

Helsinki Institute for Information Technology HIIT

Tietotekniikan tutkimuslaitos HIIT

Forskningsinstitutet för Informationsteknologi HIIT

Helsinki Institute for Information Technology HIIT is a joint research institute of Aalto University and the University of Helsinki for basic and applied research on information technology. HIIT's mission is to conduct top-level research, seamlessly moving between fundamental methods and technologies to novel applications and their impact on people and society. HIIT is a strategic partnership of the two universities to take Helsinki IT to the world class and keep it there. HIIT's research is interdisciplinary, operating across departments and with industry. The current foci of research are computational modelling and data analysis, and ubiquitous ICT in the modern networked world.

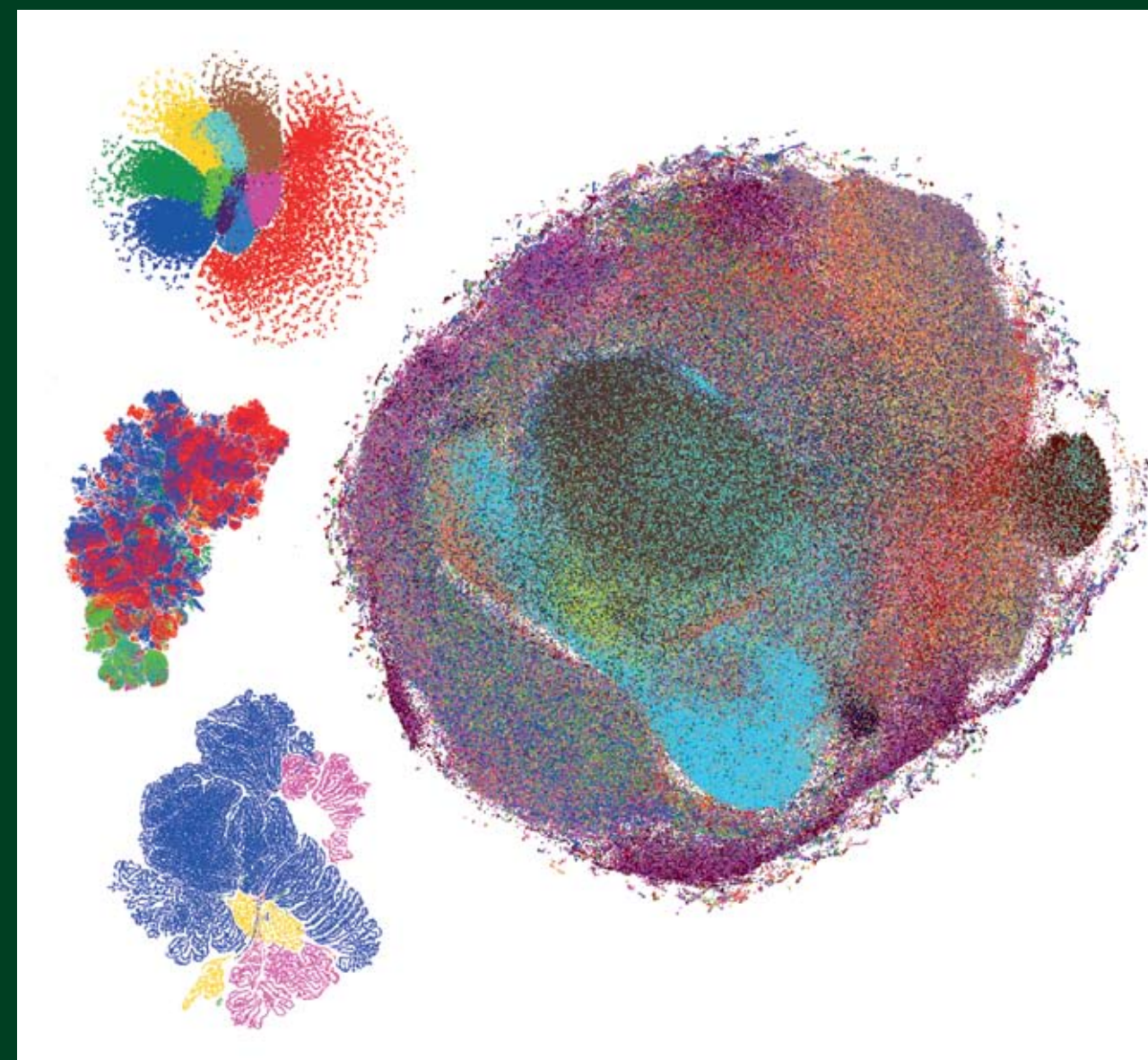
HIIT groups belong to several national Centers of Excellence and co-operate with the information industry and with sciences applying information technology. HIIT works in a multidisciplinary way, with scientists from computer, natural, behavioural and social sciences, as well as from humanities and design. HIIT partners with several international and Finnish companies as well as with universities and research institutions in Europe, North America and Asia.

HIIT was founded in 1999 and now has a budget of 10 million euros, realising in total about 154 person-years of work by about 300 researchers. HIIT is located in two different sites in the Helsinki Metropolitan area: at Aalto University's Otaniemi campus and University of Helsinki's Kumpula campus, overlapping with several departments at both places. HIIT's operating principle is that it adds to other departments' activities, instead of competing with them. The institute is led by Professor Samuel Kaski. HIIT's research is funded by Aalto University, the University of Helsinki, Tekes, the Academy of Finland, the European Union, private companies, and foundations financing Finnish hi-tech research. The research by HIIT has been assessed and deemed excellent by its Scientific Advisory Board in 2008 and 2012; in the Aalto University Research Assessment Exercise in 2009; and in the Evaluation of Research and Doctoral Training at the University of Helsinki in 2011.

www.hiit.fi

Helsinki Institute for Information Technology HIIT

Annual Report 2013



Helsinki Institute for Information Technology HIIT

Annual Report 2013

Ella Bingham (ed.)



Contact Information

Helsinki Institute for Information Technology HIIT
Tietotekniikan tutkimuslaitos HIIT (in Finnish)
Forskningsinstitutet för Informationsteknologi HIIT (in Swedish)

hiit-info@hiit.fi
www.hiit.fi

Otaniemi Site

Postal address at Open Innovation House (OIH):
Helsinki Institute for Information Technology HIIT
PO Box 15600, FI-00076 Aalto, Finland

Street address:
Aalto University, Open Innovation House (OIH), Otaniementie 19-21, Espoo
Telephone: +358 9 47001

Postal address at Computer Science Building:
Helsinki Institute for Information Technology HIIT
PO Box 15400, FI-00076 Aalto, Finland

Street address:
Computer Science Building, Konemiehentie 2, Espoo
Telephone: +358 9 47001

Kumpula Site

Postal address:
Helsinki Institute for Information Technology HIIT
PO Box 68, FI-00014 University of Helsinki, Finland

Street address:
University of Helsinki, Department of Computer Science, Exactum
Gustaf Hållströmin katu 2b, Helsinki
Telephone: +358 294 1911
Fax: +358 2941 51120

Unigrafia, Espoo 2014
Copyright © 2014 HIIT
Cover image by Zhirong Yang, Jaakko Peltonen and Samuel Kaski:
Scalable Optimization of Neighbor Embedding for Visualization
Lay-out: Anna Bergman

Table of Contents

Review of Year 2013	5
What's new	5
Collaboration and impact	7
Views for 2014	9
Algorithmic Data Analysis (ADA)	11
Computational Inference (CI)	20
Distributed and Mobile Cloud Systems (DMC)	25
Network Society (NS)	31
Administration	36
Personnel and funding	36
Board	36
Scientific Advisory Board (SAB)	38

Helsinki Institute for Information Technology HIIT is a joint research institute of Aalto University and the University of Helsinki for basic and applied research on information technology. HIIT's mission is to conduct top-level research, seamlessly moving between fundamental methods and technologies to novel applications and their impact on people and society. HIIT is a strategic partnership of the two universities to take Helsinki IT to the world class and keep it there. HIIT's research is interdisciplinary, operating across departments and with industry. The current foci of research are computational modelling and data analysis, and ubiquitous ICT in the modern networked world.

HIIT groups belong to several national Centers of Excellence and co-operate with the information industry and with sciences applying information technology. HIIT works in a multidisciplinary way, with scientists from computer, natural, behavioural and social sciences, as well as from humanities and design. HIIT partners with several international and Finnish companies as well as with universities and research institutions in Europe, North America and Asia.

HIIT was founded in 1999 and now has a budget of 10 million euros, realising in total about 154 person-years of work by about 300 researchers. HIIT is located in two different sites in the Helsinki Metropolitan area: at Aalto University's Otaniemi campus and University of Helsinki's Kumpula campus, overlapping with several departments at both places. HIIT's operating principle is that it adds to other departments' activities, instead of competing with them. The institute is led by Professor Samuel Kaski. HIIT's research is funded by Aalto University, the University of Helsinki, Tekes, the Academy of Finland, the European Union, private companies, and foundations financing Finnish hi-tech research. The research by HIIT has been assessed and deemed excellent by its Scientific Advisory Board in 2008 and 2012; in the Aalto University Research Assessment Exercise in 2009; and in the Evaluation of Research and Doctoral Training at the University of Helsinki in 2011.

Total numbers	2009	2010	2011	2012	2013
Funding M€	9,0	12,1	10,5	11,5	10,0
External/competitive funding	71 %	83 %	78 %	83 %	79 %
Person-years	185	175	135	144	154
Refereed international publications	177	222	213	194	236



Review of Year 2013

What's new

Two main goals of HIIT are to (1) conduct high level strategic research bridging the two universities, and (2) coordinate big common ICT issues across departments to form a common "Helsinki ICT". Year 2013 again witnessed success along both lines.

Signs of recognition of funding bodies to research collaboration

In 2013 HIIT contributed to four Centres of Excellence of the Academy of Finland: Computational Inference (COIN, 2012-2017, overlapping with HIIT's CI programme), Inverse Problems (2012-2017), Cancer Genetics (2012-2017), and Algorithmic Data Analysis (Algodan, 2008-2013, overlapping with HIIT's ADA programme). ReKnow, Revolution of Knowledge Work, is one of the two first big strategic openings of Tekes started in 2013, and coordinated by HIIT. The partners are University of Helsinki, Aalto University and Institute for Occupational Health. Additionally in 2013, Professor Samuel Kaski was selected as a new group leader in Biocentrum Helsinki, which is an umbrella organization of selected top level research groups working on molecular biology, biomedicine, biotechnology and bioinformatics at University of Helsinki or Aalto University.



ERC Starting Grant to a HIIT group leader

Professor Petteri Kaski obtained an ERC starting grant for his work on "Theory and Practice of Advanced Search and Enumeration (TAPEASE)". In 2013 HIIT started systematic coaching of potential ERC grant applicants among HIIT's postdoctoral faculty to increase their success rate.



Nominations

HIIT's postdoctoral researchers Travis Gagie, Eve Hoggan and Leena Salmela got an Academy Postdoctoral Researcher position of the Academy of Finland, and Dr Arto Klami was appointed Academy Research Fellow. Dr Jukka Suomela was appointed as Assistant Professor at Aalto University, and Dr Teemu Roos and Dr Mikko Koivisto were appointed as Assistant Professors at University of Helsinki. Dr Jaakko Peltonen was nominated as Associate Professor at University of Tampere. HIIT's Research Coordinator, Dr Ella Bingham, was appointed as a part-time Executive Director of the Research Foundation of TKK which supports the research and teaching of technical disciplines at Aalto University.

HIIT-Wide Focus Area



As of 2011, HIIT has begun focusing research on a HIIT-wide focus area which builds on the existing excellences, touches most of HIIT research topics, and is expected to have direct impact in several other sciences and, through use of the developed methods, in industry and other parts of society. In the focus area "Augmented Science", IT methods and pilot tools are developed for making the scientific process more cumulative which is expected to transform in particular data-driven fields such as modern biology. Improving the general problem solving method of science, in collaboration with the other fields, is perhaps the best way for HIIT to contribute to solving the grand challenges of the humanity. The work focuses on two case studies: general augmented research, and computational cumulative biology. The focus area is now in full speed and several big joint projects have started: Tekes big strategic opening "ReKnow", EU FP7 funded "MindSee" project and Academy of Finland's "Multivire" project.

Helsinki Distinguished Lecture Series on Future Information Technology



HIIT's lecture series continues to attract world-class speakers and a wide audience. Speakers during 2013 were Eric W. Brown (IBM Research), Leslie Valiant (Harvard), Christos Papadimitriou (UC Berkeley), Henry Tirri (Nokia), and Klaus-Robert Müller (TU Berlin). The first speaker of 2014 was Corinna Cortes, Head of Google Research, New York.



A new research programme Distributed and Mobile Computing Systems

The Scalable Networks programme was refocused in early 2013, with a new name Distributed and Mobile Computing Systems (DMC) and a new director, professor Keijo Heljanko. The focus of the new programme is to bridge the gap between mobile devices and the cloud based server backend systems into a single seamless distributed and mobile computing platform.

A new research group Data Mining

Professor Aristides Gionis, a new Associate Professor at Aalto University, joined HIIT in the beginning of 2013 with his group Data Mining. Professor Gionis also started as the Director of the ADA research programme. His research focuses on data mining and algorithmic data analysis, with a particular interest in algorithms for graphs, social-network analysis, and algorithms for web-scale data. Before joining Aalto University, Gionis was a senior research scientist in Yahoo! Research.



Evaluation of HIIT groups

HIIT conducted the first internal evaluation of its groups in the spring 2013, to get an overview of the groups' activities for strategic planning, and to give feedback to the groups. The results of the evaluation were taken into account when allocating HIIT's internal resources. The main KPI's were publication activity and quality, degrees supervised, competitive funding earned, and societal impact.

Collaboration and impact

HIIT collaborates actively both within Helsinki area and internationally, and acts as a link between researchers, companies and public administration.

Collaboration in Helsinki IT

HIIT coordinates common activities of computer science research at Helsinki area in several ways: doctoral education, recruitments, and joint initiatives.

In doctoral education, HIIT coordinates the Helsinki ICT doctoral training network HICT that continues the successful tradition of collaboration in Helsinki area. Altogether 57 supervisors and 210 students belong to the network that operates at 4 departments at Aalto University and one department at University of Helsinki.

Academy of Finland's doctoral programmes are gradually coming to an end, three of which were coordinated at HIIT or led by HIIT professors: Finnish Doctoral Programme in Computational Science (FICS), Future Internet Graduate School (FIGS) and Helsinki Doctoral Programme in Computer Science – Advanced Computing and Intelligent Systems (Hecse).



HIIT also collaborates with Institute for Molecular Medicine Finland FIMM in recruiting PhD students to rotate between research groups at FIMM and HIIT. The collaboration is part of the Nordic EMBL Partnership for Molecular Medicine and the call is able to attract a large number of excellent doctoral candidates from all over the world.

Recruitment is an important means of collaboration. HIIT arranges joint Helsinki IT postdoc and senior researcher calls together with several departments in the field of computer science at Aalto University and University of Helsinki. These joint calls leverage on and further enhance the brand of Helsinki as a hub of computer science, and are able to attract a wide range of high-level applicants.

Digi Platform is Aalto University's new collaborative initiative in the field of digitalization. In the Digi platform, expertise across Aalto is brought together to maximize the potential of IT to boost other sciences, technologies and society. Currently in Aalto the field is very large but scattered, and the platform aims to coordinate and raise the profile of Aalto. University of Helsinki participates in the platform via HIIT, and HIIT is actively building up the platform.

HIIT is active in the ICT SHOK Digile (Strategic Centre for Science, Technology and Innovation), shaping the collaboration between universities and companies in Finland. HIIT's researchers are coordinating Digile's research programmes Next Media, Internet of Things, and Data to Intelligence D2I. HIIT researchers participate also in the Health field SHOK SaWe.

The universities and Ministry of Education are more and more using the quality of publications as a measure of success. The Publication Forum (Julkaisufoorumi) is a collective effort of the scientific community in Finland to rank publication venues into 4 levels. In the spring 2013 HIIT made a coordinated effort to suggest level changes and new forums in the field of computer science at Julkaisufoorumi. A group of 32 experts at HIIT and at neighboring departments at Aalto and University of Helsinki participated in this effort which was welcomed by the Publication Forum.

International collaboration

All HIIT's activities are inherently international, mostly in a bottom-up manner. Cooperation with key international research institutes and universities is active: Berkeley (ICSI and UC Berkeley), MIT, Centre for Computational Statistics and Machine Learning (CSML) at UCL, European Bioinformatics Institute EBI, and Human Technology Lab (HTLab) at University of Padova, to name a few.

EIT ICT Labs is an European initiative intended to turn Europe into a global leader in ICT innovation. It aims to fulfill this mission by establishing a new type of partnership between leading companies, research centres, and universities in Europe. One of the 5 nodes is Helsinki, and Professor



Marko Turpeinen of HIIT is Helsinki Node Director. Several HIIT researchers have research projects with EIT ICT Labs and are thus contributors to the establishment and ramp up of the EIT ICT Labs activities in Helsinki.

HIIT's researchers are active in organizing high-profile international conferences and workshops: 16th International Conference on Theory and Applications of Satisfiability Testing (SAT 2013), 6th ACM International Conference on Web Search and Data Mining (WSDM 2013), and 6th Workshop on Information Theoretic Methods in Science and Engineering (WITMSE 2013).



WITMSE 2013

Societal impact

HIIT's alumni establish approximately one startup company per year, one of the newest ones being Niko Vuokko's Metrify.io.

HIIT aims to boost the competitiveness of the Finnish information industry and information society in the long run by linking scientific research with the long-term high-risk research and development of the information industry, and by contributing to the development of the information society. HIIT both anticipates developments of the information society and influences them. New technologies developed at HIIT are taken into use at the ICT sector or other fields in the commercial and public sector.



HIIT researchers have been active in issues related to public policy having the profile of a vocal and visible participant in the public debate related to various important themes of information society, especially the discussion on emerging social forms of media use, immaterial rights, privacy and trust in the network society.

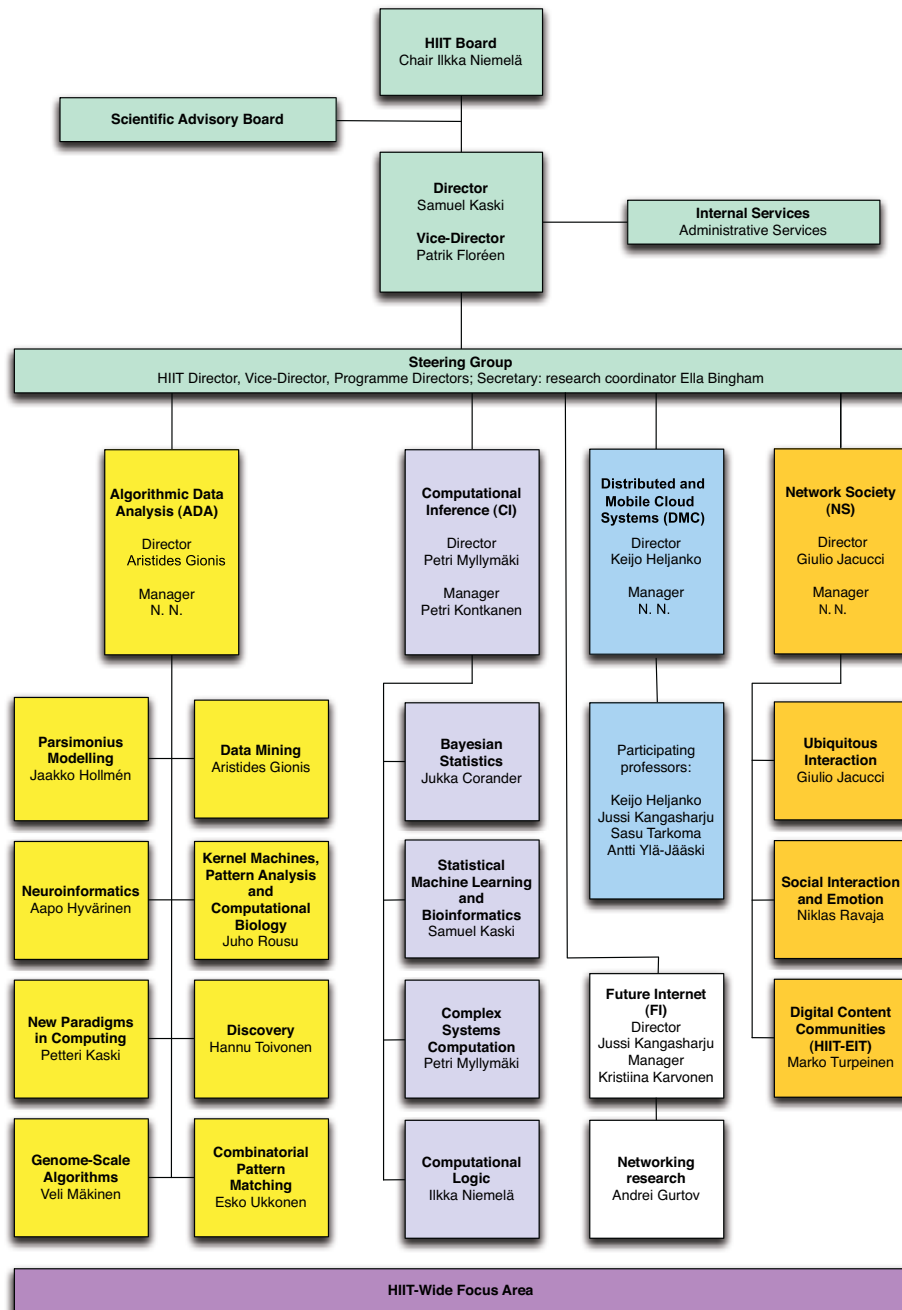
Views for 2014

In 2014 HIIT researchers will participate in organizing main international conferences such as 17th International Conference on Artificial Intelligence and Statistics (AISTATS 2014), IEEE International Conference on Data Mining (ICDM 2014), 13th International Symposium on Intelligent Data Analysis (IDA 2014), and Discovery Science (DS 2014).



HIIT will have a new Board from April 2014 onwards, as described in more detail in Chapter 3 of this book. HIIT expects to strengthen its role as a collaboration body, for example in the Digi Platform of Aalto University, also linking University of Helsinki to the platform.

Helsinki Institute for Information Technology



Algorithmic Data Analysis (ADA)

The mission of the Algorithmic Data Analysis research programme at HIIT is to develop useful algorithmic data analysis methods for other sciences and for industry. The work involves both basic research in computer science and applied work on problems arising from applications. The Finnish Center of Excellence in Algorithmic Data Analysis Research (Algodan, 2008-2013) formed the majority of HIIT's Algorithmic Data Analysis programme in 2013. The research of the Algodan CoE lies in the algorithmic and modeling problems of combinatorial pattern matching, data mining, and machine learning. The work in Algodan is strongly interdisciplinary: we cooperate constantly with application experts in various application areas, formulating novel computational concepts and ways of attacking the scientific and industrial problems of the application areas. Developing new concepts and algorithms is an iterative process consisting of interacting extensively with the application experts, formulating computational concepts, analyzing the properties of the concepts, designing algorithms and analyzing their performance, implementing and experimenting with the algorithms, and applying the results in practice. The main application areas of the Algodan CoE are in biology, medicine, telecommunications, environmental studies, linguistics, and neuroscience.

ALGODAN
Algorithmic Data Analysis



Combinatorial Pattern Matching, *Professor Esko Ukkonen*

The combinatorial pattern-matching group develops combinatorial algorithms for pattern search and synthesis problems for sequential and higher-dimensional data. The team is interested in the basic research of the theoretical aspects of the area as well as in various applications such as genome structure, metabolic modelling, gene regulation, and information retrieval.

Key publications:

- Emanuele Giaquinta, Szymon Grabowski and Esko Ukkonen. Fast Matching of Transcription Factor Motifs Using Generalized Position Weight Matrix Models. *Journal of Computational Biology* 20(9):621–630, 2013.
- Emanuele Giaquinta, Kimmo Fredriksson, Szymon Grabowski, Alexandru I. Tomescu and Esko Ukkonen. Motif matching using gapped patterns. *CoRR abs/1306.2483*, 2013.
- Arttu Jolma, Jian Yan, Thomas Whittington, Jarkko Toivonen, Kazuhiro R. Nitta, Pasi Rastas, Ekaterina Morgunova, Martin Enge, Mikko Taipale, Gonghong Wei, Kimmo Palin, Juan M. Vaquerizas, Renaud Vincentelli, Nicholas M. Luscombe, Timothy R. Hughes, Patrick Lemaire, Esko Ukkonen, Teemu Kivioja, and Jussi Taipale. DNA-Binding Specificities of Human Transcription Factors. *Cell* 152(1–2): 327–339, 2013.
- Emanuele Giaquinta, Kimmo Fredriksson, Szymon Grabowski, and Esko Ukkonen. Motif matching using gapped patterns. In *Combinatorial Algorithms – 24th International Workshop (IWOCA 2013)*, volume 8288 of *Lecture Notes in Computer Science*, pages 448–452. Springer, 2013.

Data Mining, *Professor Aristides Gionis*

The data-mining group focuses on developing novel methods to extract knowledge from data, designing algorithms to summarize large volumes of data efficiently and effectively, and exploring new ways of using the extracted information. Specific areas of interest include: pattern discovery, clustering and outlier detection, graph mining, social-network analysis, analysis of information networks and social-network dynamics, analysis of smart-city sensor data.

Key publications:

- Aristides Gionis, Flavio Junqueira, Vincent Leroy, Marco Serafini, and Ingmar Weber. Piggybacking on social networks. *Proceedings of the VLDB Endowment (PVLDB)*, 6(6), pages 409–420, 2013.
- Charalampos Tsourakakis, Francesco Bonchi, Aristides Gionis, Francesco Gullo, and Maria Tsiarli. Denser than the densest subgraph: Extracting optimal quasi-cliques with quality guarantees. *The 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2013)*, pages 104–112, 2013.
- Gianmarco De Francisci Morales, Francesco Bonchi, Aristides Gionis, and Antti Ukkonen. Activity preserving graph simplification. *Data Mining and Knowledge Discovery*, 27(3), pages 321–343, 2013.
- Aristides Gionis, Evimaria Terzi, and Panayiotis Tsaparas. Opinion maximization in social networks. *Proceedings of the 13th SIAM International Conference on Data Mining (SDM)*, pages 387–395, 2013.
- Mahashweta Das, Gianmarco De Francisci Morales, Aristides Gionis, and Ingmar Weber. Learning to question: leveraging user preferences for shopping advice. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2013)*, pages 203–211, August 2013.
- Konstantin Kutzkov, Albert Bifet, Francesco Bonchi, and Aristides Gionis. Strip: stream learning of influence probabilities. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 275–283, August 2013.



Discovery Group: Data Mining and Computational Creativity, Professor Hannu Toivonen



The Discovery group develops novel methods and tools for data mining and computational creativity. Our focus is on algorithmic methods for discovering links and patterns in data, and recently also on their use in creative systems. Computational creativity is interesting on its own right but also as an application area for data mining methods. We work on verbal creativity, i.e., computational poetry and humour, and also on automatic composition of music. We are developing novel methods that minimize the need for manually coded or language-specific knowledge.

Key publications:

- Laura Langohr, Vid Podpecan, Marko Petek, Igor Mozetic, Kristina Gruden, Nada Lavrac, Hannu Toivonen. Contrasting subgroup discovery. *The Computer Journal* 56 (3): 289-303. 2013.
- Jiuyong Li, Jixue Liu, Hannu Toivonen, and Jianming Yong. Effective pruning for the discovery of conditional functional dependencies. *Computer Journal*, 56(3):378–392, 2013.
- Juhani Huovelin, Oskar Gross, Otto Solin, Krister Linden, Sami Petri Tapio Maisala, Tero Oittinen, Hannu Toivonen, Jyrki Niemi, and Miikka Silfverberg. Software newsroom – an approach to automation of news search and editing. *Journal of Print Media Technology research*, 2(3):141–156, 2013.
- Jukka Toivanen, Matti Järvisalo, Hannu Toivonen. Harnessing constraint programming for poetry composition. In *The Fourth International Conference on Computational Creativity (ICCC)*, Sydney, Australia, June 2013.
- Alessandro Valitutti, Antoine Doucet, Jukka Toivanen, and Hannu Toivonen. "Let everything turn well in your wife": Generation of adult humor using lexical constraints. In *The 51st Annual Meeting of the Association for Computational Linguistics (ACL)*, Volume 2: Short Papers, pages 243–248, 2013.
- Jukka Toivanen, Hannu Toivonen, and Alessandro Valitutti. Automatical composition of lyrical songs. In *The Fourth International Conference on Computational Creativity, (ICCC)*, Sydney, Australia, June 2013.
- Oskar Gross, Antoine Doucet and Hannu Toivonen. Named Entity Filtering Based on Concept Association Graphs. *Research in Computing Science* 70: 33–43, 2013.
- Hannu Toivonen, Oskar Gross, Jukka M. Toivanen, and Alessandro Valitutti. On creative uses of word associations. In *Synergies of Soft Computing and Statistics for Intelligent Data Analysis, Part 1, Advances in Intelligent Systems and Computing*, pages 17–24. Springer, 2013.



Neuroinformatics, Professor Aapo Hyvärinen

Neuroinformatics is widely defined as the cross-fertilization of information-processing and mathematical sciences on the one hand, and neural and cognitive sciences on the other.

Our group works on different aspects of neuroinformatics related to machine learning. We model the visual system in the brain by analyzing the statistical structure of the natural input images. We apply machine learning models on neuroimaging data, in particular MEG. We also develop the relevant theory of statistical machine learning, typically unsupervised.



Key publications:

- Cristina Campi, Lauri Parkkonen, Riitta Hari, and Aapo Hyvärinen. Non-linear canonical correlation for joint analysis of MEG signals from two subjects. *Frontiers in Neuroscience*, 7(107), 2013.
- Michael U. Gutmann and Aapo Hyvärinen. A three-layer model of natural image statistics. *Journal of Physiology (Paris)*, 107(5):369–398, 2013.
- Antti Hyttinen, Patrik Hoyer, and Frederick Eberhardt. Experiment selection for causal discovery. *Journal of Machine Learning Research*, 2013(14):3041–3071, 2013.
- Aapo Hyvärinen. Independent component analysis: recent advances. *Philosophical transactions – Royal Society. Mathematical, Physical and engineering sciences*, 371(1984), 2013.
- Aapo Hyvärinen and Pavan Ramkumar. Testing independent component patterns by inter-subject or inter-session consistency. *Frontiers in Human Neuroscience*, 7, 2013.
- Aapo Hyvärinen and Stephen M. Smith. Pairwise likelihood ratios for estimation of non-Gaussian structural equation models. *Journal of Machine Learning Research*, 14:111–152, 2013.
- Jukka-Pekka Kauppi, Lauri Parkkonen, Riitta Hari, and Aapo Hyvärinen. Decoding magnetoencephalographic rhythmic activity using spectrospatial information. *NeuroImage*, 83:921–936, 2013.
- Hiroaki Sasaki, Michael U. Gutmann, Hayaru Shouno, and Aapo Hyvärinen. Correlated topographic analysis. *Machine Learning*, 92(2-3):285–317, 2013.
- Doris Entner, Patrik Hoyer, and Peter Spirtes. Data-driven covariate selection for non-parametric estimation of causal effects. In *Proceedings of the 16th International Conference on Artificial Intelligence and Statistics (AISTATS'13)*, pages 256–264, 2013.
- Antti Hyttinen, Patrik Hoyer, Frederick Ederhardt, and Matti Järvisalo. Discovering cyclic causal models with latent variables. In *Proceedings of the Twenty-Ninth Conference on Uncertainty in Artificial Intelligence (UAI'13)*, pages 301–310, 2013.

Parsimonious Modeling,

Dr. Jaakko Hollmén



The research group Parsimonious Modelling develops novel computational data analysis methods and applies these methods on two application fields: cancer genomics and environmental informatics. Parsimonious modeling aims at simple, compact, or sparse models as a result of learning from data in the presence of very little or no a priori information about the modeled problem. Simplicity of the models facilitates understanding of the problem domain by humans.

Both application fields present similar challenges to the data analysis problems: the high dimensionality of observed data and the presence of moderate or large noise levels are both factors that bear fundamental problems for any data analysis. Seeking new areas of application and interfacing the newest application domains with lots of novel types of generated data helps in finding new, unsolved settings of problems.

Key publications:

- Prem Raj Adhikari and Jaakko Hollmén. Mixture models from multiresolution 0-1 data. In Johannes Fürnkranz, Eyke Hüllermeier, and Tomoyuki Higuchi, editors, Proceedings of Sixteenth International Conference on Discovery Science (DS 2013), pages 1–16, Berlin Heidelberg, 2013. Springer.
- Albert Bifet, Jesse Read, Bernhard Pfahringer, Geoff Holmes, and Indre Zliobaite. CD-MOA: Change detection framework for massive online analysis. In Proceedings of the 20th International Symposium on Intelligent Data Analysis, pages 92–103, October 2013.
- Albert Bifet, Jesse Read, Indre Zliobaite, Bernhard Pfahringer, and Geoff Holmes. Pitfalls in benchmarking data stream classification and how to avoid them. In Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECMLPKDD), pages 465–479, September 2013.
- Dino Ienco, Albert Bifet, Indre Zliobaite, and Bernhard Pfahringer. Clustering based active learning for evolving data streams. In Proceedings of the 16th International Conference on Discovery Science, pages 79–93, October 2013.
- Indre Zliobaite and Jaakko Hollmén. Fault tolerant regression for sensor data. In Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECMLPKDD), pages 449–464, 2013.

Kernel Machines, Pattern Analysis and Computational Biology,

Professor Juho Rousu

The group develops machine learning methods, models and tools for computational sciences, in particular computational biology. The group focuses on kernel methods and regularized learning for multiple and structured targets, multiple views and ensembles. Applications of interest include network reconstruction and labeling, gene functional classification, mass spectrometry informatics and biomarker discovery (Rousu et al. 2013). In mass spectrometry informatics, with the support of Academy of Finland research grant (MIDAS, 2013-2017), we focus on metabolite identification from tandem mass spectrometric data, an important bottleneck problem in metabolomics. In our most recent approach, we combine molecular fragmentation models with multiple kernel learning, thus achieving state-of-the-art predictive performance (Shen et al. 2014). In machine learning, we have developed new methods for multilabel classification, based on ensemble learning on a collection of random output graphs imposed on the multilabels, and a kernel-based structured output learner as the base classifier (Su and Rousu, 2013). This work is continued in collaboration with University College London and Laval University, Canada. Another example of our recent work is structured prediction for predicting subgraphs of a network responding to a stimulus (Su et al. 2014).



Key publications:

- Juho Rousu, Daniel D. Agranoff, Olugbemiro Sodeinde, John Shawe-Taylor, and Delmíro Fernández-Reyes. Biomarker discovery by sparse canonical correlation analysis of complex clinical phenotypes of tuberculosis and malaria. *PLOS Computational Biology*, 9(4):1003018, 2013.
- Hongyu Su and Juho Rousu. Multilabel classification through random graph ensembles. In *Proceedings, 5th Asian Conference on Machine Learning (ACML2013)*, pages 404–418, November 2013.



Genome-Scale Algorithms,

Professor Veli Mäkinen



We develop algorithms and data structures for the analysis of genome-scale data. Such data is abundant due to modern molecular biology measurement techniques like high-throughput sequencing. We are especially interested in applications of compressed data structures, that make it possible to analyse the often highly redundant data within the space of their information content. We also study other scalability aspects like distributed computation/storage around genome-scale data.

An example of our recent developments is an extension of Burrows-Wheeler transform to finite automaton representing reference genome together with its common variations among the population. This enables a space-efficient index structure to be constructed to support efficient read alignment to a rich model of the population. Finite automaton representation enables good control over the richness of the model as one can e.g. create paths representing different haplotype blocks, a property not easy to handle using e.g. HMM-based aligners.

Key publications:

- Veli Mäkinen and Jani Rahkola. Haploid to diploid alignment for variation calling assessment. *BMC Bioinformatics*, 14((Suppl 15):S13), 2013.
- Alexandru I. Tomescu, Anna Kuosmanen, Romeo Rizzi, and Veli Mäkinen. A novel min-cost flow method for estimating transcript expression with RNA-Seq. *BMC Bioinformatics*, 14((Suppl 5):S15), 2013.
- Djamal Belazzougui, Fabio Cunial, Juha Kärkkäinen, and Veli Mäkinen. Versatile succinct representations of the bidirectional Burrows-Wheeler transform. In *Proc. ESA 2013*, Springer, LNCS 8125, pages 133–144, 2013.
- Travis Gagie, Danny Hermelin, Gad M. Landau and Oren Weimann. Binary Jumbled Pattern Matching on Trees and Tree-Like Structures. In *Proc. ESA 2013*, Springer, LNCS 8125, pages 517–528, 2013.
- Travis Gagie, Kalle Karhu, Gonzalo Navarro, Simon Puglisi, and Jouni Sirén. Document Listing in Repetitive Collections. In *Proc. CPM 2013*. Springer, LNCS 7922, pages 107–119, 2013.
- Alexandru I. Tomescu, Anna Kuosmanen, Romeo Rizzi and Veli Mäkinen. A Novel Combinatorial Method for Estimating Transcript Expression with RNA-Seq: Bounding the Number of Paths. In *Proc. WABI 2013*, Springer, LNCS 8126, pages 85–98, 2013.

New Paradigms in Computing,

Professor Petteri Kaski

The group performs basic research at the intersection of core computer science (algorithm design and analysis) and discrete mathematics, with an emphasis towards novel techniques and less studied models of computation. We invest substantial effort to high-risk, high-yield research problems of relatively broad theoretical interest, selected on both problem and method driven basis. However, we also aim at rapid publication of more specific, smaller observations. We particularly seek and value solid results with mathematical elegance and simplicity.

The current research themes of the group are threefold. (1) Exploring the interplay between algebraic, combinatorial, and geometric techniques in the design of exact deterministic algorithms. For example, many combinatorial problems can be cast in algebraic form, whereby a nontrivial algebraic algorithm yields a more efficient solution compared with direct combinatorial tools. (2) Restricted models of computation and tradeoffs in resources and/or objectives. For example, one fundamental limitation in modern large-scale distributed systems is the infeasibility of central control. In practice, the system must be operated by a distributed algorithm in which each computational node operates based on the information available in its local neighbourhood only. Assuming this setting, is it possible to achieve globally optimal or near-optimal operation? What is the tradeoff between the available information and the degree of approximation for the optimum? (3) While we are a theory group, we occasionally engage in practical algorithm implementation. Examples include attacks on combinatorial classification problems and applications in computational geometry (e.g. air traffic management).

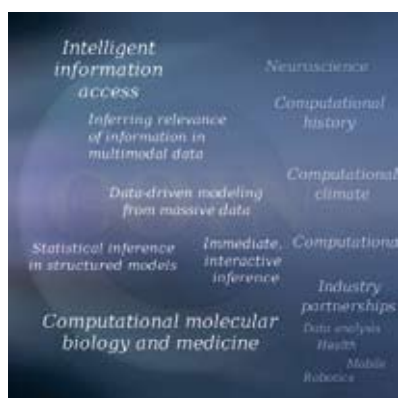


Key publications:

- Andreas Björklund and Petteri Kaski. Counting closed trails. *Information Processing Letters*, 113(1–2):1–3, 2013.
- Fedor V. Fomin and Petteri Kaski. Exact exponential algorithms: Surprises in the face of intractability. *Communications of the ACM*, 56(3):80–88, 2013.
- Mika Göös, Juho Hirvonen, and Jukka Suomela. Lower bounds for local approximation. *Journal of the ACM*, 60(5), 2013.
- Pekka Parviainen and Mikko Koivisto. Finding optimal Bayesian networks using precedence constraints. *Journal of Machine Learning Research*, 14:1387–1415, 2013.
- Jukka Suomela. Survey of local algorithms. *ACM Computing Surveys*, 45(2):24, 2013.
- Pierre Fraigniaud, Mika Göös, Amos Korman, and Jukka Suomela. What can be decided locally without identifiers? In *Proc. PODC'13*, pages 157–165, 2013.
- Janne H. Korhonen and Pekka Parviainen. Exact learning of bounded tree-width Bayesian networks. In *Proc. AISTATS 2013*, pages 370–378, 2013.
- Teppo Niinimäki and Mikko Koivisto. Treedy: A heuristic for counting and sampling subset. In *Proc. UAI-13*, pages 469–477, 2013.



Computational Inference (CI)



The four groups of the CI programme are all members of the Finnish Centre of Excellence in Computational Inference Research (COIN), and the objectives of the programme are closely intertwined with those of COIN.

The main objective of CI is to develop methods for transforming the data produced by the current data revolution into useful information. The key methodology for achieving this goal is statistical and computational inference based on the data. The emphasis is on large data collections and computationally demanding modelling and inference algorithms. Our mission is to push the boundary towards both more complex problems, requiring more structured data models, and towards extremely rapid inference. We plan to address a set of carefully chosen interdisciplinary "grand challenge" -level problems with high societal impact where solving the data intensive problems requires novel methodologies that can only result from combining the expertise of separate subfields.

Our mission brings up four partially overlapping methodological focus areas: Learning of massive data-driven models; Learning from multiple sources; Statistical inference in highly structured stochastic models; and Extreme inference engine.

In applied research our work is also motivated by the big data and ubiquitous computing vision, where adaptivity, context-awareness and personalisation are key enablers. We see that our four methodological research areas support strongly each other, and they all address from a different perspective the key technological problems we face in our future "big data" information society. In our applied research we wish to link our strong basic research work in machine learning and constraint reasoning to well-motivated applied research activities involving prototype applications and real-world deployments.

Statistical Machine Learning and Bioinformatics,

Professor Samuel Kaski

We develop new methods for machine learning, computational inference, and probabilistic modeling. We focus on models for learning from multiple data sources, including multi-view learning, multi-task learning, and multi-way learning, and methods combining mechanistic models and probabilistic inference. Our primary application areas are computational systems biology and medicine, bioinformatics, proactive information retrieval and multimodal interfaces, as well as brain signal analysis and neuroinformatics. Our main lines of work include Bayesian methodology for modeling dependencies between multiple data sets with co-occurring samples, model-driven methods for retrieving and visualizing data sets, and Gaussian process differential equation models and inference methods for short genomic time series data.



Key publications:

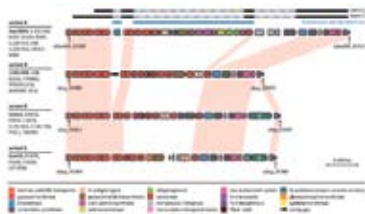
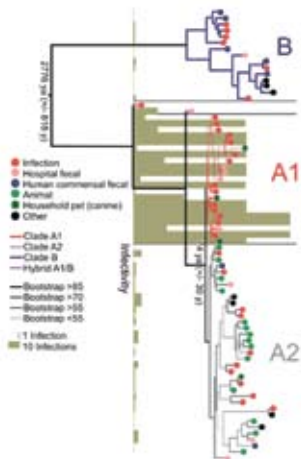
- Arto Klami, Seppo Virtanen, and Samuel Kaski. Bayesian canonical correlation analysis. *Journal of Machine Learning Research*, 14:965–1003, 2013.
- Miika Koskinen, Jaakko Viinikanoja, Mikko Kurimo, Arto Klami, Samuel Kaski, and Riitta Hari. Identifying fragments of natural speech from the listener's MEG signals. *Human Brain Mapping*, 34(6):1477–1489, 2013.
- Pekka Marttinen, Jussi Gillberg, Aki Havulinna, Jukka Corander, and Samuel Kaski. Genome-wide association studies with high-dimensional phenotypes. *Statistical Applications in Genetics and Molecular Biology*, 12(4):413–431, 2013.
- Mehmet Gönen, Suleiman A. Khan, and Samuel Kaski. Kernelized Bayesian matrix factorization. In *Proc. ICML, 2013*, vol 28 of *JMLR W&CP*, pages 864–872. *JMLR*, 2013.
- Antti Honkela, Magnus Rattray, and Neil D. Lawrence. Mining regulatory network connections by ranking transcription factor target genes using time series expression data, pages 59–67. *Methods in Molecular Biology*. 2013.
- Tuukka Ruotsalo, Jaakko Peltonen, Manuel Eugster, Dorota Glowacka, Ksenia Konyushkova, Kumaripaba Athukorala, Ilkka Kosunen, Aki Reijonen, Petri Myllymäki, Giulio Jacucci, and Samuel Kaski. Directing exploratory search with interactive intent modeling. In *Proc. CIKM*, pages 1759–1764, *ACM*, 2013.

Bayesian Statistics, Professor Jukka Corander



We do research on theoretical and applied machine learning, biometry, bioinformatics and forensic statistics. A specific area of interest is statistical inference in highly structured stochastic models. Our main application area is computational biology, in particular models for evolution and transmission of bacteria and viruses. We have developed a novel type of adaptive sequential Monte Carlo method for dynamic models with intractable likelihoods, which uses population information in the adaptation phase (Numminen et al. 2013). Our recent research highlights have shown the origins and course of horizontal evolution for multiresistant bacteria causing nosocomial and extraintestinal persistent infections (McNally et al 2013, Lebreton et al. 2013, de Been et al. 2013). These findings have been enabled by our better scalable Bayesian inference methods for the analysis of bacterial whole-genome data that are several orders of magnitude faster than recent other methods based on standard Bayesian computation.

In the forensics context, we introduced a completely novel predictive principle for comparing oil sample spectra (Blomstedt et al. 2014), which is shown to yield substantially more robust and legally sound results than either the current practice of using frequentist tests or Bayesian standard tests, the latter of which were shown to be highly unreliable. The method is based on the general predictive comparison theory developed in Blomstedt and Corander (2013) and it is expected to become the standard for forensic comparisons in future. For further details and complete publication list see www.helsinki.fi/bsg.



Key publications:

- Elina Numminen, Lu Cheng, Mats Gyllenberg, and Jukka Corander. Estimating the transmission dynamics of *Streptococcus pneumoniae* from strain prevalence data. *Biometrics*, 69(3):748–757, 2013.
- Jukka Corander, Tomi Janhunen, Jussi Rintanen, Henrik Nyman, and Johan Pensar. Learning chordal Markov networks by constraint satisfaction. In *Advances in Neural Information Processing Systems*, Volume 26, page 1357, 2013.
- François Lebreton et al. Emergence of epidemic multi-drug resistant *Enterococcus faecium* from animal and commensal strains. *mBio*, 4(4), doi: 10.1128/mBio.00534–13, 2013.
- Mark de Been, Willem van Schaik, Lu Cheng, Jukka Corander, and Rob J. Willems. Recent recombination events in the core genome are associated with adaptive evolution in *Enterococcus faecium*. *Genome Biology and Evolution*, doi: 10.1093/gbe/evt111, 2013.
- Alan McNally, Lu Cheng, Simon R Harris, and Jukka Corander. The evolutionary path to extra intestinal pathogenic, drug resistant *Escherichia coli* is marked by drastic reduction in detectable recombination within the core genome. *Genome Biology and Evolution* 5(4), 699–710, 2013.
- Paul Blomstedt, Romain Gauriot, Niina Viitala, Tapani Reinikainen, and Jukka Corander Bayesian predictive modeling and comparison of oil samples. *Journal of Chemometrics* 28(1), 52–59, 2014.

Complex Systems Computation (CoSCo),

Professor Petri Myllymäki

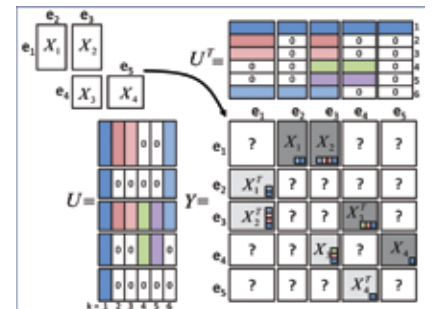
The Complex Systems Computation Research Group investigates computational modeling issues in complex systems, and the related implementation aspects, focusing on prediction and model selection tasks. The work has both a strong basic research component, being at the intersection of computer science, information theory and mathematical statistics, and an applied component where the results are used for solving problems in various disciplines from social sciences, criminology, ecology and medicine to industrial engineering.

The research areas addressed include probabilistic modeling and data analysis, information theoretical approaches to inference, constraint reasoning and ubiquitous computing. Central themes in the applied research include information retrieval, context-awareness for ubiquitous computing, modeling for location-aware services and sensor networks.



Key publications

- Matti Järvisalo and Allen Van Gelder (editors). Theory and Applications of Satisfiability Testing – SAT 2013. Volume 7962 of Lecture Notes in Computer Science, Springer 2013.
- Marijn Heule, Matti Järvisalo, and Armin Biere. Covered clause elimination. In short papers for 17th International Conference on Logic for Programming, Artificial Intelligence and Reasoning, pages 41–46, 2013.
- Antti Hyttinen, Patrik Hoyer, Frederick Eberhardt, and Matti Järvisalo. Discovering cyclic causal models with latent variables: A general SAT-based procedure. In Proc. 29th Conference on Uncertainty in Artificial Intelligence (UAI 2013), pages 301–310. AUAI Press, 2013.
- Brandon Malone and Changhe Yuan. Evaluating anytime algorithms for learning optimal Bayesian networks. In Proceedings of the 29th Conference on Uncertainty in Artificial Intelligence (UAI 2013). AUAI Press, 2013.
- Changhe Yuan and Brandon Malone. Learning optimal Bayesian networks: A shortest path perspective. Journal of Artificial Intelligence Research, 48:23–65, 2013.
- Antti Oulasvirta, Teemu Roos, Arttu Modig, and Laura Leppänen. Information capacity of full-body movements. In Proc. SIGCHI Conference on Human Factors in Computing Systems, pages 1289–1298, 2013.
- Arto Klami. Bayesian object matching. Machine Learning 92(2):225–250, 2013.
- Sourav Bhattacharya, Santi Phithakkitnukoon, Petteri Nurmi, Arto Klami, Marco Veloso, and Carlos Bento. Gaussian process-based predictive modeling for bus ridership. In Proc. 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication, pages 1189–1198, 2013.
- Tuukka Ruotsalo, Jaakko Peltonen, Manuel Eugster, Dorota Glowacka, Ksenia Konyushkova, Kumaripaba Athukorala, Ilkka Kosunen, Aki Reijonen, Petri Myllymäki, Giulio Jacucci, and Samuel Kaski. Directing exploratory search with interactive intent modeling. In Proceedings of the 22nd ACM International Conference on Information & Knowledge Management, pages 1759–1764, New York, NY, USA, 2013.



Computational Logic, *Professor Ilkka Niemelä*



The group develops automated reasoning techniques for solving challenging computational problems in engineering and science. The current focus is on developing efficient computational methods for solving large constraint satisfaction problems expressed, e.g., as Boolean satisfiability (SAT) or using rule-based constraints of answer-set programming (ASP). The group has a strong track record in research on verification and testing of automation systems and software, as well as applying formal methods in the analysis of distributed systems.

In 2013 we investigated the problem of learning Markov networks from data and showed how such networks can be described with constraints and optimized using solvers with optimization capabilities. Moreover, we published new techniques for normalizing cardinality rules used in ASP as well as translating ASP into SAT extended by bit vectors.

Our research on constraint-based planning and scheduling led to novel polynomial-time methods for finding upper bounds of lengths of transition sequences based on the decomposition of state-variable dependency graphs. We also investigated the impact of imperfect information in scheduling, demonstrating that quantified Boolean formulas (QBFs) are required to handle the most important forms of incompleteness arising during the execution of a schedule.

Moreover, we continued our work on the analysis of safety critical systems involving real-time constraints. In 2013 we developed a symbolic encoding method for a fragment of the specification logic MITL, enabling the verification of properties involving timing requirements, e.g., by using bounded model checking techniques.

Key publications:

- Jori Bomanson and Tomi Janhunen. Normalizing cardinality rules using merging and sorting constructions. In *Logic Programming and Nonmonotonic Reasoning*, pages 187–199. Springer, 2013.
- Jukka Corander, Tomi Janhunen, Jussi Rintanen, Henrik Nyman, and Johan Pensar. Learning chordal Markov networks by constraint satisfaction. In *Advances in Neural Information Processing Systems*, Volume 26, pages 1349–1357. 2013.
- Roland Kindermann, Tommi Junttila, and Ilkka Niemelä. Bounded model checking of an MITL fragment for timed automata. In *13th International Conference on Application of Concurrency to System Design (ACSD)*, pages 216–225. IEEE, 2013.
- Mai Nguyen, Tomi Janhunen, and Ilkka Niemelä. Translating answer-set programs into bit-vector logic. In *Applications of Declarative Programming and Knowledge Management*, pages 95–113. Springer, 2013.
- Jussi Rintanen. Scheduling with contingent resources and tasks. In *Proceedings of the International Conference on Automated Planning and Scheduling (ICAPS 2013)* pages 189–196. AAAI Press, June 2013.
- Jussi Rintanen and Charles Orgill Gretton. Computing upper bounds on lengths of transition sequences. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 2365–2372, 2013.

Distributed and Mobile Cloud Systems (DMC)

The focus of the new programme is to bridge the gap between mobile devices and the cloud based server backend systems into a single seamless distributed and mobile computing platform. The main motivation behind this is that mobile devices are by their very nature very resource constrained in available battery power, CPU, memory, network, as well as storage capacity compared to the server hardware available in the cloud backend systems. Thus mobile devices need to be tightly integrated to the cloud backend systems in order to do computational tasks that are too heavy for them. However, this basic setup is not yet sufficient for highly interactive applications. The wide area network (WAN) communication latencies between the mobile device and the possibly quite physically remote cloud backend can often be too large for interactive mobile applications, e.g., for interactive augmented reality applications such as Google Glass, as well as computationally intensive mobile intelligent information access applications. Therefore, sometimes an additional layer of computing called cloudlets is called for, that is a collection of local cloud servers that acts as local computing service for mobile devices in its own network neighborhood area. This idea is also known as cyber foraging in the literature.



Distributed Networking and Security, *Professor Antti Ylä-Jääski*

Distributed networking and security research group is working on two research areas, namely (a) energy aware computing and communications and (b) distributed systems and services.

Energy-aware computing and communications is a very timely and important topic. Energy consumption is a concern with today's mobile devices. While the capabilities of the devices have improved rapidly over the last ten years transforming also the way these devices are being utilized, battery technology has not been able to keep up with this evolution. As a consequence, there is an increasing gap between the battery capacity and the amount of energy required for typical usage. We address these challenges by building models of energy consumption through experiments and measurements based on which we develop new more energy-efficient protocols and services.

Distributed systems and services address architectures, platforms, and protocols for flexible, scalable, and easily usable services. Cloud computing uses virtual resources in the Internet for computing and storage, and is able to elastically scale to match changing resource needs. Merging cloud technologies with mobile domain has potential to offer new technology innovations and business opportunities for operators, vendors, and developers, as well as novel services for end users.

Key publications

- Yong Cui, Hongyi Wang, Xiuzhen Cheng, Dan Li, and Antti Ylä-Jääski. Dynamic scheduling for wireless data center networks. *IEEE Transactions on Parallel and Distributed Systems*, 2013.
- Yu Xiao, Yong Cui, Petri Savolainen, Matti Siekkinen, An Wang, Liu Yang, Antti Ylä-Jääski, and Sasu Tarkoma. Modeling energy consumption of data transmission over wi-fi. *IEEE Transactions on Mobile Computing*, 99(1):1–14, 2013.
- Ming Li, Andrey Lukyanenko, Sasu Tarkoma, Yong Cui, and Antti Ylä-Jääski. Tolerating path heterogeneity in multipath TCP with bounded receive buffers. In *Proceedings of the ACM SIGMETRICS, International Conference on Measurement and Modeling of Computer Systems*, pages 375–376, 2013.
- Ming Li, Andrey Lukyanenko, Antti Ylä-Jääski, and Sasu Tarkoma. The delayed ACK evolution in MPTCP. In *IEEE Global Communications Conference, GLOBECOM 2013*, 2013.
- Sumanta Saha, Andrey Lukyanenko, and Antti Ylä-Jääski. Cooperative caching through routing control in information-centric networks. In *The 32nd IEEE International Conference on Computer Communications (INFOCOM 2013)*, pages 100–104, Turin, Italy, April 2013. IEEE.

Distributed Computing, Professor Keijo Heljanko

The group has two main research topics: computer aided verification of distributed systems and distributed computing. On the computer aided verification side the group focuses on symbolic model checking techniques, dynamic symbolic execution based testing techniques, parallel and distributed verification methods, and techniques for the development and quality assurance of safety critical systems. On the distributed computing side the main focus is on cloud computing, the underlying distributed algorithms and technologies, as well as applications of cloud computing in key areas of science and engineering.



Key publications:

- Lauri Ahlroth, Olli Pottonen, and André Schumacher. Approximately uniform online checkpointing with bounded memory. *Algorithmica*, 67(2):234–246, 2013.
- Lauri Ahlroth, André Schumacher, and Pekka Orponen. Online bin packing with delay and holding costs. *Operations Research Letters*, 41(1):1–6, 2013.
- Tuomas Kuismin and Keijo Heljanko. Increasing confidence in liveness model checking results with proofs. In *The 9th International Haifa Verification Conference (HVC 2013)*, pages 32–43, November 2013.
- Antti Siirtola and Keijo Heljanko. Parametrised compositional verification with multiple process and data types. In J. Carmona, M. T. Lazarescu, and M. Pietkiewicz-Koutny, editors, *Proceedings, 13th International Conference on Application of Concurrency to System Design (ACSD)*, pages 67–76. IEEE, July 2013.
- Siert Wieringa and Keijo Heljanko. Asynchronous multi-core incremental SAT solving. In *Proceedings of the 19th International Conference on Tools and Algorithms for the Construction and Analysis of Systems, TACAS 2013*, pages 139–153, 2013.
- Siert Wieringa and Keijo Heljanko. Concurrent clause strengthening. In *Theory and Applications of Satisfiability Testing - SAT 2013*, pages 116–132, 2013.

Mobile Computing, Professor Sasu Tarkoma



The group investigates different aspects of wireless and mobile communications. The group has a strong focus on mobile middleware and service platforms. Recent research of the group pertains to mobile cloud computing, data and computation offloading, and energy modeling and optimization.

Our vision is that each layer of the networking stack must be aware of the environment, and the whole communication stack needs to be optimized and made adaptive. A key challenge is how to distribute service functionality and logic in the distributed environment. The aim of our work is to enable efficient, secure, always-on, and reliable connectivity irrespective of the access network and terminal device. Moreover, service access and usage must be personalized and adapted to the current operating context.

Key publications

- Ming Li, Andrey Lukyanenko, Sasu Tarkoma, Yong Cui, and Antti Ylä-Jääski. Tolerating Path Heterogeneity in Multipath TCP with Bounded Receive Buffers. *Computer Networks*, 64:1–14, February 2014.
- Weixiong Rao, Lei Chen, and Sasu Tarkoma. Toward efficient filter privacy-aware content-based pub/sub systems. *IEEE Transactions on Knowledge and Data Engineering*, 25(11):2644–2657, 2013.
- Sasu Tarkoma and Heikki Ailisto. The Internet of Things program: The Finnish perspective. *IEEE Communications Magazine*, 51(3):10–11, 2013.
- Yu Xiao, Yong Cui, Petri Savolainen, Matti Siekkinen, An Wang, Liu Yang, Antti Ylä-Jääski, and Sasu Tarkoma. Modeling energy consumption of data transmission over wi-fi. *IEEE Transactions on Mobile Computing*, 99(1):1–14, 2013.
- Ming Li, Andrey Lukyanenko, Sasu Tarkoma, and Antti Ylä-Jääski. Efficient new delayed ACK for TCP: old problem, new insight. In *Proceedings of the ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, pages 355–364, 2013.
- Ming Li, Andrey Lukyanenko, Antti Ylä-Jääski, and Sasu Tarkoma. The delayed ACK evolution in MPTCP. In *IEEE Global Communications Conference, GLOBECOM 2013*, 2013.

Collaborative Networking,

Professor Jussi Kangasharju

The Collaborative Networking (CoNe) group's research focuses on large-scale distributed systems and network applications which are based on nodes cooperating voluntarily. Examples of such systems are information-centric networks and mobile opportunistic networks. The work focuses on investigating architectures and mechanisms for designing, prototyping, and testing of future networks. The goal of the work is to understand how future networks should be designed and built. The group has also worked on data center energy efficiency and green networking.

Key publications:

- Sasitharan Balasubramaniam and Jussi Kangasharju. Realizing the Internet of Nano Things: Challenges, solutions, and applications, *Computer: a publication of the IEEE Computer Society*, 46(2): 62–68, 2013.
- Mikko Pervilä, Lassi Remes, and Jussi Kangasharju, Harvesting heat in an urban greenhouse. *SIGMETRICS Performance Evaluation Review*, 41(3): 95–97, ACM, December 2013.
- Liang Wang, Otto Waltari, Jussi Kangasharju, MobiCCN: Mobility Support with Greedy Routing in Content-Centric Networks, in *Proceedings of IEEE Globecom*, Atlanta, GA, USA, December 2013.
- Liang Wang, Jussi Kangasharju, Measuring Large-Scale Distributed Systems: Case of Bit-Torrent Mainline DHT, in *Proceedings of IEEE Conference on Peer-to-Peer Computing*, Trento, Italy, September 2013.
- Suzan Bayhan, Esa Hyytiä, Jussi Kangasharju, Jörg Ott, Seeker-Assisted Search in Mobile Clouds, in *Proceedings of ACM SIGCOMM Workshop on Mobile Cloud Computing*, Hong Kong, August 2013.
- Suzan Bayhan and Fatih Alagöz. Scheduling in centralized cognitive radio networks for energy efficiency. *IEEE Transactions on Vehicular Technology*, 62(2):582–595, 2013.



Networking Research, Professor Andrei Gurtov



We study a wide range of communication problems, starting from the medium access control, traffic analysis, multicast and IP mobility up to hierarchical peer-to-peer architectures and distributed algorithms. Our particular focus is security aspects and game-theoretic approach to model selfish user behavior in the modern Internet. We are actively involved in Internet Engineering and Research Task Force events contributing to creation of Internet standards. We are involved in projects on Internet-of-Things security and medical ICT, the locator/identifier split, web security and trust among Internet users, and smart spaces.

The group collaborates tightly with top-notch scientists from the International Computer Science Institute (ICSI) in Berkeley. The latest directions include fair bandwidth allocation in datacenter networks and resilient routing. In 2013, the group leader and two group members I. Nikolaevskiy and D. Kuptsov made a half a year visit to ICSI. Two group members, T. Polishchuk and B. Nechaev had successfully defended their PhD theses and two more dissertations were submitted for evaluation.

Key publications:

- Oscar Garcia-Morchon, Dmitriy Kuptsov, Andrei Gurtov, and Klaus Wehrle. Cooperative security in distributed networks. Elsevier Computer Communications (COMCOM), 36(12):1284–1297, July 2013.
- Huayang Cao, Peidong Zhu, Xicheng Lu, and Andrei Gurtov. A Layered Encryption Mechanism for Networked Critical Infrastructures. IEEE Network, 27(1): 12–18, January 2013.
- Dmitriy Kuptsov, Boris Nechaev, Andrei Gurtov, and Andrey Lukyanenko. A Novel Demand-Aware Fairness Metric for IEEE 802.11 Wireless Networks. In Proc. ACM SAC, pages 603–610, March 2013.
- Ilya Nikolaevskiy, Andrey Lukyanenko, Tatiana Polishchuk, Valentin Polishchuk, and Andrei Gurtov. isBF: Scalable In-Packet Bloom Filter Multicast. In Proc. ACM SAC, March 2013.
- Dmitriy Korzun and Andrei Gurtov. Hierarchical Architectures in Structured Peer-to-Peer Overlay Networks. Peer-to-Peer Networking and Applications, Springer, March 2013.

Network Society (NS)

The mission of the Network Society research programme is to empower ubiquitous users with transparent and resourceful ICT with the following overarching objectives:

- Multimodal interaction and adaptive information: resources are provided to users to interact with information ubiquitously. This includes studies of modalities such as manual ergonomics and skills, tactile interaction and cross modal rendering. Adaptation and situated computing is pursued to contribute to issues such as data deluge and persuasion.
- Emergent social media platforms and practices. The study of emergent forms of ICT including gaming, basic studies of practices and behavior in social media with special attention to knowledge work, privacy and self expression.
- Experience of Mediated Interaction. Emotional and cognitive processes during mediated social interaction and the neuroscience of social ICT.

While there are three research groups their research objectives contribute to a common agenda.

These three challenges are integrated to solve in an adjacent or combined way problems or anticipate future ICT. As examples new ubiquitous or multimodal interaction can contribute to novel forms of social media such as public and pervasive displays. The neuro-psychology of mediated interaction can contribute to understanding social media practices.

NS collaborates with other programmes, in particular in the HIIT wide focus area focusing on information exploration and knowledge work.



Digital Content Communities, *Professor Marko Turpeinen*



Our research focuses on social computing, i.e., information systems that enable and support social creativity, participatory media and distributed problem solving. However, to develop successful new technologies, and bear responsibility of design decisions, we as developers should understand and anticipate the dynamics of technology-society interaction. This requires multi disciplinary end-to-end research from technological platforms to various viewpoints to their impact on the use environment.

The goal of our research is to enable and encourage people to belong to communities of content creators, to develop service design methodology and business model development for social media, to study how people interact with each other using digital content, and to understand what new business opportunities and social structures emerge around community-generated content and technology. One of our strong focus areas is the rapidly expanding social forms of gaming and how to apply the findings from online games to other non-gaming domains. This includes virtual economies and virtual consumerism, and applying economics and business studies in virtual worlds.

Another viewpoint is legal. In relation to future technologies, it is not enough to be able to apply only traditional jurisprudential methods, but we must also collect and analyze empirical data. Therefore, in addition to conventional legal studies that focus on the analysis of statutes, court cases, etc, we also seek to understand practices and business reasons, as well as availing of user studies to find real human, business, and societal needs. Likewise, futures research methods, e.g. user scenarios, may give us important information on forthcoming issues.

Digital Content Communities is a joint research group of both HIIT and EIT ICT Labs Helsinki node.

Key publications

- Eva Heiskanen, Mikael Johnson, and Edina Vadovics. Learning about and involving users in energy saving on the local level. *Journal of cleaner production*, 48(June):249, 2013.
- Simo Järvelä, Matias Kivikangas, Timo Saari, and Niklas Ravaja. Media experience as a predictor of future news reading. *Journal of Print and Media Technology Research*, 2(3):131–139, 2013.
- Mikko Salminen, Niklas Ravaja, Kari Kallinen, and Timo Saari. Mediated cues of group emotion during knowledge-work tasks: Effects on subjective and physiological responses. *Interacting with Computers*, 25(1):60–73, 2013.
- Petri Savolainen, Sumi Helal, Jukka Reitmaa, Kai Kuikkaniemi, Giulio Jacucci, Mikko Rinne, Marko Turpeinen, and Sasu Tarkoma. Spaceify: A client-edge-server ecosystem for mobile computing in smart spaces. In *Proceedings of the 19th Annual International Conference on Mobile Computing & Networking (MobiCom '13)*, pages 211–214, 2013.

Social Interaction and Emotion (SIE),

Professor Niklas Ravaja

The mission of SIE is to increase our understanding of ICT-mediated social interaction. Our studies focus on: (a) emotional and cognitive processes during mediated social interaction (such as when playing videogames or using social network services, SNSs), (b) the neuroscience of social ICT, (c) privacy management and other socially-mediated behavioral phenomena in SNSs, (d) participation in online gift exchange and collaborative consumption, and (e) user experience of ICT.

We use a wide methodological approach, from surveys and qualitative methods to neurophysiological recordings (e.g., EEG), to study users in the laboratory as well as in real-life contexts. By providing both fundamental and applied insight into social interaction, we help to create ICT and services that take the human social and emotional processes into account.

Key publications:

- Jari Kätsyri, Riitta Hari, Niklas Ravaja, and Lauri Nummenmaa. Just watching the game ain't enough: Striatal fMRI reward responses to successes and failures in a video game during active and vicarious playing. *Frontiers in Human Neuroscience*, 7(7):1–13, 2013.
- Jari Kätsyri, Riitta Hari, Niklas Ravaja, and Lauri Nummenmaa. The opponent matters: Elevated fmri reward responses to winning against a human versus a computer opponent during interactive video game playing. *Cerebral Cortex*, 23(12):2829–2839, 2013.
- Sari Mullola, Markus Jokela, Mirka Hintsanen, Jari Lipsanen, Saija Alatupa, Niklas Ravaja, and Liisa Keltikangas-Järvinen. Associations between teacher-rated versus self-rated student temperament and school achievement. *Scandinavian Journal of Educational Research*, (1):1–26, 2013.
- Jussi Palomäki, Ilkka Kosunen, Kai Kuikkaniemi, Tetsuo Yamabe, and Niklas Ravaja. Anticipatory electrodermal activity and decision making in a computer poker-game. *Journal of Neuroscience, Psychology, and Economics*, 6(1):55–70, 2013.
- Niklas Ravaja, Outi Somervuori, and Mikko Salminen. Predicting purchase decision: The role of hemispheric asymmetry over the frontal cortex. *Journal of Neuroscience, Psychology, and Economics*, 6(1):1–13, 2013.
- Mikko Salminen, Niklas Ravaja, Kari Kallinen, and Timo Saari. Mediated cues of group emotion during knowledge-work tasks: Effects on subjective and physiological responses. *Interacting with Computers*, 25(1):60–73, 2013.
- Michiel Sovijärvi-Spapé, J. Matias Kivikangas, Simo Järvelä, Ilkka Kosunen, Giulio Jacucci, and Niklas Ravaja. Keep your opponents close: Social context affects EEG and fEMG linkage in a turn-based computer game. *PLOS ONE*, 8(11):78795, 2013.



Ubiquitous Interaction (UiX),

Professor Giulio Jacucci



Ubiquitous Interaction studies the design, development and evaluation of interactivity with users in ubiquitous and mobile computing systems. The research is carried out coupling in-depth user studies with design in the area of novel interfaces. The goal is to contribute to technology development efforts by deepening understanding of human and design perspectives. Currently, research is conducted in different areas: multitouch displays, adaptive and affective interfaces, multimodal interaction, exploratory search and social computing. Application areas range from sustainability and environmental awareness, home computing, walk-up-and-use display, information exploration, art and culture, scientific and knowledge work.

Key publications:

- Nicola Corradi, Konstantinos Priftis, Giulio Jacucci, and Luciano Gamberini. Oops, I forgot the light on! The cognitive mechanisms supporting the execution of energy saving behaviors. *Journal of Economic Psychology* 34: 88–96, 2013.
- Uta Hinrichs, Sheelagh Carpendale, Nina Valkanova, Kai Kuikkaniemi, and Giulio Jacucci and Andrew Vande Moere. Interactive Public Displays. *Computer Graphics and Applications, IEEE*, 33(2): 25–27, 2013.
- Salu Ylirisku, Sian Lindley, Giulio Jacucci, Richard Banks, Craig Stewart, Abigail Sellen, Richard Harpen, and Tim Regan. Designing web-connected physical artefacts for the 'aesthetic' of the home. In *Proceedings of the 2013 ACM Annual Conference on Human Factors in Computing Systems*, pages 909–918, ACM, 2013.
- Tuukka Ruotsalo and Matias Frosterus. Diversifying semantic entity search: Independent component analysis approach. *International Journal of Semantic Computing*, 7(4), 2013.
- Dorota Glowacka, Tuukka Ruotsalo, Ksenia Konyushkova, Kumaripaba Athukorala, Samuel Kaski, and Giulio Jacucci. Directing exploratory search: Reinforcement learning from user interactions with keywords. In *Proceedings of IUI'13, International Conference on Intelligent User Interfaces*, pages 117–128, New York, March 2013. ACM.
- Michiel Spapé, Matias Kivikangas, Simo Järvelä, Ilkka Kosunen, Giulio Jacucci, and Niklas Ravaja. Keep your opponents close: Social context affects EEG and fEMG linkage in a turn-based computer game. *PLOS ONE*, 8(11):78795, 2013.
- Eve Hoggan, Miguel Nacenta, Per Ola Kristensson, John Williamson, Antti Oulasvirta, and Anu Lehtiö. Multi-touch pinch gestures: performance and ergonomics. In *Proceedings of the 2013 ACM International Conference on Interactive Tabletops and Surfaces (ITS'13)*, pages 219–222, 2013.
- Eve Hoggan, John Williamson, Antti Oulasvirta, Miguel Nacenta, Per Ola Kristensson, and Anu Lehtiö. Multi-touch rotation gestures: performance and ergonomics. In *SIGCHI Conference on Human Factors in Computing Systems*, pages 3047–3050, 2013..



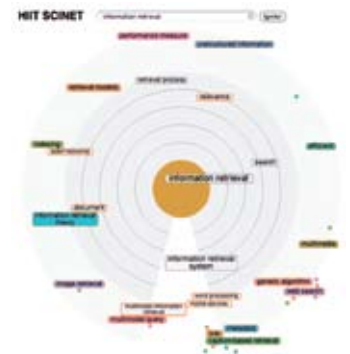
HIIT-Wide Focus Area: Augmented Science, Dr. Antti Honkela, Dr. Tuukka Ruotsalo (Coordinators)

The advances in science depend on effectively building upon the results that others have achieved, which are based on previously collected data. This brings to the forefront a challenge for computational science to better utilize the massive explosion in digital scientific data, whether it is scientific literature, or raw measurement data from previous studies. HIIT has begun focusing resources on a HIIT-wide focus area, titled Augmented Science, which builds on the existing excellences and touches most of HIIT research topics. Augmented science develops information technology methods and pilot applications for making the data-driven fields such as modern biology cumulative, and revolutionizing the way we search and access scientific resources and literature. Much of the most interesting biomedical data comes from individual people. We are developing methods that ensure the privacy of these sensitive data while allowing their use. Improving the general problem solving method of science, in collaboration with the other fields, is the best way for HIIT to contribute to solving the grand challenges of the humanity.

We aim to support scientific information access by enabling better coordination of communicating ideas and scientific results within the scientific community. Results of scientific efforts are traditionally published as articles and communicated personally as presentations or other related materials. The rapid communication of knowledge between researchers is a key success factor to enable better science. The volume of scientific output is estimated to be millions of publications worldwide per year; the growth rate of PubMed alone is more than 1 article per minute. The problem of communication that the scientific community is facing is shifting from publishing and sharing the information to finding and filtering the suitable materials to support every day work of researchers. We aim to help scientists to better find and manage the content that they use in their everyday work.

Key publications:

- Dorota Glowacka, Tuukka Ruotsalo, Ksenia Konyushkova, Kumaripaba Athukorala, Samuel Kaski, and Giulio Jacucci. Directing exploratory search: Reinforcement learning from user interactions with keywords. In Proceedings of IUI'13, International Conference on Intelligent User Interfaces, pages 117–128, New York, March 2013. ACM.
- Sohan Seth, Niko Välimäki, Samuel Kaski, and Antti Honkela. Exploration and retrieval of whole-metagenome sequencing samples. Bioinformatics, doi:10.1093/bioinformatics/btu340



Administration

Personnel and funding



HIIT is a joint research institute of Aalto University and the University of Helsinki. At Aalto University, HIIT researchers currently work in the Department of Information and Computer Science, Department of Computer Science and Engineering, and HIIT's own "department" at Open Innovation House; at University of Helsinki, our operations are at the Department of Computer Science, Department of Mathematics and Statistics, and Department of Social Sciences. The personnel of HIIT are employed by the two parent universities. Many of HIIT's personnel have double or even triple affiliations. Most common is an affiliation with other Departments of one or both of the parent universities, but there are also some who share their time between HIIT and some other organisation. Thus the total number of HIIT affiliated personnel (over 300) is much higher than the number of person-years, 154.

The total funding of HIIT in 2013 was 10 MEur, of which 79 per cent was competitive. The main funding sources were Academy of Finland, Tekes, EU and the mother universities. More details will be given in the Facts and Figures document found via <http://www.hiit.fi/abouthiit>

Board

The highest decision-making body of HIIT is the Board. It decides on HIIT's overall research strategy and research programmes. The statutory tasks of the Board are to approve the annual budget and activity plans, and follow up and comment on the work of HIIT through regular activity updates given by the Director of HIIT. In 2013 the Board convened four times.

The HIIT Board consists of nine members who are appointed for four years at a time. Each university appoints four board members, two of which are university staff and two are not employed by the university, and personal

deputies to each of them. The staff of HIIT selects one board member and his or her deputy from among their colleagues. Board members in 2013 are listed in the following, with their personal deputies in parentheses:

- Chairman: Vice President, Professor Ilkka Niemelä, Aalto (Dean, Professor Outi Krause, Aalto)
- Vice Chairman: Vice Rector, Professor Johanna Björkroth, UH (Vice Rector, Professor Jukka Kola, UH)
- Professor Olli Simula, Aalto (Professor Heikki Saikkonen, Aalto)
- Professor Hannu Toivonen, UH (Professor Jukka Paakki, UH)
- Vice President Henry Tirri, Nokia (Director Jyri Huopaniemi, Nokia)
- Docent Kari-Pekka Estola (Director Martin Mäklin, TeliaSonera Finland)
- Director Petri Vasara, Pöyry Management Consulting (Dr. Lars Gädda, Metsäklusteri)
- MSc Kimmo Kiviluoto, Webmie (Research Fellow Aimo Maanavilja, Elisa Communications)
- M.Soc.Sc. Airi Lampinen, HIIT (Docent Antti Honkela, HIIT)



The Director of HIIT Samuel Kaski was responsible for preparing and submitting propositions to the Board. Board Secretary was Research Coordinator, Docent Ella Bingham.

A new Board for the term 1 April 2014 - 31 March 2018 was nominated in early 2014 (personal deputies in parentheses):

- Chairman: Dean, Professor Risto Nieminen, Aalto (Vice Dean, Professor Jouko Lampinen, Aalto)
- Vice Chairman: Dean, Professor Jouko Väänänen, UH (Vice Dean, Professor Esko Ukkonen, UH)
- Professor Pekka Orponen, Aalto (Professor Lauri Savioja, Aalto)
- Professor Jukka Paakki, UH (Professor Jyrki Kivinen, UH)
- Vice President Hannu Kauppinen, Nokia (Director Jyri Huopaniemi, Nokia)
- Docent Kari-Pekka Estola (Innovation Director Ville Peltola, IBM)
- Director Mervi Karikorpi, The Federation of Finnish Technology Industries (Director Petri Vasara, Pöyry Management Consulting)
- CTO Harri Valpola, Zenrobotics (CEO Kimmo Kiviluoto, Enreach Solutions)
- Personnel representative: MSc Joanna Bergström-Lehtovirta, HIIT (Docent Antti Honkela, HIIT)



Scientific Advisory Board (SAB)

The SAB consists of internationally prominent scholars who are invited by the HIIT Board. The SAB convened in 2012 and its recommendations were actively taken into account during 2013. SAB Members in 2012 are listed in the following.

- Professor Randy Katz, University of California at Berkeley, Chairman of SAB
- Professor Alberto Apostolico, Georgia Tech
- Professor Christos Faloutsos, Carnegie Mellon University
- Professor Jodi Forlizzi, Carnegie Mellon University
- Professor Bengt Jonsson, Uppsala University
- Professor Martin Kersten, University of Amsterdam and CWI
- Professor Kari-Jouko Räihä, University of Tampere
- Professor Mart Saarma, University of Helsinki
- Professor John Shawe-Taylor, University College London

