees (i.e. those generated by context-free tree grammars) and hyperalgebraic trees. There have been advances in the understanding of the relationship between higher-order families of generators. Parys proved the Safety Conjecture, thus confirming the intuition that the safety constraint restricts the expressive power of recursion schemes, or equivalently that order- $n$  collapsible pushdown automata are more expressive than order- $n$  pushdown automata.

Recursion schemes are thus an appealing abstract model for model checking higher-order programs: not only are they highly expressive, the trees they generate also enjoy a rich and decidable logical theory. Although the worst-case complexity of higher-order model checking is hyper-exponential, there have been remarkable advances in the design of model checking algorithms.

## Topics

1. Extensions of HOMC: beyond simple types (e.g. untyped and recursively typed recursion schemes, higher-type Böhm trees); beyond omega-regular properties: finitary parity, omega-B, stack unboundedness, etc.
2. “Practical” HOMC algorithms. E.g. HORSATZDD, Preface, C-SHORE, etc.
3. Applications of HOMC: program verification (functional, object-orientation, concurrency), program analysis, data compression, security, etc.
4. Higher-order grammars and pushdown automata. Context sensitivity of unsafe Maslov languages and other open problems.
5. Effective denotational semantics and strategy-aware models for HOMC. Compositional approaches to HOMC.

(Excerpt from the submitted proposal)

# Architecture-Centric Modeling, Analysis, and Verification of Cyber-Physical Systems

Shonan Village Center, March 21–24, 2016

Seminar No.073

## Organizers

Shin Nakajima, National Institute of Informatics, Japan  
Jean-Pierre Talpin, INRIA, France  
Masumi Toyoshima, Denso Corporation, Japan  
Huafeng Yu, Toyota Info Technology Center, USA



## Context

The term cyber-physical system (CPS) was introduced by Helen Gill at the NSF referring to the integration of computation and physical processes. In CPS, embedded computers and networks monitor and control the physical processes, usually with feedback loops where physical processes affect computations and vice versa. The principal challenges in system design lie in this perpetual interaction of software, hardware and physics.

CPS safety is often critical for society in many applications such as transportations, power distribution, medical equipment and tele-medicine. The development of reliable CPS has become a critical issue for the industry and society.

In an architectural model, a CPS is represented by a distributed system as entities with well-defined interfaces, connections between ports on component interfaces, and specifies component properties that can be used in analytical reasoning about the model. Models are hierarchically organized: each component can contain another sub-system with its own set of components and connections between them.

An architectural specification serves several important purposes. First, it breaks down a system model into manageable components to establish clear interfaces between components. In this way, complexity becomes manageable by hiding details that are not relevant at a given level of abstraction. More importantly, an architectural model is a repository to share knowledge about the system being designed. This knowledge can be represented as requirements, design artefacts, component implementations, held together by a structural backbone.

Since all the models are generated from the same source, the consistency of assumptions w.r.t. guarantees, of abstractions w.r.t. refinements, used for different analyses becomes easier, and can be properly ensured in a design methodology based on formal verification and synthesis methods, using notions such as formally defined interfaces and contracts.

In most cases, however, quantitative reasoning in architecture modelling and CPS is predominantly parameterised by the dimension of time. In each of the viewpoints that an architecture or CPS model refers to: software, hardware, physic, time takes a different form: continuous or discrete, event-based or time-triggered. It is therefore of prime importance to mitigate heterogeneous notions of time to support quantitative reasoning in system design, either using a tentatively unified model for it, or by formalizing abstraction/refinement relations from one to another in order to mitigate heterogeneity.

## Purpose of the seminar

The proposed seminar will bring together researchers who are interested in defining precise semantics of an architecture description language, using it for building tools that generate analytical models, for the purpose of simulation and formal verification (simulation, components integration, requirements validation) code generation (interfaces, orchestration, scheduling).

To formally define or semantically interpret architectural models, we observe a significant gap between the state of the art and practical needs to handle evolvingly complex models. In practice, most approaches cover a limited subset of the language and target a small number of modelling patterns. A more general approach would most likely require semantic interpretation (an abstraction, a refinement) of the source architecture model by the target analytic tool, instead of hard-coding semantics and patterns into the model generator.

(Excerpt from the submitted proposal)

# Big Data: Challenges and Opportunities for Disaster Recovery

# 073



Shonan Village Center, March 28–31, 2016

Seminar No.082

## Organizers

Sanjay Madria, Missouri University of Science and Technology, USA  
Takahiro Hara, Osaka University, Japan  
Cyrus Shahabi, University of Southern California, USA  
Calton Pu, Georgia Tech, USA

The main purpose of this NII Shonan meeting is to bring together researchers from the multidisciplinary fields of data management and analytics; mobiles, sensors and pervasive computing; geography and urban-panning; and disaster response and recovery with public agencies and commercial entities towards using big data for better decision making and problem solving in the event of a disaster. To do so, we need to close the gaps between those who collect the data (data providers), those who could benefit from using the data (domain experts), and those who are capable of developing the methods for storing/managing/processing the data (technology enablers).

So-called “Big Data” began when the Enterprise era generated the first wave of data through various software applications such as inventory management or human resource applications. Soon the field of computer science realized that there were commonalities in how the data was being stored and accessed, which led to the development of databases. However, the second wave of data, Human-generated data (the Web), exposed the fundamental challenges resulting from data heterogeneity (Variety). The rapid growth of web applications left academics with little opportunity to identify commonalities of data usage, leading to many independent tools that focus on a narrow aspect of data preparation for a given application type and requiring human in-the-loop data extraction and preparation. This worked to some extent, as human data creation processes led to a natural gap between data generation and data consumption. Machine generated data represents the newest wave as they are generated continuously at a high rate (Velocity) from various sensors in the physical world, starting with sensor instrumentation. These three waves of data gave rise to numerous approaches benefiting from data use in critical decision-making (Big Data).

The time is ripe to embark on a fundamental approach to Big Data challenges by assembling stakeholders to review case studies, design and develop several prototypical end-to-end systems, identify the commonalities, and develop lessons learned stories. This is exactly the goal of our proposed meeting with a focus application of disaster response and recovery. This is because efficient and thorough data collection and its timely analysis are critical to any disaster response and recovery system in order to save people’s lives during disasters. However, access to comprehensive data in disaster areas and their quick analysis to transform the data to actionable knowledge are major data science challenges. Moreover, the effective presentation of the collected knowledge to human decision-makers is an open problem. Therefore, the proposed meeting is to study and share experiences in Big Data research, Education and Training as well as discuss challenges and disseminate solutions, blueprints, and prototypes focusing on the disaster recovery application domain.

Towards this end, experts from various disciplines, including application domain experts with knowledge about disaster response and recovery, need to interact and collaborate effectively. The purpose of this workshop is to bring together these experts with the common goal of improving disaster recovery through Big Data, from different countries to Japan to initiate information exchange and collaboration. Moreover, the shortage of a knowledgeable workforce presents a further challenges to the Big Data management and needs to be addressed through education and training.

(Excerpt from the submitted proposal)

# Current Trends in Combinatorial Optimization

# 074

Shonan Village Center, April 11 – 14, 2016

Seminar No.071

## Organizers

Takuro Fukunaga, National Institute of Informatics, Japan

R. Ravi, Carnegie Mellon University, USA

László Végh, London School of Economics, United Kingdom



## Description of the meeting

Combinatorial optimization is a field in the intersection of applied discrete mathematics, operations research, and computer science. The goal in the combinatorial optimization is design of efficient algorithms for computing a good quality solution to problems on discrete structures whose solution space is subject to combinatorial explosion. It has numerous applications in various fields of science, and is regarded as a fundamental technological advance in processing data for optimal solutions and structures. Clearly its importance is increasing with the growing applications of data science, and these algorithmic methods are also gathering much public attention nowadays in fields as diverse as online dating to package routing.

The theory of combinatorial optimization has recorded remarkable developments in the past fifty years. However, many fundamental questions still remain unsolved. An example of such questions is the one related to the traveling salesperson problem (TSP). This is the problem of computing a shortest tour visiting all the cities in a map, where the distance between pair of cities is given as input. Since the number of tours corresponds to the number of cyclic permutations of the cities, it is an exponentially large number and hence a victim of combinatorial explosion. Nevertheless, for the TSP with distances that obey the triangle inequality, Christofides presented an elegant 1.5-approximation algorithm in 1976. This method is guaranteed to output a tour of length at most 1.5 times the shortest length of arbitrary feasible solution in time that grows with a polynomial dependence on the number of cities. One of the most popular current questions in combinatorial optimization asks whether there is a polynomial-time  $4/3$ -approximation algorithm for metric TSP. In 2011, several groups made progress for a special case defined by the shortest path hop distances in undirected graphs. They gave approximation algorithms that achieve approximation ratios better than 1.5 for this special case. However there is still no known algorithm that outperforms Christofides' algorithm in instances with arbitrary metrics.

Another example is the bin packing problem. In the bin packing problem, we are given a set of items whose size are specified by real numbers in  $[0,1]$ , and we are asked to pack the items into the smallest number of bins of size one. Several groups of researchers proved that various greedy algorithms achieve constant approximation for the problem in 1980's. Then, in 1982, Karmarkar and Karp presented an algorithm that computes a packing using at most  $OPT + O(\log^2 OPT)$  bins, where  $OPT$  denotes the number of bins used by the optimal solutions. Their algorithm is based on a linear programming relaxation of the problem formulated by Gilmore and Gomory in 1961. Their analysis is built on a deep understanding on the structure of the relaxation. Since their work, there was no significant progress for three decades although it is conjectured that the bin packing problem has an algorithm that computes a packing with  $OPT+1$  bins. In 2013, Rothvoß improved the algorithm of Karmarkar and Karp. His algorithm computes a solution with  $OPT + O(\log OPT \log \log OPT)$ . He improved the rounding phase of method of Karmarkar and Karp using recently developed algorithmic tools from discrepancy theory.

Besides these examples, we have numerous combinatorial optimization problems for which technical breakthrough are expected. To advance development of combinatorial optimization further, wide collaboration among researchers is necessary. The purpose of this meeting is to stimulate communication among researchers who are active in this field. To enable this, we plan to invite leading and active researchers, and provide an opportunity to share the most recent information on the new findings in this field.



# Controlled Adaptation of Self-adaptive Systems (CASaS)

# 075



Shonan Village Center, April 25–28, 2016

Seminar No.077

## Organizers

Paola Inverardi, University of L'Aquila, Italy  
Nicolás D'Ippolito, Universidad de Buenos Aires, Argentina  
Kenji Tei, National Institute of Informatics, Japan

Self-adaptive systems are required to adapt its behavior in the face of changes in their environment and goals. Such a requirement is achieved by developing self-adaptive systems as closed-loop systems following a Monitor Analyse Plan Act (MAPE) scheme. MAPE loops are a mechanism that allows systems to monitor their state and produce changes aiming to guarantee that the goals are met. In order to achieve the desired goals, self-adaptive systems must combine a number of MAPE loops with several responsibilities and abstraction levels.

Higher-level goals require decision-level mechanisms to produce a plan in terms of the actions to be performed. Several mechanisms can be adopted for automatically generating decision-level plans (e.g. event-based controller synthesis) providing guarantees on the satisfaction of hard goals, and improvements on soft goals.

Lower-level goals, on the other hand, require control mechanisms that sense the state of the environment and react to it thousands of times per second. Standard solutions to this problem are based on classical control theory techniques such as discrete-time control.

Indeed, system's requirements cannot be achieved by isolated MAPE loops, on the contrary, a combination of coordinated low-level and decision-level MAPE loops must be considered. Hence, multiple layers of control combining low latency discrete-time controllers with decision-level event-based controllers are a sensible approach to guarantee satisfaction of sophisticated goals in self-adaptive systems.

However, self-adaptive systems community has neglected the use of control theory as a central tool to guarantee hard goals while aiming to better satisfy soft goals. Very recently, software engineering community has started to study the application of control theory and the formal guarantees it provides in the context of software engineering. The Shonan meeting we propose focuses on formal guarantees that can be provided in self-adaptive system via the use of the broad area of control theory (e.g. event-based controller synthesis, discrete-time control).

In this meeting, we plan to involve a group of very active researchers in key areas such as Self-Adaptive Systems, Control theory, Game theory, Software Engineering, and Requirements Engineering, setting an adequate environment to discuss current and future applications and possibilities of control theory as a mechanism to provide formal guarantees on self-adaptive systems (e.g. convergence, safety, stability). We expect that having a number of participants from a wide variety of research areas will bring to light the benefits of incorporating the application capabilities and formal framework provided by control theory to self-adaptive systems.

Our meeting focuses on the central role of control theory in providing different level of assurance to self-adaptive systems. In this context, we plan to explore the possibilities of interdisciplinary collaboration between control theory, self-adaptive systems and requirements engineering experts. In order to take most of the meeting and facilitate deeper discussions, we plan to have a good mix of participants coming from these research areas.

We envisage the meeting to be organized in two main parts. First, a “stating common grounds” part where participants will give a talk stating their perspectives on their fields and key aspects where they visualize relations between their research area and other participants' fields. Then, a second part “discussions and work” where the most relevant topics selected by the participants will be discussed in work groups. Finally we plan to summarize the workshop findings and propose future actions.

We expect the meeting to encourage participants from different areas and communities to interact and actively search for common interests that may result in possible collaborations and joint work.

(Excerpt from the submitted proposal)

# 076 Theory and Applications of Geometric Optimization

Shonan Village Center, May 30–June 2, 2016

Seminar No.079

## Organizers

Siu-Wing Cheng, The Hong Kong University of Science and Technology, Hong Kong

Yoshio Okamoto, The University of Electro-Communications, Japan

Otfried Cheong, KAIST, Korea



There are many algorithmic, geometric, combinatorial, and implementation challenges that arise in geometric optimization. Such challenges have prompted relevant research and progress in algorithm design and analysis. Geometric optimization has wide applications in and connections to various areas, including computer graphics, computer-aided design and manufacturing, robotics, computer vision, spatial databases, geographical information systems, machine learning, and scientific computing. A lot of the theoretical results in geometric optimization fall within the area of computational geometry, which is a vibrant and mature field of research, with a flagship annual international conference and several dedicated international journals. Some of the results also appear in theoretical computer science conferences and journals. This meeting will focus on current interests and future trends in geometric optimization.

For example, data analysis is a popular research topic. The support vector machine paradigm has successfully linked many data analytics problems to rigorously defined convex optimization problems, which have a strong geometric flavor. Indeed, finding a sparse approximate solution to such a convex optimization problem is closely related to the concept of core-sets in computational geometry. It has also been discovered that core-sets are strongly related to the well-known Frank-Wolfe greedy optimization method. The problem of analyzing massive data size has also prompted researchers to study the computation of a good approximation by examining only a small subset of the input. This is a particularly popular research theme in the data streaming environment.

There has also been recent algorithmic progress on the shape matching problem under rigid and affine transformations. The problem calls for placing two shapes using the allowed transformations in order to maximize some similarity measure. Recently, fast approximation algorithms have been obtained for matching polygonal shapes under rigid motions, matching convex sets in arbitrary dimensions under scaling and translations, and finding the largest common point set under rigid motions in 2D. Good progress has also been obtained on computing the Frechet distance of two polygonal curves. Still, there is a lot to be done for the shape matching problem; for example, matching polyhedral shapes in 3D under rigid motions, finding the largest common point set under rigid motions in 3D, etc. There are also many GPS trajectory problems that are closely related to shape matching.

Finding a shortest path is a classical geometric optimization problem that finds applications in computer graphics, motion planning, geographical information systems, and seismic simulations. The traditional objective is to minimize the Euclidean path length, but there has also been research work on modeling the varying difficulties in traversing different regions. Recently, there has been progress in the weighted region model, in combining path length with height constraints, and in handling situations in which the speed of travel is direction-sensitive. Still, there are many open problems; for instance whether there is an FPTAS for the weighted region problem in 3D; or whether one can handle more general cost functions for navigating a terrain. Closely related is the evacuation problem, which finds applications in planning evacuation routes for crowds of people when an emergency arises. There has been progress on the problem when the underlying network is a path. More research is needed to handle more general graph topologies.

There are many other important geometric optimization problems, in addition to the ones mentioned above. To advance the state of the art, an extensive collaboration among researchers is highly desirable. This meeting serves to trigger such collaboration on advanced research in this area.

(Excerpt from the submitted proposal)

# Recent Advances in Randomized Numerical Linear Algebra

# 077



Shonan Village Center, July 25–28, 2016

Seminar No.080

## Organizers

Ravindran Kannan, Microsoft Research India, India  
Michael Mahoney, University of California, Berkeley, USA  
David Woodruff, IBM Research Almaden, USA

## Introduction and overview

Randomized Numerical Linear Algebra (RandNLA) is an interdisciplinary research area that exploits randomization as a computational resource to develop improved algorithms for large-scale linear algebra problems. From a foundational perspective, RandNLA has its roots in theoretical computer science (TCS), with deep connections to mathematics (convex analysis, probability theory, metric embedding theory) and applied mathematics (scientific computing, signal processing, numerical linear algebra). From an applied perspective, RandNLA is a vital new tool for machine learning, statistics, and data analysis. Well-engineered implementations have already outperformed highly-optimized software libraries for ubiquitous problems such as least squares (LS), with good scalability in parallel and distributed environments. RandNLA promises a sound algorithmic and statistical foundation for modern large-scale data analysis.

The great interdisciplinary strength of RandNLA is also one of the main challenges to its future progress. Researchers from diverse areas don't speak a common language, they don't publish in the same publication venues, they often can't evaluate the significance of important contributions of researchers from other areas. In this proposed workshop, we have assembled an interdisciplinary group of researchers who have made fundamental contributions to different aspects of RandNLA.

Our main goal is to advance mathematical aspects of RandNLA and to solve fundamental algorithmic and statistical challenges in making this theory useful in large-scale machine learning and data analysis applications. In particular, by bringing together these researchers in this workshop, we would like to develop a multi-faceted common theoretical foundation for state-of-the-art RandNLA methodologies and a detailed understanding of the complementary algorithmic and statistical issues that arise in their application to modern large-scale data analysis.

## Background

Matrices are ubiquitous in computer science, statistics, and applied mathematics. An  $m \times n$  matrix can encode information about  $m$  objects (each described by  $n$  features), or the behavior of a discretized differential operator on a finite element mesh; an  $n \times n$  positive-definite matrix can encode the correlations between all pairs of  $n$  objects, or the edge-connectivity between all pairs of nodes in a social network; and so on. Motivated largely by technological developments that generate extremely large scientific and Internet data sets, recent years have witnessed exciting developments in the theory and practice of matrix algorithms. Particularly remarkable is the use of randomization typically assumed to be a property of the input data due to, e.g., noise in the data generation mechanism as an algorithmic or computational resource for the development of improved algorithms for fundamental matrix problems such as matrix multiplication, LS approximation, low-rank matrix approximation, Laplacian-based solvers, etc.

Each of the three organizers has written a separate NOW Publishers monograph on different aspects of randomized numerical linear algebra, including: "Spectral Algorithms" by Kannan and Vempala in 2009, "Randomized Algorithms for Matrices and Data" by Mahoney in 2011, and "Sketching as a Tool for Numerical Linear Algebra" by Woodruff in 2014.

These monographs explain how sampling, sketching, and other randomized algorithmic techniques can be used for speeding up classical solutions to numerical linear algebra problems. Since it is an inherently interdisciplinary area, RandNLA can be approached from several different, yet complementary, perspectives, such as "Pure and applied mathematics", "Computer science", "Statistics and machine learning" and "RandNLA in the field".

(Excerpt from the submitted proposal)

# Dynamic Networks Visual Analytics: Approaches to Facilitate Visual Analysis of Complex and Dynamic Network Data

Shonan Village Center, August 1–4, 2016

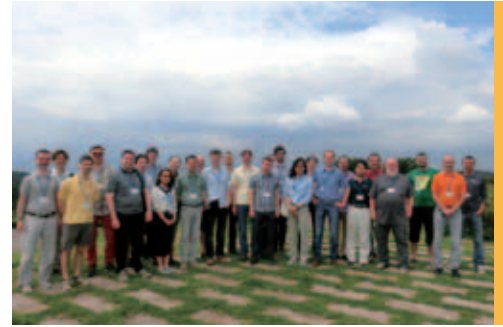
Seminar No.085

## Organizers

Takayuki Itoh, Ochanomizu University, Japan

Karsten Klein, University of Konstanz, Germany

Giuseppe Liotta, University of Perugia, Italy



## Introduction

There is an abundance of complex dynamic data across many domains. Often such data has a natural interpretation as a network, e.g. for applications in biology, the social sciences, telecommunication, finance, energy, and information security. Analysis of these dynamic networks is key to the understanding of complex processes both in fundamental research and for decision making. Use-cases include real-time monitoring of events and processes as well as a posteriori analysis of recorded data to gain insight. This insight then can be used to explain the principles of an underlying system, to detect important artifacts, e.g. cases of emergency, or to predict future changes. Efficient tools and methods for the visual analysis of network changes will facilitate the human understanding of dynamic networks, but the complexity and volume of the network dynamics poses a challenge to research. Today's telecommunication networks are huge, highly dynamic and heterogeneous, social media channels can have hundreds of thousands of messages per second, and microsecond trading at financial markets creates millions of data points per second. The development of sensor networks e.g. for electric grids and smart homes is advancing fast, and will lead to huge amounts of network data that need to be analyzed. Recently, approaches to combine multiple sources of structured and unstructured data have become more and more popular, e.g. to support decision making and fraud detection at the capital markets by combining financial data with sentiment analysis of social media data. The collection, processing, and visualization of complex network data for analysis purposes presents a huge challenges for all of those areas. While visualization of changes is an important task, interfaces that allow the user to intuitively interact with the visualization in an efficient manner are a further challenge. An important aspect is to tailor the representation towards the actual use-case. Dynamic network visual analytics poses a research challenge not only due to the dynamic nature of the data, but also its volume and complexity.

## Topics and Aim of the Seminar

We want to discuss dynamic network visual analytics starting with a few core real world examples, where practitioners present the setting and challenges and discuss requirements and constraints for potential solutions with visualization and analysis experts. Our aims for the seminar are as follows:

- Bring together researchers from several communities that can contribute to dynamic network visual analytics research.
- Identify core challenges based on both ongoing research and real-world examples.
- Examine where existing approaches can be extended and where completely new methods need to be developed to cope with the recent huge rise in complexity, velocity and volume of dynamic network data for network analysis and interactive visualization.
- Investigate the embedding of dynamic network data and visualizations into the interfaces and visualizations commonly used in practice today.
- Investigate how we can adapt existing visual analytics techniques for dynamic graphs to interact with and be executed on a cloud computing framework.
- Investigate the use of alternative representations than node-link visualizations for dynamic networks.
- Study efficient scalable approaches to detect and visualize repetitive behavior.
- To define and document research challenges for the future.

(Excerpt from the submitted proposal)

# Web-based Molecular Graphics

# 079



Shonan Village Center, September 5–8, 2016

Seminar No.086

## Organizers

Andrea Schafferhans, Technical University of Munich, Germany

Séan O'Donoghue, CSIRO, Garvan Institute, Australia

Haruki Nakamura, Osaka University, Japan

## Summary

Molecular graphics is a well-established discipline used in many scientific fields, including commercial research. A wealth of methods and tools have been developed, and are still actively developed—however until recently activity has focused on stand-alone software. This meeting will explore an emerging new frontier for molecular graphics, namely deployment in a server-client environment, which promises to make molecular structural information much more easily accessible to scientists. In addition, the recent hybrid approaches have revealed structures, functions and dynamics of very large molecules and cellular machineries, which are displayed as both atomic structures and molecular/cellular images, greatly extending the roles of molecular graphics. This meeting will bring together key players in this emerging field to explore the challenges and opportunities raised by web-based molecular graphics and new applications.

## Background

The amount of data of molecular structure is steadily growing in size as well as in complexity: more protein structures, larger protein structures (big complexes) and time-resolved data, such as molecular dynamics trajectories. In the near future, these data may increase in scale to include, for example, complex cellular machineries and whole living cells. A key challenge is to make all of these data accessible to scientists, in ways that enable interpretation and the generation of new research insights.

Fortunately, nowadays almost all scientists—even students—have easy access to computer graphics capabilities powerful enough to show a wide variety of molecular structures. With modern bandwidth speeds, it is even feasible to stream complex, interactive molecular graphics into simple devices such as an iPad or iPhone, via GPU-rendering on a remote server. Thus, it is now possible to depict and explore 3D models in regular browsers. This allows to bring 3D molecular visualization tools to non-expert end users without the need to install specialized software. However, in times of a steadily growing amount of data, improved visualization methods and tools are urgently needed, which can make the more complex data accessible and interpretable.

## Aims of the Seminar

The seminar aims to help give this emerging field direction and clarity, by bringing together global efforts in making molecular structure information available to scientists via the web.

We aim to discuss strategies for effectively managing very large molecular structures. In addition, many of the scientific problems being investigated cannot be solved with molecular graphics alone. Instead, data on molecular structures needs to be exchanged with other resources, for example resources that integrate data derived from structures with other information, and display the result using network graphs, or using other visualization paradigms. Therefore, an approach is required which is compatible to different modeling environments. JavaScript seems best suited to be used to interactively combine the website's content with the molecular 3D visualization via WebGL. The seminar will discuss pros and cons of this architecture and will define desirable interaction paradigms and communication methods.

## Audience

In addition to inviting key players in web-based molecular graphics, we will invite members from each of the laboratories behind the main traditional 'stand-alone' molecular graphics systems. We would also include prominent scientists that work on applying molecular graphics to the life sciences, biomedicine, chemistry, and material sciences. Furthermore, we will include members of the data visualization community—especially those interested in web-based graphics, as we expect they will inject 'out-of-the-box' thinking.

(Excerpt from the submitted proposal)

# Bidirectional Transformations

Shonan Village Center, September 26–29, 2016

Seminar No.091

## Organizers

Anthony Anjorin, University of Paderborn, Germany  
Zinovy Diskin, McMaster University and the University of Waterloo, Canada  
Meng Wang, University of Kent, United Kingdom  
Yingfei Xiong, Peking University, China



Bidirectional transformations (BX) is a common pattern of computing: transforming data from one format to another, and requiring a transformation in the opposite direction that is in some sense an inverse. The most well-known instance is the viewupdate problem from relational database design: a view represents a kind of virtual database table, computed on the fly from concrete source tables rather than being represented explicitly, and the problem comes when mapping an update of the view back to a corresponding update on the source tables. In a similar way, the problem is central to model transformations, playing a crucial role in software evolution: having transformed a high-level model into a lower-level implementation, one often needs to restore consistency between the two parts that may have evolved separately.

Giving this widespread applicability, research on BX naturally spans multiple disciplines: (1) Programming Languages (PL), (2) Graph Transformations (GT), (3) Software Engineering (SE), and (4) Databases (DB).

1. In PL research, the goal is to design languages that are suitable for programming BX (i.e., two opposing transformations are encoded in a single definition rather than with two separate definitions). Notable techniques includes lenses—collection of combinators, which can be read in two ways, and bidirectionalization—program transformation that constructs bidirectional programs from unidirectional ones.
2. In GT research, a rule-based approach is taken by specifying consistency as a direction-agnostic graph grammar, i.e., a set of high-level graph transformation rules that generate the language of consistent pairs of graphs. Incremental forward and backward synchroniser with desirable properties are then automatically derived by operationalising this grammar.
3. In SE research, the goal is to support different software engineering activities with BX. In software development, people usually create different kinds of artefacts, and it has been a long-standing problem to synchronise these artefacts and keep them consistent. By applying BX to these problems, SE researchers have contributed to different aspects of BX research, such as synchronisation of object-oriented models, delta-based BX, and exploring the design space of BX.
4. In DB research, BX concerns the classical view-update problem of relational databases, the dual problem of incremental view maintenance, and modern manifestations of synchronization problems like data exchange and provenance. XML transformations is another active research area of BX, effectively allowing queries to be made bidirectional.

The Shonan meeting proposed in 2016 will build on the momentum and results generated at previous meetings of similar nature (Dagstuhl 2011 and Ban 2013) to further develop cross-disciplinary research agenda and integration effort.

# Cognitive Development and Symbol Emergence in Humans and Robots

# 001



Shonan Village Center, October 3–7, 2016

Seminar No.092

## Organizers

Tadahiro Taniguchi, Ritsumeikan University, Japan

Emre Ugur, Bogazici University, Turkey

George Konidaris, Brown University, USA

Recent advances in modern robotics and machine learning, including Bayesian nonparametrics and deep learning, have brought the possibility of creating embodied computational intelligences that behave adaptively in real-world environments once more to the fore. Symbol emergence in robotics is an emerging research field that attempts to solve interdisciplinary problems related to the use of “symbols”, using a constructive approach.

Symbol systems have been an important and problematic topic in both artificial intelligence and cognitive science. Human beings make use of symbol systems to recognize various phenomena in the world and to communicate and collaborate with other entities, including robots and other humans. Newell proposed the physical symbol system hypothesis, and this notion had been fundamental to conventional artificial intelligence and cognitive science. The main misunderstanding in the physical symbol system hypothesis is that they assume that a symbol system exists without any real-world information. Owing to the missing link between the symbol system and the real-world, the physical symbol system is ungrounded, and therefore unable to function appropriately in complex environments. The “symbol grounding problem” and “intelligence without representation” arguments challenged the conventional naive understanding about and implementation of a symbol system. In cognitive science, Barsalou shed the light on a perceptual aspect of symbol system and proposed the perceptual symbol system.

The fundamental problem is that many researchers have been conflating the kinds of symbols that usually appear in programming languages, and human meaning-oriented symbols. Steels call them c-symbols and m-symbols, respectively, pointing out the ill-posed characteristics of the original symbol grounding problem. To understand the human meaning-oriented symbol system and develop an intelligent robot that can communicate and collaborate with humans, we have to model the dynamics of the symbol system on the basis of embodied cognition and semiotic communication among the agents in a bottom-up and data-driven manner.

In the developmental and cognitive sciences, there are still many unsolved problems in human infant language acquisition and concept formation processes. An integrative model that can explain the process of language acquisition and the use of the obtained language simultaneously is still missing. To understand the process of language acquisition, we require a description of the dynamic process of language-related capabilities, e.g., learning vocabulary, constructing mutual belief, forming multimodal concepts, and learning syntax. A computational model based on machine learning provides a promising model for understanding cognitive phenomena related to language acquisition. Similarly, an embodied computational model for language evolution is important in evolutionary linguistics.

This Shonan meeting focuses on the constructive approach towards embodied language acquisition, symbol emergence and cognitive development in autonomous systems, including human and robots. To foster this, we aim for interdisciplinary discussions with a wide viewpoint from various research fields including not only related computer science and robotics but also cognitive science, developmental psychology, linguistics, and other fields in the humanities. The scope of this meeting will include language acquisition, developmental robotics, human-robot interaction, machine learning, cognitive science, artificial intelligence, neural networks, multimodal sensory experience (auditory, speech, gestures), language evolution, and semiotics.

(Excerpt from the submitted proposal)

# 002 Perception in Augmented Reality

Shonan Village Center, November 14–18, 2016

Seminar No.088

## Organizers

Christian Sandor, Nara Institute of Science and Technology (NAIST), Japan  
Dieter Schmalstieg, The Institute of Computer Graphics and Vision,  
Graz University of Technology, Austria  
J. Edward Swan II, Mississippi State University, USA



Over the years, research on head-worn Augmented Reality (AR) has been complemented by work on new platforms such as handheld AR and projector-camera systems. With the rapid advent of applications on cell phones, AR has become almost mainstream. However, researchers and practitioners are still attempting to solve many fundamental problems in the design of effective AR. Although many researchers are tackling registration problems caused by tracking limitations, perceptually correct augmentation remains a crucial challenge.

Some of the barriers to perceptually correct augmentation can be traced to issues with depth and illumination that are often interconnected, or by issues related to the appearance of an environment. These problems may cause scene and depth distortions, and visibility issues, which can lead to poor task performance. Most of the issues are caused by limited understanding or by inadequate methods for displaying information.

In the mid-90s, Drascic and Milgram attempted to identify and classify these perceptual issues. Focusing on stereoscopic head-worn displays (HWDs), they provided useful insights into some of the perceptual issues in AR. Since then, considerable research has provided new insights into perceptual factors; e.g. Kruijff et al. (2010). Even though HWDs are still the predominant platform for perceptual experiments, the emphasis on a broader range of AR platforms has changed the problem space, resulting in the need to address new issues.

We believe that an overarching approach is needed to address these issues by bringing together both researchers who focus on technology issues, and researchers who focus on related psychology and cognitive science areas. We strongly feel that bringing both sides into an open meeting holds substantial promise for developing a research agenda that can address the serious challenges in AR, and therefore accelerate the promise of AR technology.

In 1950, Alan Turing introduced the Turing Test, an essential concept in the philosophy of Artificial Intelligence (AI). He proposed an “imitation game” to test the sophistication of AI software. At its core is a precise test protocol, where human participants are to determine whether a conversation partner is an actual human, or is instead an AI simulation. Similar tests have been suggested for fields including Computer Graphics and Visual Computing. We strongly believe that an AR Turing Test must be added to the AR research agenda. However, it is not straightforward to define such a test. Nevertheless, it is crucial, in order to have a measurable goal for the attempts of others and ourselves to erase the boundary between real and virtual, which requires perceptually correct augmentations, that we need to be able to test for. We think that this proposed seminar can be a crucial stepping stone for defining an AR Turing Test.

## Topics

- Information Presentation
- Input
- Output
- User Experience Design

(Excerpt from the submitted proposal)



# Implicit and Explicit Semantics Integration in Proof Based Developments of Discrete Systems

088



Shonan Village Center, November 22–25, 2016

Seminar No.090

## Organizers

Yamine Ait-Ameur, IRIT Enseeiht, France  
Shin Nakajima, National Institute of Informatics, Japan  
Dominique Mery, Loria, Université de Lorraine, France

The idea of the meeting arose from two observations, concerned with the nature of proof-based systems and the way in which they are used by formal methods applied to the development of software.

Firstly, most proof based systems are based on the definition and use of theories (logic, algebraic, types, etc.) in order to support the expression of proofs in formal developments that can become highly complex. In general, these proof systems are based on theories composed of axioms and an inference system. The formal developments conducted in such systems contribute to the definition of new theories or to the enrichment (extension) of existing ones through new deduced theorems using automatic and/or semi-automatic proof procedures.

Another observation is related to the way in which a hardware and/or software system is developed using formal methods and how the proofs associated with these developments are conducted. Such systems are usually designed by decomposing into/composing subsystems in a top/down, bottom/up and mixed process. The refinement/abstraction and instantiation development operations are applied during these processes.

The objective of the meeting is to discuss on mechanisms for reducing model heterogeneity induced by the absence of explicit semantics expression in the formal techniques used to specify these models. More precisely, the meeting targets to highlight the advances in the state in handling both implicit and explicit semantics in formal system developments. Discussions, presentations and more generally, contribution shall address one or more topics described below:

- Show that when making explicit the domain knowledge in formal models, several relevant hidden (because they are not explicitly modeled in classical formal modelling languages) properties can be handled.
- How knowledge models, like ontologies, can be handled in formal system developments?
- What are the candidate formal modelling languages and techniques to model such domain knowledge? What are the reasoning capabilities entailed by these modelling languages?
- Define relevant case studies that illustrate the need to make explicit the domain properties.
- Define composition mechanisms to handle domain knowledge in formal modelling techniques.

Beyond the technical results targeted by the proposed Shonan meeting, social, economic and resilience impacts are expected. These impacts are built on the foundations of heterogeneity reduction and formal model alignment. Finally, we mention that once this meeting is completed, it is planned to publish a book gathering the main contributions and results.

(Excerpt from the submitted proposal)

# 084 Algorithmics for Beyond Planar Graphs

Shonan Village Center, November 28–December 1, 2016

Seminar No.089

## Organizers

Seok-Hoe Hong, The University of Sydney, Australia

Takeshi Tokuyama, Tohoku University, Japan



## Summary

Many existing graph algorithms have a strong assumption that the input graph is planar, however, most graphs such as social networks or biological networks in real-world applications are non-planar. The main goal of the workshop is to promote Graph Algorithm research in Asia-Pacific region, and form a research community to collaboratively solve complex problems arising in a variety of application domains such as social networks or systems biology. In particular, special emphasis on the “Beyond Planar Graphs” will be addressed. This workshop aims to identify research opportunities on Beyond Planar Graphs, focusing on the Asia-Pacific context. More specifically, we will develop innovative algorithms to handle sparse non-planar graphs with specific application of large and complex network visualization:

- 1)  $k$ -planar graphs: a graph is  $k$ -planar if it has a drawing in which no edge crosses more than  $k$  edges.
- 2)  $k$ -skew graphs: a graph is  $k$ -skew if it has a drawing in which deletion of  $k$  edges makes the resulting graph planar.
- 3)  $k$ -quasi-planar graphs: a graph is  $k$ -quasi-planar if it has a drawing in which no  $k$  edges mutually cross each other.

## Aims

This workshop aims to bring world-renowned researchers on Graph Algorithm, Graph Drawing, Computational Geometry, Graph Theory, and Combinatorial Optimization, and collaboratively develop innovative theory and algorithms for sparse non-planar topological graphs with specific applications of large and complex network visualization. More specifically, we have the following aims:

- 1) Structural properties: We aim to characterize classes of sparse non-planar topological graphs. We aim to find important structural properties of such graphs.
- 2) Testing algorithm: We want to determine the complexity of the problem of testing whether a given graph satisfies such topological constraints is NP-hard. If not, we will design efficient (polynomial time) algorithms for testing such properties.
- 3) Drawing algorithm: We aim to design polynomial time algorithms to construct a straightline drawing of an embedding that satisfies topological constraints.

## Objectives

- 1) We will form a broader research community with cross-disciplinary collaboration, between Computer Science as well as Mathematics.
- 2) We will foster greater exchange between Graph Drawing community, Computational Geometry community and Graph Theory community, and to draw more researchers in the Asia-Pacific region to enter this rapidly growing area of research.

## Significance and Innovation

Existing planar graph drawing algorithms have made little impact on visualization of real-world complex networks. This workshop will be the first general and extensive investigation of the algorithmic questions arising from sparse non-planar topological graphs for real-world network visualization. We believe that this workshop will set a new research agenda in Graph Drawing and Geometric Graph Theory.

## Expected Outcomes

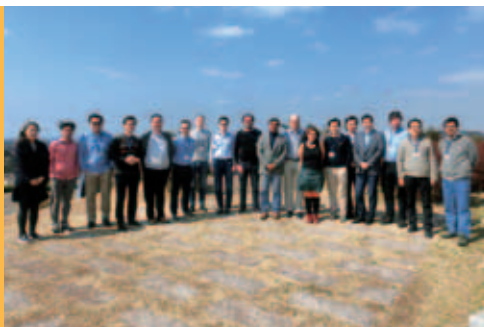
Innovative techniques and solutions for Beyond Planar Graphs, which will be used by domain experts and end users in various application.

- 1) Joint publications at the top conferences and journals in Algorithms and Theory, jointly authored by researchers in Graph Algorithm, Graph Drawing, Computational Geometry, Combinatorial Optimization, and Graph Theory.
- 2) Joint funding applications for long-term research collaboration for continuation of the research collaboration beyond 2016.

(Excerpt from the submitted proposal)

# Microfluidic Biochips: Bridging Biochemistry with Computer Science and Engineering

# 085



Shonan Village Center, February 27–March 2, 2017

Seminar No.081

## Organizers

Shigeru Yamashita, Ritsumeikan University, Japan  
Tsung-Yi Ho, National Tsing Hua University, Taiwan  
Robert Wille, Johannes Kepler University Linz, Austria  
Krishnendu Chakrabarty, Duke University, USA

Advances in microfluidic technologies have led to the emergence of biochip devices for automating laboratory procedures in biochemistry and molecular biology. These devices enable the precise control of nanoliter-scale biochemical samples and reagents. Therefore, Integrated Circuit (IC) technology can be used to transport a “chemical payload” in the form of micro-or nano-fluidic carriers such as droplets or as bulk flow in microchannels. As a result, non-traditional biomedical applications and markets are opening up for ICs and systems. This represents a More than Moore-approach.

However, continued growth (and larger revenues resulting from technology adoption by pharmaceutical and healthcare companies) depends on advances in chip integration and design-automation tools. Thus, there is a need to deliver the same level of Computer-Aided Design (CAD) support to the biochip designer that the semiconductor industry now takes for granted. Also, the design of efficient CAD algorithms for implementing biochemistry protocols to ensure that biochips are as versatile as the macro-labs that they are intended to replace. This is therefore an opportune time for the software and semiconductor industry as well as circuit/system designers to make an impact in this emerging field.

Recent years have therefore seen growing interest in design methods and design-automation tools for the digital microfluidic platform with special issues of IEEE Transactions on CAD and IEEE Design & Test of Computers, special sessions at DAC, ISPD, ASPDAC, and ICCAD, as well as workshops/tutorials at ISCAS, ICCAD, SOCC, and DATE. A number of CAD research groups worldwide (US, Germany, Taiwan, Denmark, China, Japan, India, etc.) have initiated research projects on CAD for microfluidic biochips.

The goal of the meeting is to bring together experts in order to present and to develop new ideas and concepts for the design automation algorithms and tools for microfluidic biochips. This meeting also provides a unique opportunity to bring chip designers, bioengineers, biochemists, and theoretical computer scientists together for comprehensive discussions on the research tasks as well as commercial prospects in this domain. Areas ranging from architecture, synthesis, optimization, verification, testing, and beyond should be covered.

As possible results, we expect to see a better understanding of the respective areas, new impulses for further research directions, and ideas for areas that will heavily influence research in the domain of design automation on microfluidic biochips within the next years. The meeting will facilitate greater interdisciplinary interactions between researchers in chip designers, bioengineers, biochemists, and theoretical computer scientists.

In order to ensure high-quality presentations and lively discussions, we carefully selected experts for invitation to the meeting. All of them have established for themselves a stellar reputation in the respective domains. While researchers working on design automation and optimization of microfluidic biochips build the majority of the invitees, also some experts from surrounding research areas should be invited. For example, researchers and industry practitioner’s people working on emerging architectures and applications of microfluidic biochips shall provide the needed insight for the discussions about the practical problem formulation for commercialized products. Computer scientists with a focus on computer-aided design shall enrich the discussions about the top-down design methodology and optimization of large-scale components like mixers and routing channels. Therewith, the unique concept of Shonan meetings is applied in order to bring researchers from different domains together so that also the interdisciplinary topics can be discussed and progress in these areas is made.

(Excerpt from the submitted proposal)

# Mining Software Repositories: Accomplishments, Challenges and Future Trends (NII Shonan School)

Shonan Village Center, March 6–10, 2017

Seminar No.095

## Organizers

Emad Shihab, Concordia University, Canada

Akinori Ihara, Nara Institute of Science and Technology (NAIST), Japan



## Invited Lecturers

Ahmed E. Hassan, Queen's University

Daniel German, University of Victoria

Alberto Bacchelli, Delft University of Technology

Shane McIntosh, McGill University

The field of Mining Software Repositories (MSR) has been steadily growing over the past decade. Since its first workshop in 2004, the MSR community has steadily grown to be one of (if not) the biggest co-located events with the International Conference on Software Engineering, the flagship conference in Software Engineering.

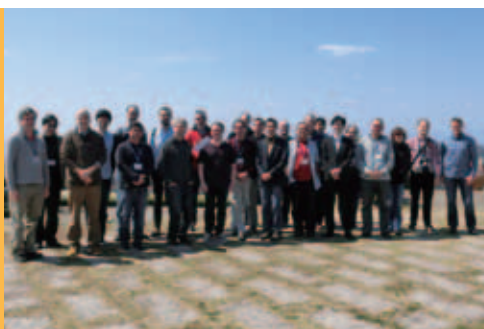
The field of informatics concerns the processing of information and engineering of information systems. The main goal of the MSR community is to leverage development data, often stored in software repositories, hence it can be seen as a sub category of the informatics field, which specifically concerns the processing of software development data. Examples of these repositories are: source control repositories, which store source code changes, defect tracking repositories, which store software defect reports and communication repositories, which store developer communications such as emails. These repositories contain a wealth of information that is available for most software projects. The MSR community has proposed techniques to effectively mine repository data, leverage such data to improve requirement, quality and traceability of software project and empirically study the impact of several development phenomena.

However, now MSR is at a critical point where many accomplishments have been made, but many challenges remain due to the changing landscape of software engineering. For example, the availability of data was one of the biggest challenges for MSR in the past, whereas now the availability of too much data is causing challenges. Recent technology advancements in complementary areas such as machine learning and artificial intelligence has enabled MSR researchers to develop more accurate techniques, however usability remains as an open challenge. The widespread use of mobile devices has led to more mobile-related software, often known as mobile apps, is also a new trend that the MSR community has recently started to target.

Given the recent accomplishments, challenges and trends in MSR, we plan to organize a NSS to teach and train future generations about the accomplishments and future challenges of MSR. The NSS will serve as a forum where students can learn from leaders in the MSR field and discuss potential solutions for such challenges. The NSS will be different than the first MSR workshop and the current MSR conferences, which focus mainly on encouraging researchers to showcase their techniques to process repository data and to present their findings from mining such data. In short, the MSR workshop and conference are meant for established researchers, whereas the NSS is meant to train future generations of the MSR field.

We plan the lectures at the NSS to have both, hands-on and lecture sessions. The lectures will be given by a mixed set of well-established and emerging leaders in the MSR field. In addition, we plan to have a day where students and young researchers will present and get feedback on their latest work. To maximize interaction and provide useful feedback for the students and young researchers, the poster session will be a draw your research on a poster session, where presenters will be given a poster paper and a felt pen and asked to draw out their research.

# 007 Computational Metabolomics



Shonan Village Center, March 20–23, 2017

Seminar No.093

## Organizers

Sebastian Böcker, Friedrich Schiller University Jena, Germany

Masanori Arita, National Institute of Genetics, Japan

Steffen Neumann, Leibniz Institute of Plant Biochemistry, IPB Halle, Germany

Metabolomics has been referred to as the apogee of the omics-sciences, as it is closest to the biological phenotype. Metabolites are not only responsible for tasks such as growth, development, and reproduction, but also directly relevant to structure, signaling, and chemical interactions with other organisms. Metabolomics also plays an essential role in the investigation of novel drug leads or profiling metabolites of pharmaceutical compounds to detect and understand side effects. With advances in instrumentation, metabolomics is currently at the edge of becoming a “big data” science.

Mass spectrometry is the predominant analytical technique for detecting and identifying metabolites and other small molecules in high-throughput experiments. Huge technological advances in mass spectrometers and experimental workflows during the last decades enable novel investigations of biological systems on the metabolite level. But these advances also resulted in a tremendous increase of both amount and complexity of the experimental data, and the data processing and metabolite identification form the largest bottlenecks in high-throughput analysis. Unlike proteomics, where close cooperation between experimental and computational scientists have been established over the last decade, such cooperation is still in its infancy for metabolomics.

The key goal of this seminar is to foster the exchange of ideas between the experimental (analytical chemistry and biology) and computational (computer science and bioinformatics) communities. State-of-the-art methods from computer science, statistics, analytical and biological experiments will be presented, along with problems arising from these techniques. Brainstorming sessions and breakout groups will discuss individual topics in greater detail, to initiate new collaborations between participants who have not yet worked together. This exchange of expertise is needed to form a scientific community to advance computational metabolomics.

A selection of topics to initiate discussions at the seminar include:

- Searching in molecular structure databases: How can the promising approaches MetFrag, MAGMa, FingerID, LipidBlast, CFMID and others be improved?
- Identification statistics: What statistics can be incorporated to improve identification quality of metabolites, such as False Discovery Rates?
- Experimental frontiers: Incorporation of experimental strategies such as data-independent acquisition (DIA), ultrahigh resolution, imaging mass spectrometry, 2-dimensional chromatography etc. in metabolomics.
- Labeling: Development of novel computational methods for analyzing the data from labeling experiments to gain learn about metabolic transformations.
- Quantification and biomarker discovery: Many computational challenges remain to be discussed in these fields.
- Incorporating experimental knowledge into computational methods: How can experimentalists add their knowledge into automated procedures?
- Screening methods and metabolite prediction: Can we improve the methods to help “look” for metabolites rather than performing non-target identification?
- Data exchange and public reference data: How can metabolomics researchers be encouraged to provide additional training data that covers a sufficient breadth of the expected molecular space?
- Publication standards for computational methods: Can current standards be improved, consolidated and harmonized, to ensure consistent presentation of methods and publicly available reference data?

(Excerpt from the submitted proposal)

# Semantics of Effects, Resources, and Applications (NII Shonan School)



Shonan Village Center, May 15–19, 2017

Seminar No.103

## Organizers

Marco Gaboardi, University at Buffalo, SUNY, USA

Shinya Katsumata, Kyoto University, Japan



Programs are dynamic entities that interact with their environment in a complex way. Understanding and controlling this behavior is one of the central themes in programming language research. Several approaches have been proposed with the aim of representing the behavior of a program as some abstract mathematical structure that is manageable to program analysis. Formal semantics has provided the needed tools to make this abstraction effective and useful. The advantage of this approach is that program can then be statically analyzed in order to verify that their behavior meet given specifications.

One such approach that has been particularly successful is the one based on the use of types for representing specifications and of type-checking and type-inference as program analyses guaranteeing that the programs respect these specifications. The success of this approach is also due to the strong connections between functional programming languages, logic, semantics and mathematics, represented by Curry-Howard isomorphism.

The technological advancement in type systems has brought in recent years new tools that are useful to reason in an abstract way about non-functional aspects of program behavior such as *side effects and resource consumption*.

Side effects, such as I/O, store updates, using randomness, and performing nondeterministic choice, are fundamental components of modern programming languages. The side effects caused by programs can be understood and controlled by using type systems that contain additional information useful to describe them. Type systems of this kind are often referred as effect systems. Effect systems are commonly used as framework to statically estimate program's side effects and to enforce particular management policies for them. An interesting aspect of this approach is that effect systems can be directly relates to a rigorous representation of various side effects as monads, a concept from category theory.

Resource consumption is another fundamental aspect of program's behavior. Several techniques have been proposed to describe resources in programs semantics. A technique that have found numerous applications in resource analysis is Linear Logic. Its key idea is to distinguish reusable (non-linear) data from single-use (linear) data using a command, another concept from category theory. This approach can be generalized to explicitly track different kind of data usage such as execution time, memory space, information flows, etc. Crucially, the approach proposed by linear logic can be easily integrated in type systems, providing new techniques for the analysis of resource usage of programs. Moreover, this approach has generated several important results in research areas where the notion of resource is a key component, such as implicit computational complexity, and sensitivity analysis in the context of differential privacy.

The semantics understanding of effects and resources has also permitted the integration of these refined aspects of the programs behavior in the formal reasoning about practical applications. As an example, notions of resources and effect are now widely used when reasoning about the correctness of compilation and implementation of programs. Similarly, semantics models for resources and effects are used to reason about randomized algorithms.

The goal of the Shonan School is to share the cutting-edge techniques and various research problems in the semantic study of effects and resources in programming languages. We expect students to engage with other students and with the lecturers and organizers on different research topics.

(Excerpt from the submitted proposal)

# Algorithms and Optimization under Uncertainty



Shonan Village Center, May 22–25, 2017

Seminar No.094

## Organizers

Niv Buchbinder, Tel-Aviv University, Israel  
Nikhil Devanur, Microsoft Research, USA  
Debmalya Panigrahi, Duke University, USA

### Description of the meeting

In the classical offline computational model, an algorithm operates on a specified set of input data to produce a desired output. While this model has propelled much of computer science, modern applications typically do not afford the luxury of complete knowledge and certainty of the input data. Thus, the typical real world algorithm today has to make optimization decisions with incomplete information.

In this workshop, we plan to bring together world-class researchers from multiple domains of algorithm design with the goal of discussing and furthering research in this area of algorithms under uncertainty. The main focus will be on optimization problems that are central in algorithmic research, and our objective will be to promote activity in theoretical research on these problems in limited information settings.

### Online algorithms and Competitive analysis

The field of online computation in which the input is presented piece-by-piece to the algorithm played an important role in the design of data structures, scheduling, memory management, and other fundamental areas in computer science. Recently, there has been a renewed interest in this area due to new applications that emerged in online systems, and especially Internet advertising.

In addition to these myriad applications, recent research interest in the online algorithms community has also been piqued by the development of fundamentally new ideas and techniques—in particular, the import of the linear programming toolkit applicable to a broad class of online problems.

### Beyond worst-case analysis and connections to machine learning

There is a growing belief among researchers that we also need to explore possibilities beyond worst case (competitive) analysis. For example, in online advertising, while future demands are not exactly known in advance, reasonably sound predictions based on historically observed data is commonly used in practice. This has led to a growing body of work in these domains that focuses on average case analysis, encompassing various stochastic models including i.i.d. models with (known/unknown) input distributions and random permutation models. Improvements in algorithmic performance have been noted in these average-case settings over the corresponding worst-case models, and have been successfully employed in real world online systems.

Recent research has suggested that such gains can be extended to other domains as well, most notably in the broad class of optimization problems that can be encoded as packing/covering linear programs. From a technical perspective, this research is bringing together two powerful toolkits in theoretical computer science—online learning and combinatorial optimization. At the heart of online learning is the bandits framework capturing the “explore-exploit tradeoff” that characterizes optimization with incomplete information. While optimization under uncertainty and online learning have similar goals at a high level, there are crucial differences in the techniques commonly used, and successful transition of technical tools between the two areas is a recent and nascent phenomenon.

In this workshop, we plan to bring together experts in these two areas for the exchange of ideas and problems. For example, we want to encourage the following question: can no-regret bounds that are ubiquitous in online learning also be obtained in classical online problems? We believe that this is a promising direction for overcoming the limitations of worst-case competitive analysis, where one of the key definitional obstacles has been in its performance metric that compares against an offline optimum. We believe that ideas from queuing theory and other related domains could also be useful in this regard, and we plan to bring together researchers from these areas to enable cross-fertilization of ideas.

(Excerpt from the submitted proposal)

# Language Integrated Queries: Towards Standard Logics for Big Data Analytics

Shonan Village Center, May 29–June 1, 2017

Seminar No.098

## Organizers

Laurent Daynes, Oracle Labs, France

George Fletcher, Eindhoven University of Technology, The Netherlands

Wook Shin Han, Pohang University of Science and Technology, South Korea



## Context

Database management systems (DBMSs) are typically optimized for a particular data model (e.g., relational, semi-structured, graph-based) and interfaced with a unique query language (e.g., SQL, HiveQL, XQuery, JSONiq, SPARQL). In contrast, database applications are written in general-purpose programming languages that offer developers a large choice of libraries (e.g., to simplify presentation to the end-users, the writing of business logic, etc.).

For various architectural reasons, database applications execute in an environment distinct from that of the DBMS, i.e., on a client machine (e.g., a connected mobile device) or on a middle-tier, or even within the database itself. This situation causes two main problems that have been the focus of research for several decades: (1) how to better integrate database querying with application programming languages to eliminate impedance mismatch while retaining the full power of the database querying capabilities, and (2) how to minimize the traffic, both in terms of number of interactions and volume of data exchanged, between the database and its applications.

## Towards an Algebra for data analytics

In the last decade, we have witnessed a continued explosion of research and development of data intensive systems and languages for big data analytics. These range, for example, from distributed computing frameworks such as Apache Spark and Apache Flink to document-centric data stores such as MongoDB or Microsoft Azure DocumentDB. To bridge the gap between the specification of analytic tasks by clients of these systems, on one hand, and compilation of optimized execution plans, on the other, an analogue of the relational algebra for big data analytics processing is called for. Although each of the systems in the contemporary data engineering landscape to some degree realizes its own flavor of a query algebra, there is currently no recognized logical language which serves this role. Recent efforts such as Apache Calcite are a step in the right direction, but are still focused on the relational paradigm.

As discussed above, on the other end of the spectrum there have been efforts on integrating native data querying capabilities into languages (aka language integrated querying) such as .NET, Java, PHP, and JavaScript. Such efforts extend the various languages by the addition of query operators and expressions which often go beyond the expressiveness of the relational algebra.

A broad community discussion of the features and design of extended algebras for big data analytics, as integrated in general-purpose programming languages, is crucial to bring big data analytics solutions to the next level of maturity.

As data processing platforms equip themselves with support for new programming languages (e.g., JavaScript support in Postgres and Microsoft Azure DocumentDB; Python support in Amazon Redshift; R in Oracle, Microsoft SQL server, IBM dashDB, SPARK, etc.), the need for a data provider interface capable of accepting multi-lingual expressions in queries regardless of particular query syntax will only continue to grow in importance.

## Goals and outcomes of the meeting

The goal of this meeting is to take the first steps towards elaborating solutions for (1) a standard language-, data-model-, and platform-independent declarative interface to data providers which is able to leverage available multi-lingual capabilities of data providers; and, (2) corresponding compilation and execution strategies. For this broad discussion, we aim to bring together relevant leading researchers from both academia and industry, across the domains of programming languages, data management systems, and distributed and parallel systems. In addition to an in-depth report on the discussions and results of the seminar, other possible outcomes include a community white paper and concrete action plans for collaborations in research and longer-term international projects of broad ambition.

(Excerpt from the submitted proposal)



# Processing Big Data Streams

# 091



Shonan Village Center, June 5–8, 2017

Seminar No.087

## Organizers

Vladimir Braverman, John Hopkins University, USA

David Woodruff, IBM Almaden, USA

Ke Yi, The Hong Kong University of Science and Technology, Hong Kong

The more data we have, the more data we need to process. Whether it's Internet traffic or biological data, the hardware is never fast enough. The aim of this workshop is to focus on analyzing data under new models: when the data cannot be stored (e.g. identifying viruses in the Internet traffic), when we use multiple cores to analyze the data, and when we are generating short sketches of the data to be sent and analyzed by someone else. We believe a new set of algorithmic techniques (which rely mostly on statistics and on data structures) can be used in these models, and wish to find such techniques and employ them. An important focus of this workshop is to find algorithms which are elegant, and thus can also be used in practice. To develop these algorithms, it is critical to connect the key subareas of the algorithmic research on Big Data. These key subareas include streaming, sketching, and sampling. The goal of the workshop is to bring researchers together from these different sub-areas and to establish strong collaborations among the attendees.

### Streaming model

The main motivation for this model is cases where there is not enough memory (or storage space) to hold all the input. Thus, the input is examined sequentially, character by character, and each character is being processed on arrival. Algorithms in this model have sub-linear storage, and typically use only poly-logarithmic space and time (per character). Although the streaming model is quite restrictive, major breakthroughs were achieved for several sketching and statistical problems. Some of the methods rely on metric embeddings, pseudo-random computations, sparse approximation theory, and communication complexity. Finding other problems that can be solved in this model, and new techniques for solving them, are fundamental open questions. An important and natural relaxation of the streaming model is the semi-streaming model. In this model, the algorithm is now allowed near-linear space, and sometimes we also allow it multiple passes over the input (usually a constant number of passes). This model has been mainly used for solving graph problems, in which it is presented as a stream of edges. Due to the space limitations only the vertices and a portion of the edges  $l$  can be stored. Reducing the space from quadratic to linear is a very important practical achievement of this model. We also consider a stream of edit operations—in this model we maintain a list of counters, and the meaning of the element  $(i,v)$  is to add  $v$  to the  $i$ 'th counter, e.g., a stream of IP packets, which can be viewed as  $(\text{src ip addr.}, \text{data length})$ .

### Sampling

The concept of a representative random sample is a central tool in data gathering and analysis. It has been a key concept in statistics and probability theory. More recently, with the progress made in randomized algorithms, it has proved to be a very useful and versatile tool in algorithm design (e.g., in computational geometry, graph optimization and data structures). The usefulness of random sampling in algorithms design comes from the fact that a sample often inherits important properties of the algorithm's input, while at the same time it is very small. Thus one can discover these properties quickly by analyzing the sample, and then use this knowledge to improve the running time of the algorithm operating on the whole input.

### Sketching

Sketching is a method of saving data using an extremely concise representation, which allows us to retrieve the information required for the processing application without going over the data again. These techniques are not standard compression techniques since they are often linear so that they can be easily updated in a streaming context. They create much shorter descriptions than compression methods achieve. Rather, sketches do not allow retrieving the original data (or even a close approximation of it), but only allow a very specific processing to be done on the data. There are two possible ways to reduce sketch size. One is by using randomized methods that allow answering the query correctly with high probability, and the second is to allow an approximate rather than exact result.

(Excerpt from the submitted proposal)

# Towards Engineering Free/Libre Open Source Software (FLOSS) Ecosystems for Impact and Sustainability

Shonan Village Center, June 26–29, 2017

Seminar No.099

## Organizers

Brian Fitzgerald, University of Limerick, Ireland

Audris Mockus, University of Tennessee, USA

Minghui Zhou, Peking University, China



Free/libre open source ecosystems such as the Linux kernel, have had a tremendous impact on computing and society and have captured the attention of businesses, researchers, and policy makers. Millions of participants, from independent volunteers to those representing companies or government organizations, have created and maintain massive numbers of software projects, ranging from individual scratch space or classroom assignments, to critical infrastructure projects such as the Linux Kernel, OpenStack, Docker or Android.

The spectrum and scale of FLOSS has substantially expanded in recent years, as has its popularity. The combination of distributed version control and social media features have created “transparent” environments that facilitate the scaling up of the ecosystems to millions of repositories and developers. Despite the substantial amount of research on FLOSS in disciplines such as software development, organizational science, management, and social sciences, it remains unclear how and why FLOSS ecosystems form, how they achieve their impact, or how they sustain themselves. The open nature of these communities and the associated vast collections of operational data represent a tantalizing possibility to discover the mechanisms by which such ecosystems form and operate. Achieving such understanding would inform approaches to structuring future open source communities, and could reveal ways to nudge the behavior of individuals and groups involved towards greater sustainability of FLOSS ecosystems.

This proposed Shonan meeting will bring together a blend of established and young researchers involved in studying the FLOSS phenomenon from software engineering, human computer interaction, computer-supported cooperative work, data mining, cognitive science, psychology, operations research, organization management, and complex systems domains. Industry practitioners with experience in various FLOSS aspects will also be included. The participants will discuss fundamental questions that are related to the impact and sustainability of FLOSS ecosystems. More specifically, the meeting will have the following sessions:

1. How does an ecosystem form? How do different stakeholders work together to form a community that develop and maintain valuable and freely available software, and how does an ecosystem with millions of repositories and developers operate given the lack of centralized planning?
2. How is the ecosystem organized? How do the teams cooperate to resolve the issues (workflow), and what are the typical relationships between the code and the team?
3. How does the ecosystem evolve in response to the environment as technology and needs evolve over time?
4. What distinguishes ecosystems that sustain themselves from ecosystems that disappear? How can an ecosystem be sustained? Under what circumstances should it be sustained?
5. How do the newcomers learn the protocols and practices of an ecosystem? How would they sustain the ecosystem? What is the relationship between people sustainability and ecosystem sustainability?
6. What kinds of research methods might be utilized (e.g., what qualitative and quantitative methods) to achieve research goals?

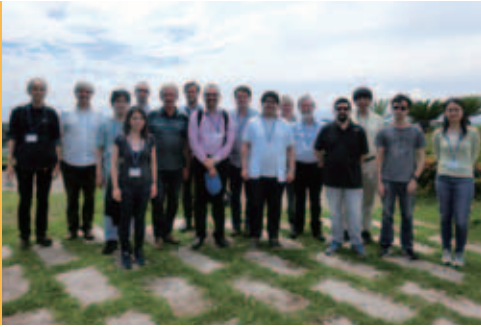
The expected outcome of the meeting is to achieve the following:

- a. Have provided an opportunity for the established and upcoming researchers and leading practitioners to exchange the ideas in the process of defining the agenda for FLOSS sustainability.
- b. Have framed the most critical research questions related to FLOSS ecosystem sustainability and impact.
- c. Determined the most relevant theoretical frameworks and methodological approaches to achieve research goals.
- d. Have made substantial progress towards identifying actions that would help sustain FLOSS ecosystems and reduce risks to the critical FLOSS infrastructure.

(Excerpt from the submitted proposal)

# Reverse Execution in Testing—Improving Security and Reliability

# 093



Shonan Village Center, July 3–6, 2017

Seminar No.102

## Organizers

Michael Kirkedal Thomsen, DIKU, University of Copenhagen, Denmark

Kazutaka Matsuda, Tohoku University, Japan

Mohammad Reza Mousavi, Halmstad University, Sweden

## Introduction

Implementing programs without errors is important for reliability and security of software. But though it is important it is also a hard task. Many approaches to this have over the years been developed, exemplified by numerous testing techniques, formal verification, and static program analysis. These techniques can be used to identify problematic parts of programs or to some degree verify the correctness of them. However, achieving any guarantees is often impossible or (at the least) very cumbersome, which to some extent is due to the complexity of the conventional deterministic computation model. So using a more restricted computation model has the potential to improve this.

A way to achieve this is to use a computation model that has a notion of reverse execution. Here, the ability to compute from any reached state back to any previous state is the restriction over a deterministic model; but at the same time the restriction gives the possibility for better reasoning within the model. There exist several different notions of which one can achieve reverse execution. Two of the most widely researched areas are Reversible Computations and Bidirectional Transformations.

Reversible Computations are often designed to be completely information lossless and information generating. The field is often motivated by Landauer's work that based on a gedankenexperiment found that, theoretically, the energy consumption of computations is not correlated with how much information is processed, but only with the amount of information that is lost during the computations. A result that has since been experimentally verified. However, modern research in the area has a larger focus on applications where the information preservation has an advantage.

Bidirectional Transformations are transformations, or more generally mechanisms, that keep the consistency of two or more data. Originally, bidirectional transformation was studied mainly around 80s in the database community as the view-update problem, in which one transformation is a usual query and the other transformation is translation of updates on a view to updates on the original table. Recently, the problem draws our attention again, influenced by the rise of programming languages and techniques for bidirectional transformations (such as lenses), and the needs in software development processes being complex, especially in model-driven software development.

## Testing meets Reversible Computations and Bidirectional Transformations

While both reversible computations and bidirectional transformations are small but growing research fields, testing is large and well established. No one will doubt that testing is immensely important to reliability of software and new techniques are often presented; from finite state-machine based models to recent year's frameworks for automatic randomized test generation.

It is interesting that all three fields share the possibility to improve reliability of programs, though there is a difference in the foundations behind. This is why we find that increased cross-collaboration can result in interesting new research. This both researching test algorithms for reversible and bidirectional programming models, and exploiting the structures of backward execution models in testing subsets programs without backwards execution.

## Goals and results of the meeting

Within the European networking project COST Action IC1405 on reversible computations some of this work has been initiated. Initial collaborations crossing between these fields have been established, also involving organizers of this Shonan meeting. However, this is primarily a European project.

A focused Shonan meeting will, firstly, give a better possibility to push the ideas and discussion out to a wider audience and, secondly, give better involvement of the international community (and more specifically the Japanese) in the fields. The four days is the perfect opportunity for networking between the areas. We expect that the invited experts will bring back a view into a new area and ideas for future collaborations.

(Excerpt from the submitted proposal)

# The 2nd Controlled Adaptation of Self-adaptive Systems (CASaS2017)

# 094

Shonan Village Center, July 24–28, 2017

Seminar No.110

## Organizers

David Garlan, Carnegie Mellon University, USA

Nicolás D’Ippolito, Universidad de Buenos Aires, Argentina

Kenji Tei, National Institute of Informatics, Japan



Self-adaptive systems are required to adapt its behavior in the face of changes in their environment and goals. Such a requirement is typically achieved by developing a system as a closed-loop system following a Monitor-Analyse-Plan-Act (MAPE) scheme. MAPE loops are a mechanism that allows systems to monitor their state and produce changes aiming to guarantee that the goals are met. In practice it is often the case that to achieve their desired goals, self-adaptive systems must combine a number of MAPE loops with different responsibilities and at different abstraction levels.

Higher-level goals require decision-level mechanisms to produce a plan in terms of the high-level system actions to be performed. Various mechanisms have been proposed and developed for automatically generating decision-level plans (e.g., event-based controller synthesis), providing guarantees about the satisfaction of hard goals (e.g., providing a certain level of service), and supporting improvements in soft goals (e.g., doing this in an efficient or cost-effective manner). These decisions are often made at a time scale of seconds to minutes.

Lower-level goals, on the other hand, typically require control mechanisms that sense the state of the system and environment and react at a fine time granularity of milliseconds. Solutions to this problem are typically based on classical control theory techniques such as discrete-time control.

A successful adaptive system, then, must find ways to integrate these multiple levels of control, leading to an important question of how best to do that, and what concepts. Additionally, concepts from classical control theory (typically applied at low levels of control) can also be useful in understanding higher-level control.

Recently the software engineering community has begun to study the application of control theory and the formal guarantees it provides in the context of software engineering. For example, the 2014 Dagstuhl Seminar “Control Theory meets Software Engineering”, is an example of such recent interest. That seminar discussed a variety of possible applications of control theory to software engineering problems.

As in the first edition we expect to involve a group of active researchers in key areas such as Self-Adaptive Systems, Control theory, Game theory, Software Engineering, and Requirements Engineering, creating an ideal environment to discuss current and future applications and possibilities of control theory as a mechanism to provide formal guarantees for self-adaptive systems (e.g., convergence, safety, stability). Encouraged by the success of the first CASaS meeting, we expect to have a number of participants from a wide variety of research areas to further explore the benefits of incorporating the application capabilities and formal framework provided by control theory to self-adaptive systems.

Among the research questions that we expect to discuss are: How to coordinate multiple levels of adaptive control? What kinds of properties from classical control theory can be applied at higher levels to guarantee certain properties? To what extent does the domain and contest of use influence the design of a control regime for adaptation? In what ways can AI techniques of planning and machine learning be applied to adaptive systems? How can one deal with uncertainty in a systematic fashion? How can control theory inform our decisions about ways to incorporate humans into self-adaptive systems?

We envisage the 5-day meeting to be organized in two main parts. During the first day, participants will present their background and what they are interested in. Then, for the remaining four days, we will identify and discuss the most relevant topics selected by the participants in working groups.

Finally, we plan to summarize the workshop findings and propose future actions.

We expect the meeting to encourage participants from different areas and communities to interact and actively search for common interests that may result in possible collaborations and joint work.

(Excerpt from the submitted proposal)

# Software Engineering and Networked Control for Smart Cyber Physical Systems (SENCPS)

005



Shonan Village Center, August 21–24, 2017

Seminar No.104

## Organizers

Hausi Muller, University of Victoria, Canada  
Fuyuki Ishikawa, National Institute of Informatics, Japan  
Marin Litoiu, York University, Canada

Cyber physical systems (CPS) are distributed, software-intensive systems that control tightly integrated and networked computational and physical components. CPS technologies are becoming the key enablers for how we control and build smarter, context-aware and situation-aware systems, such as autonomous vehicles, smart cities and buildings, renewable energy systems, elderly healthcare, resource management, and food supply chains. The societal impact of CPS and its associated industrial revolution is enormous.

Advances in the interconnected capabilities of CPS affect virtually every engineered system. The technologies emerging from combining the cyber and physical worlds will provide an innovation and incubation engine for a broad range of industries—creating entirely new markets and platforms for years to come. CPS are advanced technology systems that require knowledge and training for their development and operation. A skilled workforce to support future CPS is a challenge in its own right and of strategic importance.

There are many challenges that must be addressed to be able to harvest CPS's rich economic opportunities. The goal of this workshop is to bring together researchers from different fields to address these challenges. The research topics we propose to discuss include, but are not limited to, control and systems science for CPS, models and co-design for CPS, models at runtime for CPS, networked control for CPS, integrating discrete and continuous control for CPS, continuous assurance for CPS, software frameworks and platforms for CPS. The proposed attached invitee list brings together researchers from the following research communities: software engineering, cyber physical systems, adaptive systems, Internetware, networked control, and Internet of Things. These researchers will bring together different perspectives on methods, techniques, technologies, and applications from their respective fields. By working collaboratively at beautiful Shonan Village, software engineering, adaptive systems, Internetware and networked control researchers ought to generate significant synergy for cyber physical systems research.

## Keywords

Cyber physical systems (CPS), Internetware, Internet of Things, Industry 4.0, smart, contextaware and situation-aware systems, models for CPS, composition, abstractions and integration of CPS systems, continuous assurance, models at runtime, networked control, integration and co-design of discrete and continuous control, uncertainty, feedback loops, software engineering, software evolution.

(Excerpt from the submitted proposal)

# Memory Abstraction, Emerging Techniques and Applications

# 096

Shonan Village Center, September 11 – 14, 2017

Seminar No.108

## Organizers

Bor-Yuh Evan Chang, University of Colorado Boulder, USA

Xavier Rival, INRIA, France

Suhyoung Ryu, KAIST, Korea



## Memory abstraction

Static analysis aims at computing semantic properties of programs, so as to verify properties such as absence of runtime errors, functional correctness, termination, security, and more. Since these properties are generally undecidable, automatic program analysis tools usually need to be conservative so as to attempt to prove properties of interest. In that process, they need to reason about all components of the semantics of programs (numeric and symbolic computations, parallelism, etc).

The notion of abstraction is central in static analysis: an abstraction defines a set of logical properties that a program analysis tool may use, their meaning, and the supporting algorithms. Numeric abstractions made strong progress early in the development of static analysis. However, to reason about complex languages (be it C, Java, ML, or JavaScript), analysis tools also need to use a careful abstraction of the structures stored in memory.

Many kinds of memory abstractions have been introduced in the last thirty years. Initially memory abstractions mainly consisted of pointer abstractions based on aliasing graphs or points-to relations. Such abstractions cannot cope precisely with data-structures of unbounded size, thus shape analysis techniques were introduced in the 1990s. Shape analysis techniques rely on more complex mathematical objects to describe unbounded memory graphs by summarizing inductive patterns. In the last decade, a large number of novel techniques and applications for memory abstractions have emerged. Even though the development of memory abstractions still seems to be lagging compared to, e.g., numeric abstraction, recent progresses seem to make this field more mature.

## Meeting objectives

The purpose of the meeting is to bring together experts in the design and in the use of memory abstractions, so as to leverage on the recent advances in this field, allow for the development of new fundamental principles and tools, and ease the use of memory abstractions in emerging applications. In the next two sections we elaborate on the two main aspects of the envisioned seminar, namely the memory abstraction techniques, and their applications. For each aspect, we explain why we think such a meeting would be timely, and outline the expected benefits.

## Novel applications of memory abstractions

A second important trend is the emergence of new applications for memory abstractions, to cope with different programming languages to analyze and families of target properties to verify.

Quite early, memory abstraction techniques have targeted imperative and object oriented programming languages (such as C, C++, Java). In the last few years, several programming languages have emerged as new targets for memory abstractions. For instance, JavaScript (and other web/scripting languages) has become very popular and raises specific verification problems: as it is untyped and very dynamic, reasoning statically over programs is extremely complex, and requires the development of novel static analysis and memory abstraction techniques.

The set of target properties of interest has also grown. Safety properties such as absence of certain classes of errors (null or invalid pointer dereferences) was a natural first target for static analysis and memory abstraction. Structural invariance properties state that certain data structures should remain well formed at all times, and form a very interesting class of (more complex) safety properties. Additionally, analyses to establish liveness properties such as termination also require precise information about memory. Moreover, in the last few years, security properties have become a huge concern (at the level of operating system components, web services, communication and cryptographic protocol implementations). These also require to first compute a precise description of the memory data structures manipulated in programs.

The envisioned seminar will allow to gather designers of memory abstractions and potential users, and will provide an opportunity for these two groups to better exchange on outstanding problems. We expect new collaborations to emerge from such a meeting.

(Excerpt from the submitted proposal)

# Logic and Computational Complexity

# 097



Shonan Village Center, September 18–22, 2017

Seminar No.096

Organizers

Yijia Chen, Fudan University, China

Rodney Downey, Victoria University of Wellington, New Zealand

The discipline of theoretical computer science has its early roots in the pioneering work of Church, Turing, and Gödel. Two important branches of theoretical computer science were already visible right from the beginning: One oriented to computational complexity and algorithms, the other to logic, semantics, and formal methods. The two branches have quite different goals and problems, each developed methods of its own, and they partly use different mathematical tools. Even though their division has been growing steadily during the last 30 years, the two branches come together from time to time as witnessed by the work in areas as descriptive theory, proof complexity, and more recently, parameterized complexity. The main focus of the planned meeting are those areas.

Probably the theorem of Büchi and Trahtenbrot characterizing the languages accepted by finite automata in terms of monadic second-order logic (MSO) can be viewed as the first main result in the area of descriptive complexity. However, the systematic development of descriptive complexity (or finite model theory at large) started with Fagin's seminal work. It shows that the complexity class NP consists precisely of the problems definable in existential second-order logic (ESO). It is well known that all major complexity classes have such a characterization in terms of an appropriate logic.

Of course, also in **proof complexity** the central open problem is the P versus NP question of whether there exists a polynomial time method of recognizing tautologies. A related research area is the question of proof lengths. In this area, the central questions concern the minimum lengths of proofs needed for tautologies in proof systems. A proof system P is polynomially optimal if for any other proof system P' there is a polynomial algorithm transforming every proof in P' of a tautology into a proof in P of the same tautology. In particular, the length of the proof in P is polynomially bounded in the length of the proof in P'. Recently, it was shown that the existence of a polynomially optimal proof system for tautologies is equivalent to the fact that a certain logic, introduced by Blass and Gurevich, is a logic for P.

**Parameterized complexity theory** provides a framework for a refined analysis of hard algorithmic problems. A specific structural property of a given problem is identified (called the parameter). It is expected to be small in typical instances of the problem. Then, the (parameterized) complexity of the problem is measured in terms of its parameter and input length.

Logic shows up in this area in many different ways. For example, logic yields the framework for algorithmic meta-theorems. These theorems give sweeping explanations for the existence of many efficient algorithms on special graph classes. For instance, Courcelle's Theorem yields linear time algorithms for all problems definable by MSO on graphs of bounded tree-width. The area of algorithmic meta-theorems uses deep tools from both model theory and structural graph theory.

Furthermore, computational problems from logic, e.g., weighted satisfiability problems for propositional logic and model-checking problems for first-order logic, are used to define (or, to characterize) classes of parameterized intractability.

In the planned Shonan meeting, we aim to bring together researchers from both communities, complexity and logic, working in the areas mentioned above. So they can share their recent work and discuss research problems. The meeting will consist a number of tutorials, survey talks, and research talks.

(Excerpt from the submitted proposal)

# Analysis and Verification of Pointer Programs

# 098

Shonan Village Center, October 2–5, 2017

Seminar No.100

## Organizers

Marieke Huisman, University of Twente Enschede, The Netherlands

Thomas Noll, RWTH Aachen University, Germany

Makoto Tatsuta, National Institute of Informatics, Japan



Computer software is ubiquitous in today's information society, and ensuring its correctness is of great importance. This is particularly true for safety-critical systems, which occur in transportation, communication, healthcare, and many other application domains. Consequently, software failures can have severe consequences. To tackle the problem of correctness and reliability of such software, formal methods are increasingly being employed. They enable the exhaustive and mathematically founded analysis of all possible behaviours of a computer program and the verification of properties such as functional correctness. They also allow to reduce the effort and, thus, the cost of testing activities. Due to their benefits, they are increasingly becoming an integral part of the development cycle of safety-critical systems.

Many software bugs can be traced back to the erroneous use of pointers, i.e., references to memory addresses. They constitute an essential concept in modern programming languages, and are used for realising (dynamic) data structures like lists, trees etc., which are organised in the computer's memory as a so-called heap. Pointers are also abundantly present in object-oriented software such as Java collections, albeit in the somewhat implicit form of object references. Pointer-handling operations occur in device drivers, operating systems, and all kinds of application codes including those implementing safety-critical systems. The analysis of such software is a highly demanding and important task, as memory leaks, dereferencing null pointers or the accidental invalidation of data structures can cause great damage especially when software reliability is at stake. Moreover, the increasing presence of concurrency in modern computing raises additional problems such as the non-synchronised access to memory areas, which can entail so-called data races. Even worse, the formal analysis of concurrent software poses additional challenges caused by the non-deterministic execution order (interleaving) between different strands of concurrent activities.

In consequence, the complexity of state spaces arising from dynamic data structures, recursive method or procedure calls, and dynamic creation of and interleaving between concurrent threads imposes challenges which cannot be handled by standard verification algorithms such as finite-state model checking. This problem has been a topic of continuous research interest since the early 1970s. A common approach are abstraction techniques that employ symbolic representations of sets of program states using suitable formalisms. Various such formalisms have been considered for this purpose, which can be distinguished with regard to their expressiveness, precision, efficiency, and automatability.

## Aims of the meeting

This Shonan meeting shall bring together both theoreticians and practitioners working on different techniques for heap abstraction and pointer program analysis. It aims to provide a broad understanding of the various techniques to support the exploitation of their commonalities such that they can benefit from each other.

More specifically, the following questions will be discussed:

- What are the commonalities and differences between the various techniques?
- Is it possible to exploit interrelations between different methods? In particular, by utilising results from related areas such as graph theory, is it possible for tools supporting separation logic to reach a similar degree of automation as state-of-art static analyzers based on abstract interpretation?
- Some approaches are more successful in automation and extension for concurrency than others. What can these less successful techniques learn from the more successful ones?
- Which theoretical questions should be addressed in the future?

(Excerpt from the submitted proposal)



# Computation over Compressed Structured Data

# 099



Shonan Village Center, October 9–12, 2017

Seminar No.126

## Organizers

Kunihiko Sadakane, The University of Tokyo, Japan

Gonzalo Navarro, University of Chile, Chile

Data compression is used to store data in devices with small storage or to exchange data over small band-width networks. To process transferred or stored data efficiently, it is better to process them without decompression. Typical processes on compressed data are searching in compressed data, inserting into compressed data, or deleting from compressed data. Another merit of processing data without decompression is that we can reduce secondary memory access. Furthermore, we can store a large percentage of data into main memory, resulting fast processing. There is a possibility that the performance gain outweighs the computational overhead resulting from working on compressed data. We consider processing structured data such as strings, trees, or graphs.

The aim of the meeting is to bring together researchers from various research directions of compression of structured data. The meeting will inspire a discussion on theoretical results and practical implementations of compression algorithms. Especially, we consider the following topics:

- Encoding Data Structures.

It is sometimes enough to construct data structures which support some queries on data but which do not recover the actual data. The size of such data structures can be much less space than the entropy of the data. These data structures are called encodings. A typical example of an encoding is the  $2n$ -bit structure that answers range minimum queries on a permutation of  $[1; n]$ , whose entropy is  $n \lg n$  bits. This direction of research is theoretically challenging and practically relevant.

- Computation-Friendly Compression.

It is difficult in general to do computation over compressed data compressed by algorithms which focus on only compression effectiveness. Recent trends are designing computation-friendly compression schemes, which achieve both strong compression and allow for efficient computation.

- Graph Compression.

Graph compression is important due to the growth of complex networks such as social networks. There exist many results on graph compression, but there are still important open problems such as suitable definitions of graph entropies, graph algorithms on compressed graphs, designing benchmark data, etc.

- Machine Learning.

The importance of efficient algorithms for machine learning rapidly grew recently. It will be possible to accelerate learning if data are compressed.

# 100 Geometric Graphs: Theory and Applications

Shonan Village Center, October 30–November 2, 2017

Seminar No.106

## Organizers

Naoki Katoh, Kwansai Gakuin University, Japan

Hee-Kap Ahn, Pohang University of Science and Technology, Korea

Subhas C Nandy, Indian Statistical Institute, India



The main goals of the workshop are to facilitate the exchange of tools, techniques, questions, and ideas that will lead to a better understanding of geometric graphs that arise in different applications. Our objective is to gather researchers working on these topics in order to exchange the ideas for further research. The focus will be mainly on open problems and exploring possible approaches to solve them. We are also planning to arrange 3-4 plenary talks (45-60 min) from eminent researchers in these fields, and 5-6 shorter talks (25-30 min). The invited attendees of this workshop will also be requested to present open problems. The talks and presentation of open problems will be in the forenoon session. The afternoon session is left for discussion among researchers for solving those problems. During the workshop, we plan to focus on the following research areas:

### Combinatorial questions on geometric graphs

The study of geometric graphs is a classical topic in computational geometry. Many practical problems have natural models via geometric objects. For example, map labeling, problems in wireless and sensor networks and VLSI physical design, database queries, etc. Here the vertices of the graph are mapped to some geometric objects and an edge between a pair of vertices indicates the existence of some specific relation depending on the problem. Here, the objectives are three fold, namely (i) characterizing the graph to have some desired embedding in  $\mathbb{R}^d$ , for some chosen  $d$ , depending on the problem specification, (ii) combinatorial questions regarding various properties of the graph, for example, minimum number of layers required to draw the graph in planar way, coloring vertices/edges with minimum number of colors to avoid conflict among the objects/paths, showing the relationship among different parameters of the graph, namely coloring number, clique size, etc., (iii) algorithmic questions regarding polynomial time computation of some parameters of the graph, or approximating the parameters to a desired accuracy in polynomial time, etc.

### Optimization problems on complete geometric bipartite graphs

Similarly, for a geometric bipartite graph  $G$  with bichromatic point set  $R \cup B$ , where the set  $R$  (resp.  $B$ ) are red (resp. blue) points,  $|R|=|B|$ , the following optimization problems require further research: (i) computing the minimum weight maximal planar matching, (ii) computing a perfect matching with minimum number of crossing, etc.

For a geometric graph with  $n$  points with edges colored by  $k$  colors ( $k < n - 1$ ), a minimum length spanning tree with at least one edge of each color may be an interesting problem to study. In addition, there are a lot of open questions related to the bounds of the spanning ratios on such geometric graphs.


### Random Geometric Graphs

In recent years, a lot of progress has been made in the study of geometric intersection graphs, but many important combinatorial and algorithmic questions are still open. Recently, in the network science community the interest is growing towards the geometrical characterization of real network. A large real network is usually considered as a random graph (Erdos and Renyi (1960)). It is assumed that each pair of vertices in the network is connected with probability  $p$ , and is independent of the other edges of the graph. Here, the desired problems are modeled by random walk on geometric graphs for the average case analysis of the performance of the corresponding network. On the other hand, random geometric graph was first formally suggested by Gilbert (1961), whose vertices are random points in Euclidean plane and an edge between a pair of points (nodes) exists if their distance is less than a given constant  $r$ . These classes of graphs are known to have numerous applications as a model for studying communication primitives (broadcasting, routing, etc.) and topology control (connectivity, coverage, etc.) in idealized wireless sensor networks as well as extensive utility in theoretical computer science and many fields of the mathematical sciences, namely routing problems in the internet, data mining, community detection, to name a few.

(Excerpt from the submitted proposal)



INTRODUCTION  
OF  
SHONAN MEETINGS  
AND  
COMMITTEE



# NII SHONAN MEETINGS

The NII Shonan Meetings are, like the well-known Dagstuhl Seminars, another premier venue for world-class scientists and promising young researchers to come together and share their knowledge, discuss their research findings, and explore cutting-edge topics in the field of informatics.

The Meetings are held at Shonan Village Center near Tokyo under the auspices of the National Institute of Informatics (NII) in Japan. The Center offers excellent conference and training facilities and comfortable lodging. Meeting participants can relax in the open, friendly atmosphere of the Center where the emphasis is on intercultural communication. Rail transportation from Narita International Airport to the center is direct and convenient.

## **Dagstuhl-Style Seminar**

The NII Shonan Meetings have the aim of promoting informatics and informatics research at the international level. They follow the style of the well-known Dagstuhl Seminars and are basically held once a month. The academic committee of NII reviews the proposals submitted by at most three organizers who are established leaders in their field. Each proposal includes a list of invitees to the meeting. The organizers should preferably be from different institutions, and one of them must be from an Asian country. NII invites on their behalf 25 to 35 researchers of international standing from academia and industry.

## **Meetings with Productive Outcomes**

The meetings typically do not have fixed programs, and participants are not required to submit a paper or make a presentation. Instead, the topics and programs evolve from the discussions at hand, and participants are encouraged to present and discuss new ideas and work in progress.

## **Administrative Support**

The staff at NII handles the sending of invitations, making arrangements for accommodation, etc. This leaves the organizers with the main responsibility of choosing interesting topics and the best participants.

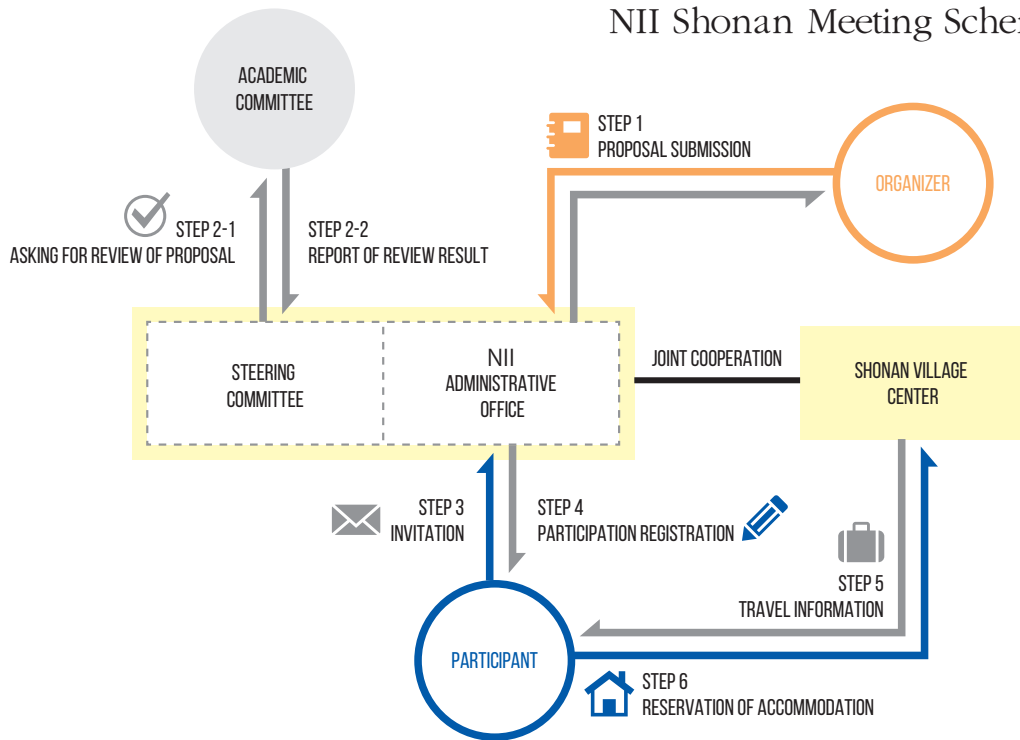
## **Reasonable Accommodations**

Overnight accommodations, including a single room and three meals (breakfast, lunch, and dinner), are provided free of charge to organizers. Participants from academic organizations are charged 8,000 yen per day, and those from industrial organizations pay 15,000 yen per day. Travel costs are not included.

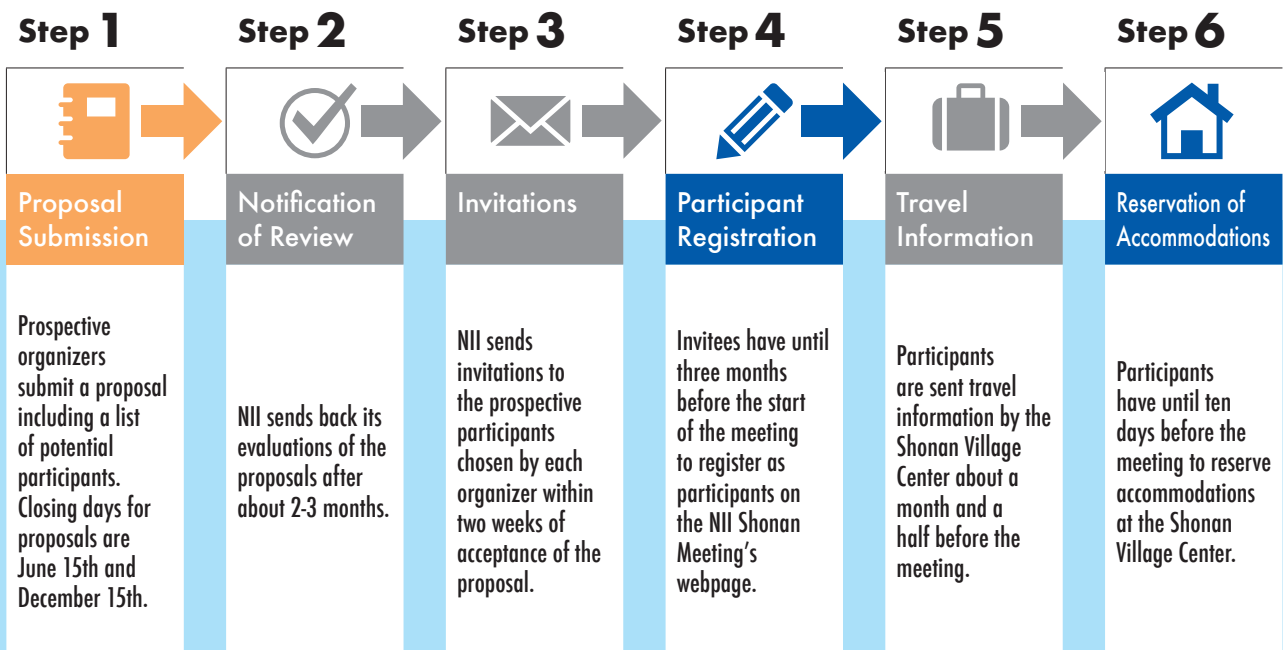




## NII Shonan Meeting Scheme



## Time Frame



**Steering Committee**

Akiko Aizawa (NII): Chair

Isao Echizen (NII)

Cheung Gene (NII)

Zhenjiang Hu (NII)

Michihiro Koibuchi (NII)

Ken-ichi Kawarabayashi (NII)

Junichi Yamagishi (NII)

Masako Suzuki (NII)

**Academic Committee**

Zhenjiang Hu (NII): Chair

Ken-ichi Kawarabayashi (NII) : Vice-Chair

Jeremy Gibbons (University of Oxford)

Michael E. Houle (NII)

Katsumi Inoue (NII)

Sadao Kurohashi (Kyoto University)

Kae Nemoto (NII)

Bashar Nuseibeh (Lero/The Open University)

Makoto Onizuka (Osaka University)

Shin'ichi Satoh (NII)



# IMPRESSION OF PARTICIPANTS



**Question 1** "Collaborated research": Please mention if you obtained any effective results after organizing the Shonan Meeting.

<b>Favorable</b>	1	Yes, collaboration and collaboration opportunities. Joint publications, follow up joint meetings	
	2	Yes. I found a potential researchers for a new joint works.	
	3	Yes, I am now working on formal proof of implicit computational complexity with people I met for the first time at the Shonan meeting on ICC.	
	4	Several joint work started after the meeting.	
	5	Yes. Thanks to the discussion I had with one of the researchers who participated the same Shannon meeting, we later collaborated on a research project, which lead to a publication at a top venue in the field.	
	6	Yes, four different collaboration areas where defined as result of the meeting activities and discussions. I am currently involved in one of these initiatives which is related to a series of special sessions and workshops in chatbots and conversational agents (WOCHAT)	
	7	This is the most important issue. However, if we have this type of meetings, such positive effects automatically come. Their levels may be different for each meeting, but I do not think it the main issue. As far as the one I organized, it was quite a bit ago, and as far as I remember, it was nice in the sense that high-level researchers in a single topic can get together and were able to make deep discussions.	
	8	Thanks to the Shonan meetings I have co-organized I could strengthen my research collaboration network. I have now stronger collaborations with colleagues in the UK and in the US. We have had some PhD students exchanges and we have worked together on some research projects.	
	9	Yes, we published a very good book in CRC on a very important emerging topic.	
	10	The two Seminars on program generation and High-performance computing have indeed fostered collaboration: we have collected Shonan Challenge problems that have inspired several implementations. There was a PEPM 2012 paper specifically about it.	
	11	Not directly, but got several new insights, which lead to some new results.	
	12	Extensions of existing collaborations, new collaborations, and joint publications.	
	13	Yes. The meeting lead to several papers at PLDI and ICFP conferences.	
	14	Yes, the Shonan meetings I've attended and run have created new collaborations ... there is enough time and enough focus to find subsets of workshop participants that have an inclination for new collaborations.	
	15	The Shonan Meeting on Big Data for Disaster Management refers to a topic that was part of my ongoing research project. The collaborations I have, consist of relationships that have been ongoing.	
	16	A book will be published from Springer.	
	17	Collaboration with Michael Witting (Munich) on retention time prediction, will result in a joint grant proposal in the near future.	
	18	I started new research collaborations at the international level. The new collaborations led to a paper accepted at the International Symposium on Graph Drawing and Network Visualization, GD 2017.	
	<b>Positive</b>	19	The meeting was very productive. The community had been able to make progress to the goal of further unifying different strands of research. For example, there has been intensive discussions across the programming languages and software engineering groups. Personally, I managed to establish new academic and industrial connections
		20	We obtained great collaboration results in the software engineering community in general and adaptive systems and cyber physical systems communities in particular.
		21	Yes. Paper about transparency instead of hiding data for privacy
		22	Yes. We have made good progress in probabilistic model checking research work.
		23	We are going to publish a book writing with about 30 researchers including participants of NII meetings. Also, my laboratory has some co-authorships with foreign researchers participated NII meetings.
		24	yes, various collaborations which lead to joint projects and publications
		25	The meeting was good opportunity to join different research groups. We severally met at conferences but had few chance to discuss deeply.
		26	The members published several papers. Among participants from Tohoku U. I know we published at least 6 papers jointly with foreign participants.
		27	We are writing a collaborative survey paper based on the discussion held in the seminar.
		28	I have published a journal paper together with Prof. xxx, who attended the Shonan seminar;
		29	Yes, working with Prof. xxx and published two papers after this meeting. I have also published a paper with collaborators from Germany who attended the meeting. I also gave a talk in Germany with one of the attendees of Crowdsourcing meeting I attended in Shonan.
		30	We have continued and intensified our collaborations after the Shonan Meeting
		31	The publication of a book with Springer
		32	I did. We produced a very interesting and relevant report that stablished the basis for future collaborations.
		33	Yes, several participants subsequently engaged in collaborative projects.
		34	Yes, I have written a paper with xxx and xxx who were participants/organizers of our Shonan Meeting in 2012. (I should mention that the last two authors were my collaborators since 2009, but collaboration with the first author, xxx, was not made possible without having the Shonan Meeting. (I first met him in the Shonan meeting) I have authored many papers with another organizer Oleg Kiselyov after we have co-organized the Shonan meetings.
		35	Yes, the seminar "Immersive Analytics" led to new research collaborations and to a book project.
		36	I started a collaborated research with Prof. xxx, who attended the Shonan meeting which I co-organized. In practice, I send one of Ph.D. candidates to his lab for six months and I invited him to stay with us as a special guest professor to Keio University. As a result, we published two papers directly from the collaborative research. The student received his Ph.D. and was employed by Microsoft.
		37	Strengthened collaboration between my research groups and colleagues at NII, leading to many individual collaborations
<b>Neutral</b>	1	Shonan meetings are like dagstuhl establishing a network and structuring the new topic. In my case transparency for privacy was influential on European privacy directive.	
	2	We discussed a collaboration with several of the attendants with the aim of creating an EU project proposal. Unfortunately, due to Brexit we could not pursue this collaboration further.	
	3	This Shonan school on Coq was not targeting new scientific results, but teaching.	
<b>Others</b>	1	We did not obtain any specific result yet.	

**Question 2** "New community formation": Please mention if you had any effective results as a consequence of the Shonan Meeting.

<b>Favorable</b>	1	I personally got introduced to several "neighbouring" research areas and got inspired by the gained insights and the introduced tools and techniques.	
	2	Established european network on new paradigms for privacy.	
	3	Yes, several new contacts and collaboration. I.e. research visits to their groups and to my group.	
	4	Yes, I have given an invited talk at DICE 2016 to present my new results since the Shonan meeting on ICC.	
	5	We hope we were able to foster a community of Coq users in Japan.	
	6	We have found new research connections between hardware-oriented and cryptography-oriented researchers.	
	7	I am not new community has been formed. But I did join some community, and developed some new research profile after joining some Shannon meeting.	
	8	Yes, the WOCHAT initiative will have its fourth event in SIGDIAL 2017 next month. A shared task has been also organized as part of the WOCHAT initiative, and up to date we have around 20 shared task participants and more than 30 papers presented at the different WOCHAT events.	
	9	We conducted a collaborative project under India-Japan cooperative science program (JSPS) from 2015 to 2017 with Prof. Pushpak Bhattacharyya, who participated in our Shonan meeting.	
	10	Yes, I could meet several people, which helped me to organize some other meetings later.	
	11	Immersive Analytics community was formed.	
	12	We started new collaborations in the area of behavioral typing and submitted research proposals together.	
	13	The continuation of the Competition on Legal Information Extraction and Entailment (COLIEE) is one important consequence of the NII meetings. At the COLIEE competition in London last June, several of the Shonan Meeting participants were present and interacted with visitors from Japan, North America, and Europe.	
	14	The area of Big Data for Disaster Management is an active area of research worldwide. The Shonan meeting contributed to its growth and vitality.	
	15	The meeting connected some people, and they continue to work together. For example, xxx continues to work with researchers in xxx University who participated in the meeting. I continue to work with xxx and xxx.	
	<b>Positive</b>	16	New Research Community on "Visual Analytics" in Asia-Pacific Region was formed. PacificVAST workshop will be collocated with IEEE PacificVis conference.
		17	The Shonan Meeting was most useful for engaging intensively, in a short time frame. The contacts I made still exist.
18		The Shonan seminar helped significantly to further develop the Computational Metabolomics community. We are now part of the CompMS group ( <a href="http://compms.org/">http://compms.org/</a> ), and there is also a task group for this topic as part of the international Metabolomics Society. In particular, all participants signed a Memorandum of Understanding to initiate long-term research cooperation, promote open data standards and open data, exchange researchers and hold joint symposia. The community building will be continued by further similar seminars, such as a Dagstuhl seminar in 12/2017.	
19		The community was already formed before the meeting. But the meeting has contributed to the continued bounding of the the community. There is a following on Dagstuhl meeting being organized as a result of the Shonan meeting	
20		We did not create a new community as a result of the seminar, but for me the main value of the seminar was to strengthen the links with the other attendants in an informal way. Thanks to the Shonan Seminar it is now much easier to collaborate with the attendants.	
21		The Shonan meeting held already three times on Engineering Adaptive Software Systems (EASSy) formed a very cohesive community. The Shonan meeting just completed in August formed a new high international community on Software Engineering and Networked Control for Smart Cyber Physical Systems (SENCPS).	
22		Not in Japan but US and Europe	
23		We have formed new collaborations with various colleagues.	
24		Writers of the above-mentioned book are making a new community.	
25		yes, extended my network of personal connections	
26		I and (young Japanese) participants start discussing about research collaboration with researchers in other country.	
27		We formed a new community based on the meeting. We organized several workshops with the members of the community including <a href="http://mlhcr2016.tanichu.com/">http://mlhcr2016.tanichu.com/</a> <a href="http://mlhcr2017.tanichu.com/">http://mlhcr2017.tanichu.com/</a> in IROS 2016 and IROS 2017. The member of the community is going to be in charge of general and program committee of ICDL-Epirob 2018 in Tokyo. <a href="http://icdl-epirob.org">icdl-epirob.org</a>	
28		Assistant Professor xxx (xxx University), who attended our seminar is currently visiting xxx, Austria (university of Prof. xxx, my co-organizer for our Shonan Seminar) for 6 months.	
29		There is some effort in forming a community around big data management and analytics. Some effort has been made by presenting some work in recent US-Japan meeting with IEEE ICDCS 2017.	
30		During the meeting contacts were made to initiate the joining of an Asian database joining an international consortium for data exchange.	
31		Most participants knew each other very well already	
32		indeed we have joined together two communities that didn't collaborate much before.	
33		Yes, there is now an active community research community engaged in the problem of evaluation of whole session information retrieval. NCTIR, TREC and CLEF evaluation campaigns have added this issue to their agendas.	
34		As a result of Shonan Meeting in 2012, we have initiated 'Shonan Challenge for Generative Programming', which consists of challenging problems in the area of generative programming (or, staging). We have set up a Github page ( <a href="https://github.com/StagedHPC/shonan-challenge">https://github.com/StagedHPC/shonan-challenge</a> ) for the problem set as well as inviting solutions written in various programming languages. This challenge actually worked very well, which attracted many researchers' interest. In fact, we were able to organize the second Shonan meeting in 2014, based on the strong interest in Shonan Challenge.	
35		Yes, the novel area of "Immersive Analytics" gained a lot from the Shonan seminar.	
36	This July, we founded the consortium of Japan-China Visualization (CJ-Vis). The representative from China side is Prof. xxx from xxx University, who was invited by the Shonan meeting which I co-organized. I am serving as the representative from Japan side.		
37	Through a series of three Shonan meetings on "Engineering Adaptive Software Systems" (EASSy), this community strengthened and made better links with the wider community in the area.		

**Question 3** "External funds raised": Please mention if you obtained any funds as a result of the Shonan Meeting.

<b>Favorable</b>	1	National project (IoT Security driven by NEDO in Japan) has started since 2016.
<b>Positive</b>	2	I successfully obtained one research grant (in the range of half million) to carry out some research idea which was formed during the meeting.
	3	Not yet, but we are currently exploring different possibilities for raising funds to support the workshops and shared task activities.
	4	The topic we organized has been continuously popular in theoretical computer science and I am sure that there have been a lot of opportunities for funding in both Japan and in the western world.
	5	After a bit modification from the topic discussed at our Shonan meeting, our proposal was accepted as a JST CREST project.
	6	I believe that xxx from NII has recently been successful in receiving a grant to support ideas in legal reasoning, arising from discussions at a Shonan meeting. I am a collaborator in that project.
	7	We applied a joint grant application, which is on the review process in Australia.
	8	I won an Australian Research Council Linkage Project in the same year. The intention to coordinate the Shonan Meeting in that year perhaps helped in the proposal assessment.
	9	See above, grant proposal planned shortly.
	10	I put together a proposal with xxx and xxx, we discussed this during the Shonan meeting but the topic was different from the topic of the meeting. This proposal for a 5-year project was funded by EPSRC.
	11	We have received multi-millions funding in security system verification.
	12	yes, a grant proposal was accepted by the Swiss National Science foundation which was based on the topic of the Shonan meeting I co-organized
	13	I have won a 2-year JSPS Bilateral Joint Research Project grant, together with Prof. Saakes from KAIST, Korea. Prof. Saakes attended our Shonan seminar. I have submitted a JSPS Visiting Professor Grant together with Prof. Swan (USA, co-organizer of
	14	We are still working towards this direction and hope to have some supplement funding.
	15	As our meeting was only this year, we have not yet received additional funding.
	16	I received funding for a 3-year project from the US NSF, aspects of which were directly influenced by results from our NII Shonan Meeting.
	17	Not many, but all my Kakenhi fundings after 2012 are somehow related to the results of Shonan meetings in 2012 and 2014. In particular I got Kakenhi (B) from 2013-2015 (17,290,000 yen).
	18	Not yet, but there is the plan for funding applications.
	19	The meetings were mentioned on grant proposals in the UK and Ireland, which were later funded, although of course the meetings were not the only reason for obtaining funding.
	<b>Neutral</b>	1
<b>Others</b>	1	We have not obtained any funds.

**Question 4** The others, please mention if you had any effective results.

<b>Favorable</b>	1	I had many. Eg stopped work on resilience from computer science. Clear term is a requirement, but not translatable in algorithms
<b>Positive</b>	2	We could adapt to the different levels of the students thanks to the usage of several rooms in parallel.
	3	A lecture by xxx on contracts alerted me to a new problem, and later let me spot the significance of one of my student's results.
	4	Acceleration in research through the discussion.
	5	We have definitively benefited from the meeting with a better network of researchers and different collaboration initiatives.
	6	Very nice to increase the reputation of Japan.
	7	We had a number of follow up academic visits between the participants after the workshop to continue research collaboration.
	8	Dagstuhl-like seminars are not known in the metabolomics community; several attendees from metabolomics (biology, chemistry, mass spectrometry) liked it a lot in the end! We had a lot of fun, we made much progress, and it was almost cherry blossom; thanks for all your support!
	9	Another effective result for me personally was the continued collaboration with Prof. xxx of xxx University. The Seminar provided the opportunity to strengthen the bonds.
	10	Many papers have been produced as a result of the three Shonan meetings I co-organized
	11	We also had closer links with researchers in Japan.
	12	Lots of interesting discussions in a very pleasing setting. Shonan is amazing.
	13	In our research community, this meeting become famous like Dagstuhl Seminar. Many researchers visit our school before/after the meeting.
	14	We have published joint papers in IEEE ICDCS and IEEE SECON 2017 conferences.
	15	Strengthened contact with community members
	16	Yes, excellent discussions with people working in the area from all over the world, new collaborations and joint paper.
	17	The Shonan meeting 046 on Computer Visualization--Concepts and Challenges gave a trigger to the visualization community; indeed we had a number of visualization-related Shonan meetings after this.
	18	Met new people in the community (e.g. on feedback/control systems) whom I did not know before.

**Question 5** Please let us know if you have any suggestions or requests regarding the coming 100th Commemorative Ceremony.

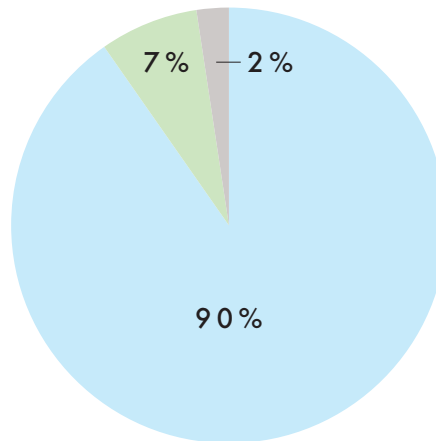
1	The topics of many seminars have character of university seminars. Very specific and excellent, but not strategic and lack inspiration and pointing in future. It is often a collection of experts. At present all sustainable solutions have interdisciplinary sources
2	Reunion of the meeting, which would be a nice chance to re-think which research direction was appropriate.
3	It would be nice to have more information about the ceremony and related events.
4	A commemoration is ok, but the most important thing is to continue this project firmly and quietly.
5	It's nice if you can broadcast the ceremony on the Internet.
6	Congratulations!
7	Invite representatives from each past seminar. Every representative could write a brief note on developments after the seminar.
8	I think it is important to convey the value of these kinds of meetings to the most senior leadership at NII (including the CEO), and to the public agencies that support Shonan. Researchers that participate in Shonan meetings, and their equivalents at Dagstuhl in Germany or BIRS in Canada all understand the value ... but this value has to be conveyed to the funders, and the 100th anniversary is a good opportunity to invite those leaders to participate.
9	It's a good idea. Collecting information on previous Shonan Meetings is a good idea. Perhaps try to search for references to Shonan Meetings in published papers (Google Scholar should have that information.)
10	It would be nice to have a big meeting with satellite workshops. Talks on new trends by invited speakers would be good.
11	Not sure. Maybe some sort of publication?
12	I have answered while i was visiting prof in NII Major problem i see for Shonan is that everybody in committee is from NII. Good people in NII. But not completely representative for future in informatics I suggest limited topics, not all that is in NII 1. security and privacy 2. fake news and AI 3. 3 big data and big data analytics
13	Don't stop
14	Please provide simultaneous interpretation for non-fluent Japanese speakers.
15	The venue, the people, the food are great. Only a clear view to mount Fuji was missing
16	I'd like to attend!
17	I really appreciate the anniversary, and I would like to thank Zhenjiang Hu and all NII people who contributed/helped organizing the Shonan meetings. Frankly speaking, organizing this kind of ceremony is not easy, as most researchers are too busy in doing his/her own research, and do not usually attend non-research events. My suggestion is to include a few research-oriented talks in the ceremony. If the topic of the talk is of broad interest (or the speaker is very popular), maybe many people would attend it. I think you can pick up some excellent speakers from the past organizers/participants.
18	Make a visible publication about the 100th seminar, maybe with contributions from all the 100 seminar topics/participants.
19	I hope to see some small leaflet or a special webpage introducing the title and organizers of all the meetings.
20	I think the key is to emphasise: 1) the wide range and number of people and disciplines represented 2) The unique opportunity to offer a reliable and high quality forum for the world wide research community and to do so in a venue in Asia (previously only Dagstuhl was available).

**Question 6** Kindly provide comments or suggestions for further development of the Shonan Meeting.

1	Insist on interdisciplinarity and strategic potential in topic selection
2	We would love to have feedback from the participants. It should of course be anonymous so that they feel free to express their real impressions of this Shonan school on Coq.
3	Please keep going! The meetings are valuable.
4	I think it will be nice to have online resources for conducting follow up meetings after the Shonan meetings. A site where researchers can share opinions, resources and ideas with other researchers that participated in the meeting.
5	The meetings are very well organized. Maybe a four days meeting is too short for people who come a long way to Japan. I would rather have a five days meeting.
6	Might be good to publish a small report combining summaries from all participants and the organizers
7	Shonan meetings proved very successful. This is also a drawback: it takes more than a year nowadays to reserve a meeting. It would be nice if the wait period were shortened (by NII booking a longer time or a bigger space).
8	I think you are on the right path, I wish you all the best for the future development. Just continue your path of improvement steadily.
9	Improve the food back to the standard of earlier years.
10	Regional collaborations (e.g., with China, Korea, and Taiwan) may help increase the activity level of Shonan Meetings.
11	Maybe 4 days are a little too long for many to participate. 3 days may be OK.
12	To get good science you need smart people sit together and talk, instead of writing another paper in their rooms separately. Moving the social area to the place where coffee is served, was definitely a good idea. It is much cozier there, we liked it. Nevertheless, Shonan is still less cozy than Dagstuhl. Anything that will make it cozier would be appreciated. Another thing is giving (particularly the younger) participants "something to do". In Dagstuhl, there is snooker, football, a music room, a sauna, bikes for free rent etc etc etc. In Shonan, there is table tennis and a pool (which most attendees do not even know about).
13	Financial support to cover travel expenses would be very much appreciated.
14	I strongly suggest to change the policy about inviting PhD students to Shonan meetings. It is my understanding that PhD students are not allowed to attend or only under exceptional circumstances. For and foremost, Japan would greatly benefit if Japanese PhD students could attend these wonderful seminars at Shonan. Secondly, international PhD students would greatly benefit in getting to know senior researchers and expanding their network in a 5-day meeting at Shonan.
15	Make committee international. Maybe keep nii dominance, but not 100%
16	All perfect.
17	I would like to welcome Ph.D. student participants. I do not think we need to prepare accommodation for students if Shonan village does not have sufficient numbers. Just we would like to invite some Ph.D. students to the meetings
18	Although I know there are strong restrictions, some of my friends who attended the Shonan meeting several times told me that we need develop other excursion programs.
19	Please get a proper coffee machine :) E.g. <a href="https://www.usa.philips.com/c-m-ho/coffee/saeco-automatic-espresso-machine">https://www.usa.philips.com/c-m-ho/coffee/saeco-automatic-espresso-machine</a> Most international researchers drink a lot of coffee. The drip coffee provided at the Shonan meeting was not good enough.
20	Keep it as is! It is a great place, I look forward to visiting again
21	I think shonan meeting organisers should be encouraged to organise longer seminars (5 days). This duration and short presentations only the first day was a great success to us. This organisation led to much fruitful discussions.
22	Nowadays, it seems really difficult to find the available dates of the venue (Shonan center). It is inevitable as Shonan meetings became very popular, but as an organizer, it is rather frustrating that we cannot say anything about the possible dates of the meeting in advance (then people in the invitee list cannot say if he/she may attend.) For the rescue, it is rather helpful to show the list of available weeks of Shonan center on the web page. When we organized the seminar in 2014, Zhenjiang Hu kindly arranged our meeting in the same week as another (already planned) meeting. In my opinion, having two Shonan meetings in the same week worked very well, as the research fields of the other meeting is the same as ours, and we were able to communicate with the participants of other meeting. In fact, some participants are 'shared' between two meetings. I guess it was an exceptional case, but maybe this is one possibility to add more available dates for Shonan meetings. (Dagstuhl seminar also allows two meetings be held in the same week. When I attended it last time, it didn't work very well as the topics of the two meetings were completely different.) Another thing: the excursion to Kamakura area is fun for the first comers, but not for all the participants. I do not think such a gorgeous excursion is necessary. Since all the participants are experts of international travel, they only need information on interesting places and restaurants, and then they would take a tour by themselves.
23	A nicer place for the evenings (for discussions but also for leisure) would be good.
24	How about explicit collaboration with Dagstuhl seminar. For example, last year, immersive analytics sub-community attempted to have two seminars in Shonan Center as well as Dagstuhl. They seem to publish a comprehensible book on the topic shortly.
25	I think it is important in the long term to have an overall academic overview of the Shonan meeting programme. Perhaps the academic committee could meet once every year or two, to review the topics and quality of the meetings, and consider the overall balance of the programme to see if certain areas are missing or duplicated.

### Question 1

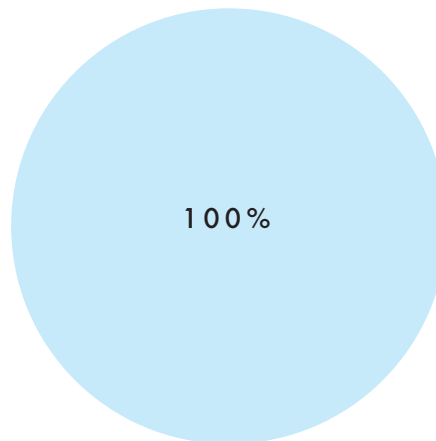
“Collaborated research”: Please mention if you obtained any effective results after organizing the Shonan Meeting.



Favorable	37
Neutral	3
Others	1

### Question 2

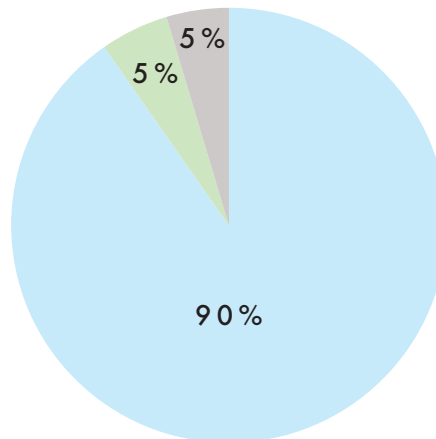
“New community formation”: Please mention if you had any effective results as a consequence of the Shonan Meeting.



Favorable	37
Neutral	0
Others	0

### Question 3

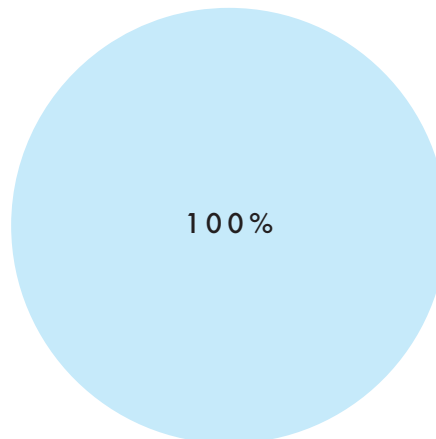
“External funds raised”: Please mention if you obtained any funds as a result of the Shonan Meeting.



Favorable	19
Neutral	1
Others	1

### Question 4

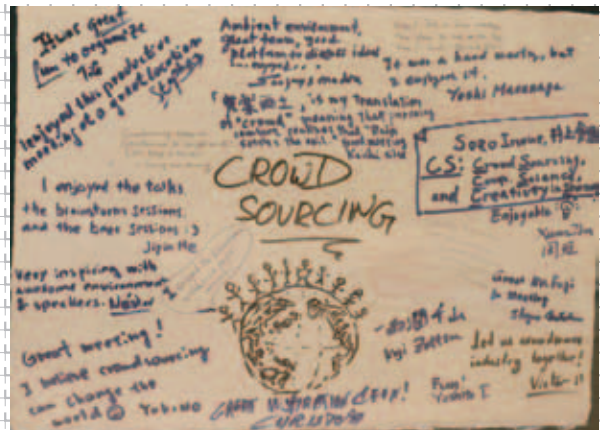
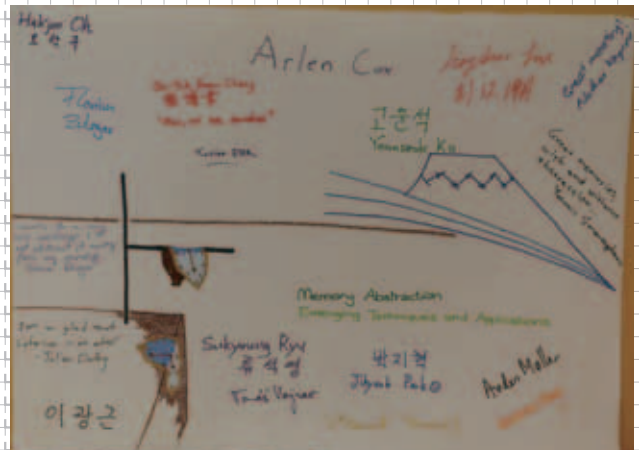
The others, please mention if you had any effective results.



Favorable	18
Neutral	0
Others	0

MEMORY  
OF  
SHONAN  
MEETINGS





SHONAN MEETING  
 PARTICIPANTS WROTE AND PAINTED  
**MESSAGE  
 BOARDS!**

**COGNITIVE SOCIAL ROBOTICS:**  
**EMBEDDED EXPERIENCES & SOCIAL INTERACTIONS**

Handwritten notes and diagrams surrounding the title, including a small drawing of a person and some flowcharts.

**GEOMETRIC GRAPHS**  
**Theory & Applications**

Handwritten notes and diagrams related to geometric graphs, including a drawing of a mountain peak and various geometric shapes like a tetrahedron and a cube.

**SENCPS**  
**Software Engineering and Mechanical Club for Cyber-Physical Systems**  
 August 21-24, 2017

Handwritten notes and diagrams surrounding the title, including a drawing of a person and some flowcharts.

**Testable Computer**  
**Testable Hardware**  
**Testing**  
**Hardware**  
**Testable Hardware**

Handwritten notes and diagrams surrounding the title, including a drawing of a pagoda and various text annotations.

**COMPUTATION**  
**OVER COMPRESSED**  
**DATA**


Handwritten notes and diagrams surrounding the title, including a drawing of a person and various text annotations.

**Logic And Computational Complexity**  
 Sept 19-22, 2017

Handwritten notes and diagrams surrounding the title, including a drawing of a landscape and various text annotations.



INTRODUCTION  
OF  
SHONAN VILLAGE  
CENTER



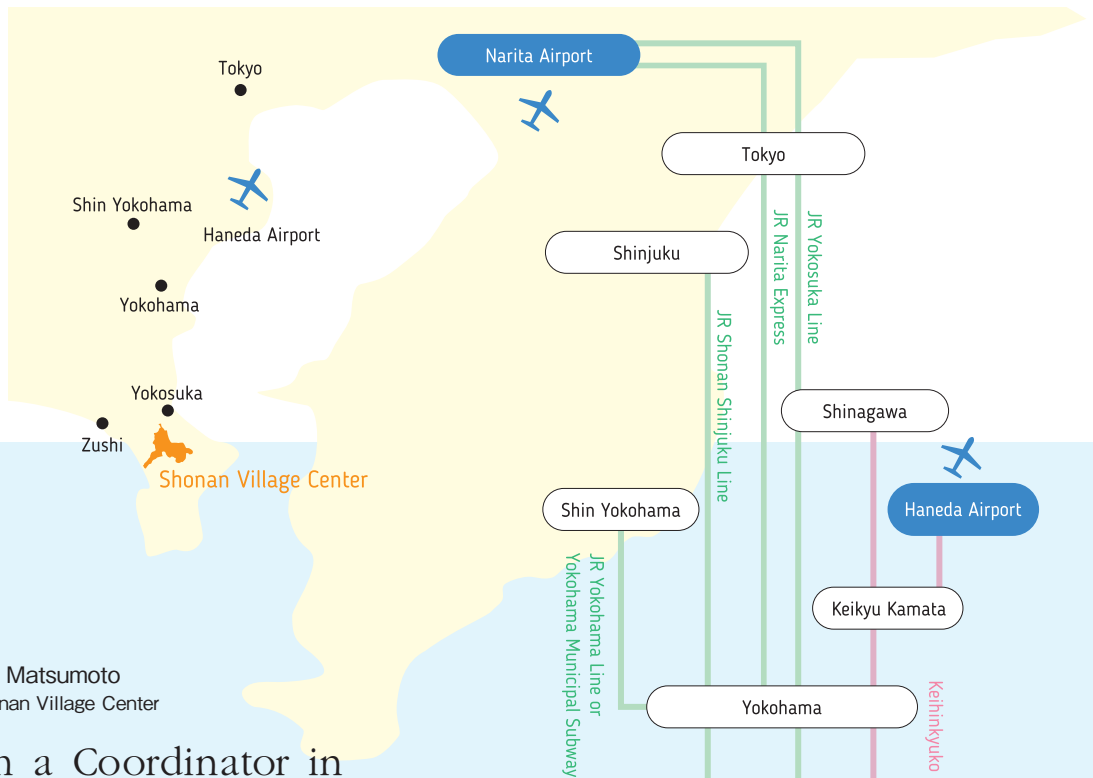


## Shonan Village Center

Shonan Village Center is set on a scenic hilltop with spectacular views of Mt. Fuji and the Shonan seaside. It offers conference goers excellent facilities and lodging in a resort-like atmosphere. The nearest train station offers direct service to Narita Airport.



# Access



Kei Matsumoto  
Shonan Village Center

## Message from a Coordinator in Shonan Village Center

This is Kei Matsumoto from Shonan Village Center, a local coordinator of the meeting. I have been communicating with a lot of participants. As a local coordinator, I would like to congratulate and give my sincere gratitude to all the organizers, participants and NII for having the Shonan Meetings at Shonan Village Center.

The very first meeting was held in February, 2011, and we were preparing for the next meeting. And then, a month later, the earthquake hit Japan and we had to postpone the meeting for 6 months. Back then, I could not imagine that the meetings would come back again. It must have been hard work for Prof. Hu, the academic committee and NII secretariat to bring the meetings back to normal again.

As of now, even the 200th meeting would be coming soon in the near future. I will still be a local coordinator of the meeting by then.



# Excursion

A half-day excursion during the seminar creates a refreshing atmosphere and deepens the communications among participants.

Kamakura, quite close to Shonan Village Center and famous for its Great Buddha, offers lots of experiences of Japanese culture and history.









Ken-ichi Kawarabayashi  
Professor

NII Shonan Meeting Academic Vice-Chair  
教授 河原林健一

Recent trend in computer science conferences is having a “mega” conference every (half) year. For example, NIPS has attracted a few thousands participants. There are many workshops co-located with these mega conferences, but they only last either a day or half a day.

Therefore, it is very difficult for participants to interact each other at a single workshop (and at a conference in general). This prompts the research community to have a “smaller” workshop, but rather last at least a few days. The famous “Dagstuhl” is the most successful of this kind. Also in mathematics, there is an “Oberwolfach” which has a long tradition. Following these two, there are a few more such workshops, such as Bertinoro, BIRS (Banff), MSRI etc.

Such a workshop focuses on some small but hot areas, so usually there are only 30-50 participants. Moreover, it lasts 4-5 days. So at the end of the workshop every participant can know all of them (and their research interest), and they can start discussing. They can even kick off new collaborations. This is the most important aspect of such a workshop.

So far we have done 100 workshops, and we have heard from many participants that they really want to come again (and again). Therefore, we feel our obligation to continue at least another 100 (and maybe another 1000). In the meanwhile, we really hope that there are some great outcomes from Shonan workshops!

