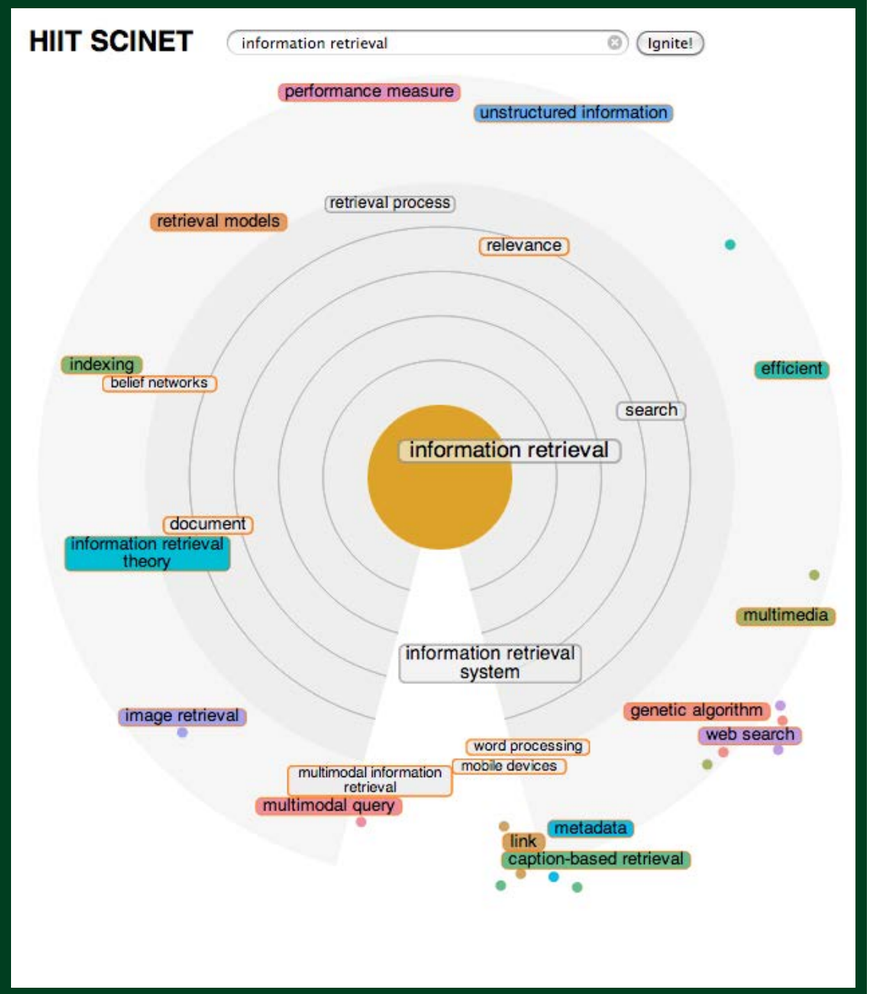
Aalto University

UNIVERSITY OF HELSINKI

HELSINKI INSTITUTE FOR INFORMATION TECHNOLOGY

Helsinki Institute for Information Technology HIIT

Annual Report 2012



Ella Bingham (ed.)

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

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. Its research ranges from fundamental methods and technologies to novel applications and their impact on people and society. The current foci of research are computational modelling and data analysis, and ubiquitous ICT in the modern networked world. HIIT groups belong to four 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 about 11 million euros, realising in total about 144 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 at University of Helsinki's Kumpula campus. 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 and other foundations financing Finnish hi-tech research, the European Union and private companies. 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 2012

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 Hällströmin katu 2b, Helsinki
Telephone: +358 9 1911
Fax: +358 9 191 51120

Redfina, Espoo 2013

Cover image by the HIIT-Wide Focus Area
Lay-out: Anna Bergman, Noora Suominen de Rios

Table of Contents

HIIT in brief

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. Its research ranges from fundamental methods and technologies to novel applications and their impact on people and society. The current foci of research are computational modelling and data analysis, and ubiquitous ICT in the modern networked world.

HIIT groups belong to four 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 about 11 million euros, realising in total about 144 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 at University of Helsinki's Kumpula campus. 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 and other foundations financing Finnish hi-tech research, the European Union and private companies.

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
Funding M€	9.0	12.1	10.5	11.5
External/competitive funding	71 %	83 %	78 %	83 %
Person-years	185	175	135	144
Refereed publications	177	222	213	209

Review of Year 2012

What's new

Scientific Advisory Board assessment of HIIT

The Scientific Advisory Board (SAB) of HIIT convened in June 2012 to assess HIIT's recent changes and future directions. The SAB's overall impression on HIIT was very positive: HIIT has a right balance between scientific rigor and applications; HIIT's research applications are inspired by research opportunities instead of being dictated by the industry; HIIT has implemented the recommendations of the 2008 SAB report; HIIT is clearly able to attract international talent; and the SAB supports the recent changes and the current direction. The SAB also considered that "HIIT provides a high return on external funding from a surprisingly modest investment in the Institute, and the SAB strongly encourages the universities to increase the level of funding". More details of the SAB report are given in the end of this Annual Report.

Photo by Olli Pitkänen

Rearrangement of HIIT programmes to align with new Centres of Excellence and with new research spearheads

HIIT has been particularly successful regarding Centres of Excellence of the Academy of Finland. During the period 2012-2017 HIIT contributes to 3 new Centres: Finnish Centre of Excellence in Computational Inference Research COIN, Finnish Centre of Excellence in Inverse Problems Research, and Finnish Centre of Excellence in Cancer Genetics Research. The fourth one is the Finnish Centre of Excellence in Algorithmic Data Analysis Research (Algodan, 2008-2013) The research programmes of HIIT were rearranged in the beginning 2012, to better align with these Centres of Excellence (CoE), and to reflect new areas of investigation. HIIT's Algorithmic Data Analysis programme now consists of groups belonging to the Algodan CoE and the new Computational Inference programme consists of groups belonging to the COIN CoE. A new programme Scalable Networks started with spearhead activities: a holistic view on energy efficiency, and new programming paradigms for massively scalable networks.

New group by professor Jukka Corander

A new group "Bayesian Statistics", led by Jukka Corander, professor of statistics in the Department of Mathematics and Statistics of University of Helsinki, started in the Computational Inference (CI) programme. Professor Corander is an ERC grant holder and was recruited to strengthen HIIT's strategic core area of probabilistic modeling.

New group by professor Juho Rousu

A new group “Kernel Machines, Pattern Analysis and Computational Biology”, led by Juho Rousu, professor in the Department of Information and Computer Science of Aalto University, started in the Algorithmic Data Analysis (ADA) programme. The group particularly focusses in learning with multiple and structured targets, multiple views and ensembles. Applications of interest in computational biology include network reconstruction, gene functional classification as well as biomarker discovery.

New group by professor Keijo Heljanko

Keijo Heljanko, professor in the Department of Computer Science and Engineering of Aalto University, started in the Scalable Networks programme. Professor Heljanko’s research group, Distributed Computation Group, has two main research topics: computer aided verification and distributed computation. Towards the end of 2012, professor Heljanko was appointed the Director of the Scalable Networks programme, and started to refocus the activities of the programme.

Open Innovation House (OIH)

In summer 2012 a welcomed change took place as HIIT researchers located at Innopoli business park moved to new premises more centrally located at Aalto University’s Otaniemi campus, sharing premises with the EIT ICT Labs Helsinki node.

Helsinki Distinguished Lecture Series on Future Information Technology

During 2012 HIIT started a “Nobel laureate level” lecture series which is open to public and whose aim is to raise the visibility of HIIT and its mother universities. The first speaker was professor Tuomas Sandholm from Carnegie Mellon University.

HIIT-Wide Focus Area

As of 2011, HIIT has begun focusing resources 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 Research, IT methods and pilot tools are developed for making the research 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.

Other key achievements in 2012

Researchers from HIIT and Institute for Molecular Medicine Finland FIMM, led by HIIT's professor Samuel Kaski, excelled in predicting cancer drug sensitivity in an international crowdsourcing challenge. The team developed a winning solution for predicting responses of breast cancer cells to a set of cancer drugs. The prediction is based on the genomic profiles of the cancer cells. Harnessing genomic profiles of cells in choosing the best treatment is considered the holy grail of personalised medicine. The team participated in the seventh annual DREAM competition organised by the U.S. National Cancer Institute (NCI) and The Dialogue for Reverse Engineering Assessments and Methods (DREAM).

HIIT's postdoctoral researchers Pekka Marttinen and Simon Puglisi got an Academy Postdoctoral Researcher position of the Academy of Finland, and Dr Antti Honkela was appointed Academy Research Fellow.

SAT Challenge 2012 (SC 2012), a competitive event for solvers of the Boolean Satisfiability (SAT) problem, was co-organized by HIIT postdoc Matti Järvisalo, as a satellite event to the 15th International Conference on Theory and Applications of Satisfiability Testing (SAT 2012). The Fifth Workshop on Information Theoretic Methods in Science and Engineering (WITMSE) was co-organized by HIIT researchers Petri Myllymäki, Teemu Roos and Jorma Rissanen. 13th Scandinavian Symposium and Workshops on Algorithm Theory (SWAT 2012) was co-organized by professor Petteri Kaski of HIIT.

Collaboration

Internationality is an integrated part of all operations. Cooperation with key international research institutes and univarrersities is active: Berkeley (ICSI and UC Berkeley), Tsinghua University, MIT, NICTA and UCL, to name a few. Especially noteworthy is the long collaboration with Berkeley where several HIIT researchers spend 3-12 months each year.

Photo by Olli Pitkänen

HIIT has been successful in recruiting excellent postdocs from abroad, indicating a research environment that is fully competitive in international comparison. In turn, the graduates from HIIT have typically obtained very good positions in industry or in academia both in Finland and abroad. The same applies to HIIT alumni: Heikki Mannila, Director of HIIT in 2009 and Chairman of the Board from 2010 to early 2012, starts as the President of the Academy of Finland in Spring 2012.

The Strategic Centres for Science, Technology and Innovation (SHOKs) have played a major role in shaping the collaboration between universities and companies in Finland. HIIT researchers are active in the SHOK on Information and Communication Technologies (ICT-SHOK Tivit), having prominent roles as academic coordinators in Future Internet and Next Media, as well as in the new programmes Internet of Things and From Data to Intelligence (D2I). The D2I programme, with 46 companies and universities participating, builds on data analysis and modelling, a core competence of HIIT. HIIT researchers participate also in the Health field SHOK SalWe.

EIT ICT Labs is a new 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 6 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.

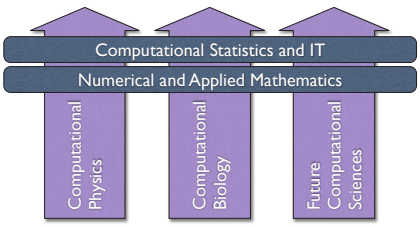
Examples of other international collaboration are numerous. Many HIIT groups participate actively in the activities of the EU Network of Excellence PASCAL2 in which HIIT professors Petri Myllymäki and Samuel Kaski are Members of the Steering Committee.

Photo by Olli Pitkänen

Societal impact

HIIT aims to boost the competitiveness of the Finnish information industry and information society in the long term 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. HIIT has also contributed to policy development both at the national and EU level by co-operating with the various branches of public administration. HIIT research has also direct societal impact through collaborative research done with different public organizations such as the Finnish Federation of the Visually Impaired, Finnish Association for Intellectual and Developmental Disabilities, the Finnish National Institute of Health and Welfare, and the National Board of Patents and Registration, and the Mechanical Engineering and Metals Industry Standardization Association of Finland.



Doctoral studies

HIIT coordinates two national doctoral programmes: Finnish Doctoral Programme in Computational Science FICS (directed by Professor Samuel Kaski of HIIT) and Future Internet Graduate School FIGS (directed by Professor Jussi Kangasharju of HIIT and Professor Joerg Ott). In addition, Helsinki Doctoral Programme in Computer Science -- Advanced Computing and Intelligent Systems (Hecse) is led by HIIT programme director, professor Petri Myllymäki.

Since 2012, HIIT partners with FIMM, Institute for Molecular Medicine Finland, in recruiting doctoral 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 very high level doctoral candidates from all over the world.

Both Aalto University and University of Helsinki are in the process of reorganizing their doctoral training, and HIIT's researchers have taken an active role in these processes. HIIT has also contributed strongly in establishing the EIT ICT Labs Doctoral School, and will remain an important partner in the Helsinki node of the school.

10 HIIT researchers got their PhD during the year.

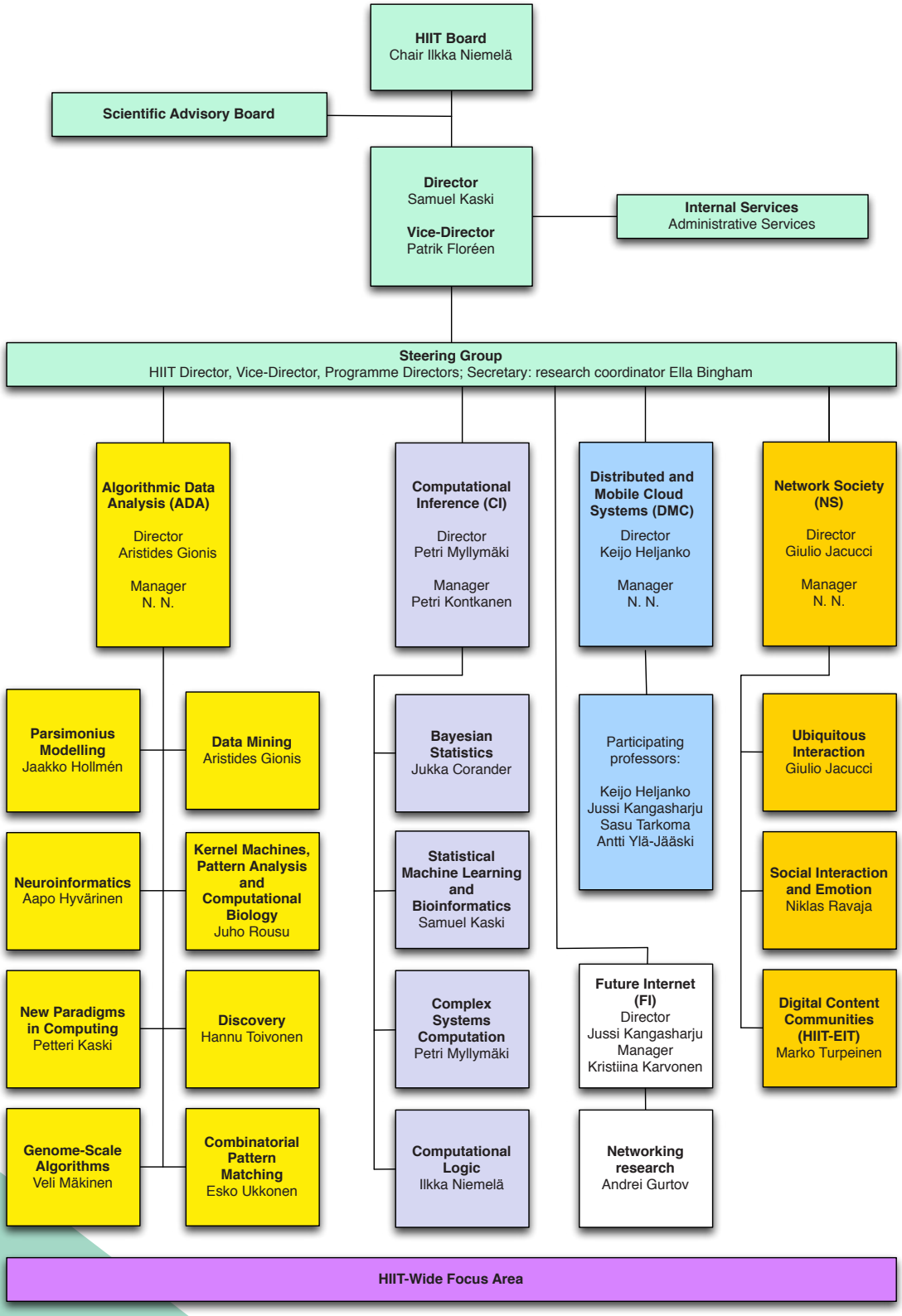
Views for 2013

A new research group named Data Mining, led by a new professor Aristides Gionis at Aalto University, will start in the beginning of 2013. Professor Gionis will also start 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.

The Scalable Networks programme will be refocused in early 2013, with a new name Distributed and Mobile Computing Systems (DMC) and a new director, professor Keijo Heljanko. The main focus of the programme is to bridge together two separate worlds into a seamless combined distributed cloud computing system: Distributed cloud based backend systems, and Mobile devices. A combined platform for mobile devices and cloud is needed because of inherent limitations in the battery capacity, CPU power, storage and networking bandwidth of current mobile devices.

The Helsinki Distinguished Lecture Series on Future Information Technology, coordinated by HIIT, will continue in 2013. Confirmed speakers include Dr Eric Brown from IBM Research, professor Leslie Valiant from Harvard University, professor Christos Papadimitriou from University of California Berkeley, and Executive Vice President and CTO Henry Tirri from Nokia.

HIIT in early 2013



Research

Algorithmic Data Analysis (ADA) Programme

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 2012.

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 group is interested in the basic research of the theoretical aspects of the area as well as in various applications, mostly in bioinformatics and information retrieval.

Recently we have focused in developing algorithms for assembling genomes from short read data. We have proposed new algorithms for correcting sequencing errors in the data. We have also developed new methods for the scaffolding phase where contiguous sequences assembled from the short read data are arranged into a linear order.

Another recent research focus is regulatory regions in genomes. We have developed new models based high throughput data for describing regulatory elements and we have proposed new algorithms for recognizing such elements in genomic sequences.

Key publications:

- V. Mäkinen, L. Salmela, and J. E. Ylinen. Normalized n50 assembly metric using gap-restricted co-linear chaining. BMC Bioinformatics, 13(255), 2012.
- L. Salmela. Average complexity of backward q-gram string matching algorithms. Information Processing Letters, 112(11):433–437, 2012.
- J. Fischer, T. Gagie, T. Kopelowitz, M. Lewenstein, V. Mäkinen, L. Salmela, and N. Välimäki. Forbidden patterns. In Proceedings of the 10th Latin American Symposium on Thoretical Informatics (LATIN 2012), Lecture Notes in Computer Science, pages 327–337. Springer Berlin Heidelberg, 2012.

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. Application areas range from link discovery in bioinformatics to computational generation of poetry.

Methodologically, we focus on analysis and exploration methods for weighted (biological) graphs. We identify relevant computational problems, develop new concepts and algorithms, and apply them. We have recently developed a range of novel methods to simplify large networks into simpler ones or for extracting relevant information from them. These methods allow more efficient and user-friendly analysis of social networks, biological networks, etc.

The more applied line of this research has produced Biomine, a search engine prototype that integrates and indexes data from several publicly available biological databases. Biomine presents the data as a weighted graph, and its query tools aid explorative discovery of non-trivial connections between biological entities, such as genes and phenotypes. See http://biomine.cs.helsinki.fi/.

A new focus area is computational creativity, interesting on its own right but also as an application area for data mining methods. We mostly work on verbal creativity, i.e., computational poetry and humour. We are developing novel methods that minimize the need for manually coded or language specific knowledge.

Key publications:

- L. Eronen and H. Toivonen. Biomine: Predicting links between biological entities using network models of heterogeneous databases. BMC Bioinformatics, 13(119), 2012.
- M. E. Skarkala, H. Toivonen, P. Moen, M. Maragoudakis, S. Gritzalis, and L. Mitrou. Privacy preservation by k-anonymization of weighted social networks. In Proceedings of IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012), pages 486–491, 2012.
- J. Toivanen, H. Toivonen, A. Valitutti, and O. Gross. Corpus-based generation of content and form in poetry. In Proceedings of the Third International Conference on Computational Creativity (ICCC 2012), pages 175–179, 2012.
- A. Tulilaulu, J. Paalasmaa, M. Waris, and H. Toivonen. Sleep Musicalization: Automatic Music Composition from Sleep Measurements. Eleventh International Symposium on Intelligent Data Analysis (IDA), LNCS 7619, pages 392-403, 2012.
- A. Valitutti, H. Toivonen, O. Gross, and J. M. Toivanen. Decomposition and Distribution of Humorous Effect in Interactive Systems. Artificial Intelligence of Humor, the AAAI Fall Symposium Series, pages 96-100, 2012.

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 statistical data analysis. We model the visual system in the brain by analyzing the statistical structure of the natural input images. We also develop the relevant theory of statistical multivariate modelling, and apply such models on different kinds of neuroscientific measurement data. We recently developed a new method for estimating unnormalized probabilistic models, based on the intuitive idea of learning to discriminate the observed data from artificial noise by logistic regression (Gutmann and Hyvärinen, 2012).

Key publications:

- M. U. Gutmann and A. Hyvärinen. Noise-contrastive estimation of unnormalized statistical models, with applications to natural image statistics. Journal of Machine Learning Research, 13:307–361, 2012.
- A. Hyttinen, F. Eberhardt, and P. Hoyer. Learning linear cyclic causal models with latent variables. Journal of Machine Learning Research, 13:3387–3439, 2012.
- A. Moneta, D. Entner, P. Hoyer, and A. Coad. Causal inference by independent component analysis with applications to micro- and macro-economic data. Oxford Bulletin of Economics and Statistics, 2012.
- P. Ramkumar, L. Parkkonen, R. Hari, and A. Hyvärinen. Characterization of neuromagnetic brain rhythms over time scales of minutes using spatial independent component analysis. Human Brain Mapping, 33(7):1648–1662, 2012.
- A. Hyttinen, F. Eberhardt, and P. Hoyer. Causal discovery of linear cyclic models from multiple experimental data sets with overlapping variables. In Proceedings of the 28th Conference on Uncertainty in Artificial Intelligence 2012 (UAI-12), 2012.
- J.-I. Hirayama, A. Hyvärinen, and S. Ishii. Structural equations and divisive normalization for energy-dependent component analysis. In Advances in Neural Information Processing Systems (NIPS 2011), Volume 24, 2012.

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.

Recently, we have developed methods to yield parsimonious models using sparsity-inducing projections (Prada et al., 2012) and search-based methods (Korpela et al., 2012).

Key publications:

- T. Niini, I. Scheinin, L. Lahti, S. Savola, F. Mertens, J. Hollmén, T. Böhling, A. Kivioja, K. H. Nord, and S. Knuutila. Homozygous deletions of cadherin genes in chondrosarcoma - an array CGH study. Cancer Genetics, 205(11):588–593, 2012.
- M. A. Prada, J. Toivola, J. Kullaa, and J. Hollmén. Three-way analysis of structural health monitoring data. Neurocomputing, (80):119–128, 2012.
- P. Raj Adhikari and J. Hollmén. Fast progressive training of mixture models for model selection. In J-G Ganascia, P Lenca, and J-M Petit, editors, Proceedings of Fifteenth International Conference on Discovery Science (DS 2012), pages 194–208. Springer-Verlag, October 2012.
- P. Raj Adhikari and J. Hollmén. Multiresolution mixture modeling using merging of mixture components. In Steven CH Hoi and Wray Buntine, editors, Proceedings of Fourth Asian Conference on Machine Learning (ACML 2012), pages 17–32, November 2012.
- J. Hollmén. Mixture modeling of gait patterns from sensor data. In Proceedings of the 5th International Conference on Pervasive Technologies and Relative to Assistive Environments. ACM, 2012.

Kernel Machines, Pattern Analysis and Computational Biology, Professor Juho Rousu

The group is part of HIIT since the beginning of 2012. The group develops machine learning methods, models and tools for computational sciences, in particular computational biology. The methodological backbone of the group is kernel methods and regularized learning.

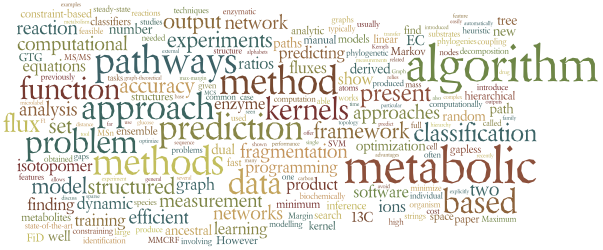
The group particularly focusses in learning with multiple and structured targets, multiple views and ensembles. Applications of interest in computational biology include network reconstruction, gene functional classification as well as biomarker discovery.

Our recent research includes a novel approach for metabolite identification, an important bottleneck problem in systems biology. Our method rely on predicting molecular fingerprints based on a large collection of tandem mass spectra and probabilistic retrieval of molecules using these fingerprints from a large molecular database (Heinonen et al. 2012a). The method has been shown to provide predictive performance competitive with state of the art.

Other major research directions are biological network inference, where we develop new machine learning approaches integrating heterogeneous biological data. In particular, dechiphering biological networks in metagenomic context, such as deep biosphere or human gut microflora, are of interest.

Key publications:

- M. Heinonen, H. Shen, N. Zamboni, and J. Rousu. Metabolite identification and molecular fingerprint prediction through machine learning. Bioinformatics, 28(18):2333–2341, 2012.
- M. Heinonen, N. Välimäki, V. Mäkinen and J. Rousu. Efficient Path Kernels for Reaction Function Prediction. Bioinformatics Models, Methods and Algorithms, 2012.
- M. Heinonen. Computational Methods for Small Molecules. PhD thesis. Report A-2012-9, Department of Computer Science, University of Helsinki, 2012



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:

- V. Mäkinen, L. Salmela, and J. E. Ylinen. Normalized n50 assembly metric using gap-restricted co-linear chaining. BMC Bioinformatics, 13(255), 2012.
- N. Välimäki, S. Ladra, and V. Mäkinen. Approximate all-pairs suffix/prefix overlaps. Information and Computation, 213:49–58, 2012.
- J. Fischer, T. Gagie, T. Kopelowitz, M. Lewenstein, V. Mäkinen, L. Salmela, and N. Välimäki. Forbidden patterns. In Proceedings of the 10th Latin American Symposium on Thoretical Informatics (LATIN 2012), Lecture Notes in Computer Science, pages 327–337. Springer Berlin Heidelberg, 2012.
- N. Välimäki. Least random suffix/prefix matches in output-sensitive time. In Proceedings of the 23rd Annual Symposium on Combinatorial Pattern Matching (CPM 2012), Lecture Notes in Computer Science, pages 269–279. Springer Berlin Heidelberg, 2012.
- N. Välimäki and S. J. Puglisi. Distributed string mining for high-throughput sequencing data. In Proceedings of the 12th International Workshop on Algorithms in Bioinformatics (WABI 2012), Lecture Notes in Computer Science, pages 441–452. Springer Berlin Heidelberg, 2012.

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.

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:

- A. Björklund, T. Husfeldt, P. Kaski, and M. Koivisto. The traveling salesman problem in bounded degree graphs. ACM Transactions on Algorithms, 8(2):Article Number: 18, 2012.
- I. Kostitsyna and V. Polishchuk. Simple wriggling is hard unless you are a fat hippo. Theory of Computing Systems, 50(1):93–110, 2012.
- A. Björklund, T. Husfeldt, P. Kaski, M. Koivisto, J. Nederlof, and P. Parviainen. Fast zeta transforms for lattices with few irreducibles. In Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms (SODA ‘12), pages 1436–1444, 2012.
- S. Gaspers, M. Koivisto, M. Liedloff, S. Ordyniak, and S. Szeider. On finding optimal polytrees. In Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence (AAAI 2012), 2012.
- M. Göös, J. Hirvonen, and J. Suomela. Lower bounds for local approximation. In Proceedings of the 2012 ACM SIGACT-SIGOPS Symposium on Principles of Distributed Computing (PODC’12), pages 175–184, 2012.
- J. Hirvonen and J. Suomela. Distributed maximal matching: greedy is optimal. In Proceedings of the 2012 ACM SIGACT-SIGOPS Symposium on Principles of Distributed Computing (PODC’12), pages 165–174, 2012.
- T. Niinimäki and P. Parviainen. Local structure discovery in Bayesian networks. In Proceedings of the Twenty-Eighth Conference on Uncertainty in Artificial Intelligence (UAI-12), pages 634–643, 2012.

Computational Inference (CI) Programme

The four groups of the CI programme are all members of the recently launched 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 models 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:

- J. Caldas, N. Gehlenborg, E. Kettunen, A. Faisal, M. Rönty, A. G. Nicholson, S. Knuutila, A. Brazma, and S. Kaski. Data-driven information retrieval in heterogeneous collections of transcriptomics data links SIM2s to malignant pleural mesothelioma. Bioinformatics, 28(2):246–253, 2012.
- P. Glaus, A. Honkela, and M. Rattray. Identifying differentially expressed transcripts from RNA-seq data with biological variation. Bioinformatics, 28(13):1721–1728, 2012.
- S. A. Khan, A.Faisal, J. P. Mpindi, J. A. Parkkinen, T. Kalliokoski, A. Poso, O. P. Kallioniemi, K. Wennerberg, and S. Kaski. Comprehensive data-driven analysis of the impact of chemoinformatic structure on the genome-wide biological response profiles of cancer cells to 1159 drugs. BMC Bioinformatics, 13:Article Number: 112, 2012.
- M. Koskinen, J. Viinikanoja, M. Kurimo, A. Klami, S. Kaski, and R. Hari. Identifying fragments of natural speech from the listener's MEG signals. Human Brain Mapping, 34(6):1477-1489, 2012.
- G. Leen, J. Peltonen, and S. Kaski. Focused multi-task learning in a gaussian process framework. Machine Learning, 89(1-2, SI):157–182, 2012.
- M. K. Titsias, A. Honkela, N. D. Lawrence, and M. Rattray. Identifying targets of multiple co-regulating transcription factors from expression time-series by Bayesian model comparison. BMC Systems Biology, 6:53, 2012.
- S. Virtanen, Y. Jia, A. Klami, and T. Darrell. Factorized multi-modal topic model. In Proc. 28th Conference on Uncertainty in Artificial Intelligence (UAI), pages 843–851, 2012.
- S. Virtanen, A. Klami, S. A. Khan, and S. Kaski. Bayesian group factor analysis. In Proc. 15th International Conference on Artificial Intelligence and Statistics, pages 1269–1277, 2012.

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. This focus area is concerned with the issue of how to do the modelling when the models are very complex in the sense of being structured by prior information and constraints. Despite the wealth of documented success stories about inference in highly structured stochastic models, awareness about the failure of standard stochastic computation techniques to handle increasing complexity of models is accumulating in the inference research community. In order to scale up existing stochastic learning methods, we will develop new types of intelligent adaptive algorithms. Our main application area is computational biology, in particular models for evolution and transmission of bacteria and viruses, where key collaborators represent world-leading institutions in infectious disease epidemiology (Broad Institute, Harvard School of Public Health, Imperial College London, MIT, Sanger Institute).

Key publications:

S. Castillo-Ramirez, J. Corander, P. Marttinen, M. Aldeljawi, W. Hanage, H. Westh, K. Boye, Z. Gulay, S. Bentley, J. Parkhill, M. Holden, and E. Feil. Phylogeographic variation in recombination rates within a global clone of Methicillin-Resistant Staphylococcus aureus (MRSA). Genome Biology, 13(12):R126, 2012.

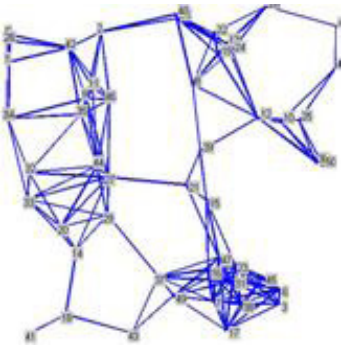
L. Cheng, A. W. Walker, and J. Corander. Bayesian estimation of bacterial community composition from 454 sequencing data. Nucleic Acids Research, 40, 5240-5249, 2012.

J. Sirén, W. P. Hanage, and J. Corander. Inference on Population Histories by Approximating Infinite Alleles Diffusion. Molecular Biology and Evolution, 30(2):457-468, 2012.

R. Gauriot, L. Gunaratnam, R. Moroni, T. Reinikainen, and J. Corander. Statistical Challenges in the Quantification of Gunshot Residue Evidence. Journal of Forensic Sciences, in press, 2012.

T. D. Laajala, J. Corander, N. M. Saarinen, K. Mäkelä, S. Savolainen, M. I. Suominen, E. Alhoniemi, S. Mäkelä, M. Poutanen, and T. Aittokallio. Improved statistical modeling of tumor growth and treatment effect in pre-clinical animal studies with highly heterogeneous responses in vivo. Clinical Cancer Research, 18(16); 4385–96, 2012.

P. Blomstedt, and J. Corander. Posterior predictive comparisons for the two-sample problem. Communications in Statistics – Theory and Methods, in press, 2012.



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, mathematics and statistics, and an applied component where the results are used for solving problems in various disciplines from sociological and historical studies to industrial engineering.

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

Key publications:

- W. Dvořák, M. Järvisalo, J.P. Wallner, and S. Woltran. Complexity-Sensitive Decision Procedures for Abstract Argumentation. In Proceedings of the 13th International Conference on Principles of Knowledge Representation and Reasoning (KR 2012), pages 54-64. AAAI Press, 2012. [KR 2012 Distinguished Student Paper Prize]
- M. Järvisalo, A. Biere, and M. J.H. Heule. Simulating Circuit-Level Simplifications on CNF. Journal of Automated Reasoning 49(4):583-619, 2012.
- A. Medlar, D. Glowacka, H. Stanescu, K. Bryson, and R. Kleta. Swiftlink: Paralle MCMC linkage analysis utilising multicore CPU and GPU. Bioinformatics, 2012.
- T. Pulkkinen, and P. Nurmi. AWESOM: Automatic Discrete Partitioning of Indoor Spaces for WiFi fingerprinting. Proceedings of the 10th International Conference on Pervasive Computing (Pervasive), pp. 271-288, 2012.
- A. Oulasvirta, A. Pihlajamaa, J. Perkiö, T. Vähäkangas, D. Ray, N. Vainio, P. Myllymäki, and T. Hasu. Long-term Effects of Ubiquitous Surveillance at Home. Proceedings of the 14th International Conference on Ubiquitous Computing (Ubicomp), September, 2012.
- J. Rissanen, Optimal Estimation of Parameters. Cambridge University Press, 2012.
- V. Lehtinen, A. Oulasvirta, A. Salovaara, P. Nurmi. Dynamic Tactile Guidance for Visual Search Tasks. Proceedings of the 25th ACM Symposium on User Interface and Software Technology (UIST, October 2012, Cambridge, MA).

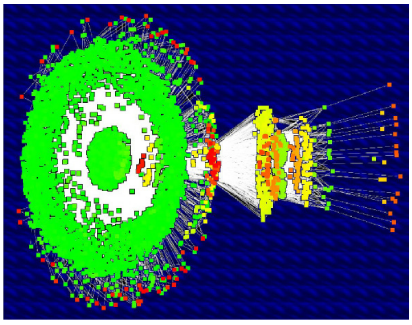
Computational Logic, Professor Ilkka Niemelä

The group develops automated reasoning techniques for solving challenging computational problems in engineering and science. The current focus is on efficient computational methods for solving large constraint satisfaction problems including SAT, satisfiability modulo theories, and rule-based constraints -- also known as 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 2012 we have continued developing novel techniques for reasoning on parity constraints, a constraint class that frequently occurs in some application areas but is not optimally handled by classic techniques such as resolution. In addition to introducing new reasoning techniques, we have also studied decomposition, classification, and simulation of parity constraint instances. Moreover, we devised a new translation from ASP into mixed integer programming (MIP) which enables the use of MIP solvers for the computation of answer sets and extends rule-based modeling by linear constraints.

In 2012 we have also continued devising methods for the model checking analysis of safety critical timed systems employing real-valued clocks. In particular, we have developed a timed version of the very competitive IC3 (or PDR) induction algorithm introduced recently by Aaron Bradley.

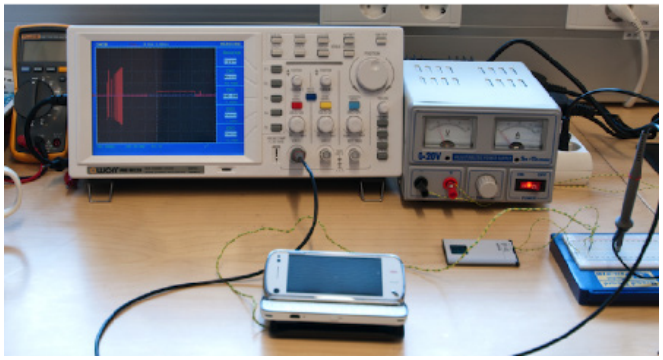
Key publications:

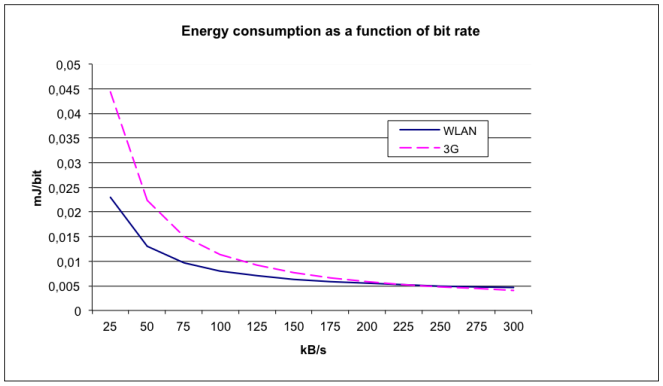


- H. Heikinheimo, J. T. Eronen, A. Sennikov, C. D. Preston, E. Oikarinen, P. Uotila, H. Mannila, and M. Fortelius. Convergence in the distribution patterns of Europe's plants and mammals is due to environmental forcing. Journal of Biogeography, 39(9):1633–1644, 2012.
- K. Heljanko, M. Keinänen, M. Lange, and I. Niemelä. Solving parity games by a reduction to SAT. Journal of Computer and System Sciences, 78(2):430–440, 2012.
- R. Kindermann, T. Junttila, and I. Niemelä. SMT-Based Induction Methods for Timed Systems. In FORMATS 2012, volume 7595 of Lecture Notes in Computer Science, pages 171-187, Springer, 2012.
- T. Laitinen, T. Junttila, and I. Niemelä. Conflict-Driven XOR-Clause Learning. In SAT 2012, volume 7317 of Lecture Notes in Computer Science, pages 383-396, Springer, 2012.
- T. Laitinen, T. Junttila, and I. Niemelä. Classifying and Propagating Parity Constraints. In CP 2012, volume 7514 of Lecture Notes in Computer Science, pages 357-372, Springer, 2012.
- G. Liu, T. Janhunen, and I. Niemelä. Answer set programming via mixed integer programming. In Principles of Knowledge Representation and Reasoning, pages 32–42, June 2012.

Scalable Networks (SN) Programme

The programme started during 2012 with an aim to work on the design, analysis and evaluation of extremely large-scale networked systems. The topics include research into novel networking technologies as well as into foundations of large scale computing systems, including cloud computing. The programme has two spearhead activities: a holistic view on energy efficiency, and new programming paradigms for massively scalable networks.





Combinatorial Algorithms and Computation, Professor Pekka Orponen

Our work in the area of combinatorial algorithms and computation theory is structured along three research themes: Modern Algorithmics, Computational Models and Mechanics, and Algorithmics for Data Security.

The Modern Algorithmics team applies combinatorial and complexity-theoretic methods to the solution of algorithmic problems in challenging operating environments, often characterised by incomplete information about the global system state. Recent areas of research have been online algorithms for production planning applications, and energy-awareness issues in ad hoc and sensor networks, such as the design and analysis of energy-efficient routing methods, and lifetime-maximising network topology control. In summer 2012, a pilot project was pursued on using techniques from distributed algorithmics to coordinate the actions of a small ensemble of low-cost swarm robots.

The Computational Models and Mechanics team studies methods for the solution of computational problems in structurally complex state spaces. The main focus is on techniques that are algorithmically relatively simple, but which adapt effectively to the characteristics of the problem instance at hand. The specific topics considered include e.g. the behaviour of graph algorithms on small-world networks and the design and analysis of stochastic search methods for complex optimisation landscapes. A recent area of interest have been algorithmic issues related to the design of self-assembling nanoscale structures.

The Algorithms for Data Security team works on advanced static and dynamic malware-detection methods on host-based systems. This includes analysis of e.g. call graphs of malware executables and dynamic sandbox execution data, provided by the team's industrial partners.

Key publications:

- E. Czeizler, A. Mizera, E. Czeizler, R.-J. Back, J. E. Eriksson, and I. Petre. Quantitative analysis of the self-assembly strategies of intermediate filaments from tetrameric vimentin. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 9(3):885–898, 2012.
- E. Czeizler, V. Rogojin, and I. Petre. The phosphorylation of the heat shock factor as a modulator for the heat shock response. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 9(5):885–898, 2012.
- S. Seki. Absoluteness of subword inequality is undecidable. Theoretical Computer Science, (418):116–120, 2012.

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. In addition, reducing energy consumption and carbon footprint has been widely recognized as a challenge for the whole ICT industry. The growing operational expenditure in ICT calls for more energy-aware networking solutions throughout the entire end-to-end communication chain. 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. Mobile cloud computing provides versatile research topics varying from the virtualization layer up to the service layer. Machine-to-machine communication, wireless sensor networks, and peer-to-peer applications are other examples of active research areas in the distributed services domain

Key publications:

- A. Lukyanenko, E. Morozov, and . Gurtov. An adaptive backoff protocol with Markovian contention window control. Communications in Statistics - Simulation and Computation, 41(7):1093–1106, 2012.
- Y. Xiao, W. Li, M. Siekkinen, P. Savolainen, A. Ylä-Jääski, and P. Hui. Power management for wireless data transmission using complex event processing. IEEE Transactions on Computers, 61(12):1765–1777, 2012.

Distributed Computing, Professor Keijo Heljanko

The group is focusing on distributed computing technologies. The work is motivated by the raise of very large computing infrastructure, called “Warehouse Scale Computing”, where the power of a very large number of computers is harnessed to a single computing platform. The work involves research into new computing paradigms, such as the Google developed Map-Reduce programming paradigm, and its refinements, such as Berkeley Spark. The approach to research is based on using these paradigms to implement scalable algorithms for science and engineering, and using the gathered know-how to improve the frameworks themselves.

Some of the challenges in the frameworks include the difficulty of executing iterative algorithms efficiently using the programming frameworks, as well as lack of good methods for running low latency computations while still retaining good scalability to high computer counts. These shortcomings can be attacked by either novel algorithms avoiding these weaknesses of the frameworks or by improving the frameworks themselves in these areas of weakness. We are investigating both approaches to solving these problems.

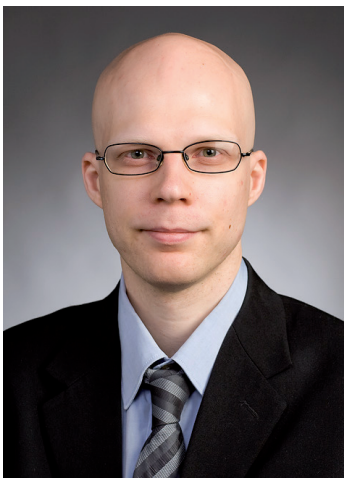
The initial areas where the group is applying the new computing paradigms are the processing of very large datasets in next generation sequencing data, and the area of computer aided verification, where the large scale computing machinery is put into use to find bugs in distributed systems in a fully automated fashion.

The work is based on research into software systems combined with the background of the group in formal methods for designing correctly functioning parallel and distributed systems.

Key publications:

- J. Dubrovin, T. Junttila, and K. Heljanko. Exploiting step semantics for efficient bounded model checking of asynchronous systems. Science of Computer Programming, 77(10-11):1095–1121, 2012.
- K. Heljanko, M. Keinänen, M. Lange, and I. Niemelä. Solving parity games by a reduction to SAT. Journal of Computer and System Sciences, 78(2):430–440, 2012.
- J. Lahtinen, J. Valkonen, K. Björkman, J. Frits, I. Niemelä, and K. Heljanko. Model checking of safety-critical software in the nuclear engineering domain. Reliability Engineering & System Safety RESS, (105):104–113, 2012.
- M. Niemenmaa, A. Kallio, A. Schumacher, and Klemelä. Hadoop-BAM: Directly manipulating next generation sequencing data in the cloud. Bioinformatics, 28(6):876–877, 2012.

Mobile Computing,
Professor Sasu Tarkoma



The Mobile Computing 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.

Our central research items include synchronous and asynchronous middleware communications, publish/subscribe, data synchronization, context-aware systems, and more recently peer-to-peer content dissemination and delivery.

Key publications:

- Y. Cui, W. Hu, H. Wang, D. Chen, S. Tarkoma, and A. Yla-Jääski. Probabilistic multi-path routing for multimedia over wireless mesh networks. Ad-Hoc & Sensor Wireless Networks, 14(3-4):205–225, 2012.
- W. Rao, L. Chen, S. Chen, and S. Tarkoma. Evaluating continuous top-k queries over document streams. World Wide Web: Internet and Web Information Systems, pages 1–25, 2012.
- W. Rao, L. Chen, and S. Tarkoma. Towards efficient privacy-aware content-based pub/sub systems. IEEE Transactions on Knowledge and Data Engineering, PP(99):1, 2012.
- M. Komu, M.Sethi, R. Mallavarapu, H. Oirola, R. Khan, and S. Tarkoma. Secure networking for virtual machines in the cloud. In Proceedings of 2012 IEEE International Conference on Cluster Computing Workshops (Cluster Workshops 2012), pages 88–96, 2012.
- W. Rao, L. Chen, P. Hui, and S. Tarkoma. Move: A large scale keyword-based content filtering and dissemination system. In Proceedings of 2012 IEEE 32nd International Conference on Distributed Computing Systems (ICDCS 2012), pages 445– 454, 2012.
- W. Rao, K. Zhao, Y. Zhans, P. Hui, and S. Tarkoma. Maximizing timely content advertising in DTNs. In Proceedings of the 9th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON 2012), pages 254–262, 2012.

Collaborative Networking, Professor Jussi Kangasharju

We investigate novel network architectures, in particular in the areas of content distribution and content management. Our current primary focus is in information-centric networking, mobile opportunistic content sharing and green networking. Our work is experimentation-oriented and in our research we combine modeling, simulations, and practical experiments. In opportunistic content sharing, we have developed the floating content paradigm where information is published to a particular location and then replicated in an ad hoc manner between mobile devices, attempting to keep the information available in the area where it was published. Our work assumes no support from the infrastructure.

We have developed in-network caching architectures and devised caching mechanisms for information-centric networks. Our solutions cover simple, decentralized caching algorithms and their evaluation. In the area of green networking, we have continued our work on data center cooling optimization solutions, with work on cold aisle containment and its evaluation.

Currently we are extending this work to include heat harvesting and we have built a greenhouse where we are experimenting if the exhaust heat from computers is sufficient to heat up the greenhouse sufficiently to allow for plants to grow.

Key publications:

- P. Sarolahti, J. Ott, and J. Kangasharju. Locations vs. identities in internet content: Applying information-centric principles in today's networks. IEEE Communications Magazine, 50(12):54–59, 2012.
- O. Karkulahti and J. Kangasharju. Surveying wikipedia activity: Collaboration, commercialism, and culture. In Proceedings of the 2012 International Conference on Information Networking (ICOIN 2012), pages 384–389, 2012.
- J. Mineraud, S. Balasubramaniam, J. Kangasharju, and W. Donnelly. Fs-PGBR: a scalable and delay sensitive cloud routing protocol. In Proceedings of the ACM SIGCOMM 2012 conference on Applications, technologies, architectures, and protocols for computer communication (SIGCOMM'12), pages 301–302, 2012.
- M. Pervilä, M. Rantanen, and J. Kangasharju. Implementation and evaluation of a wired data center sensor network. In Proceedings of the First International Workshop on Energy Efficient Data Centers (E2DC 2012), Lecture Notes in Computer Science, pages 105–116. Springer Berlin Heidelberg, 2012.
- W. Wong, L. Wang, and J. Kangasharju. Neighborhood search and admission control in cooperative caching networks. In Proceedings of GLOBECOM 2012: Next Generation Networking and Internet Symposium, pages 2876–2882, 2012.

Future Internet (FI) Programme

Networking Research, Professor Andrei Gurtov

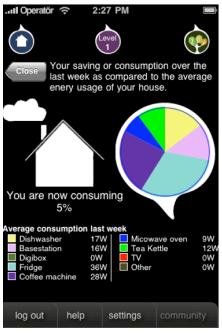
We study a wide range of communication problems, starting from the medium access control, traffic analysis, multicast and IP mobility up to hierarchical peer-to-peer architectures and distributed algorithms. Our particular focus is security aspects and game-theoretic approach to model selfish user behavior 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.

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.

Key publications:

- D. Korzun, A. Gurtov, Structured P2P Systems: Fundamentals of Hierarchical Organization, Routing, Scaling, and Security, Springer, ISBN 978-1-4614-5482-3, November 2012 (388 p).
- D. Kuptsov, B. Nechaev, A. Gurtov, Demand-Aware Flow Allocation in Data Center Networks, in Proc. of Capacity Sharing Workshop 2012 (CSWS'12), ACM CoNEXT'12, December 2012.
- M. Karl, T. Polishchuk, T. Herfet, A. Gurtov. Mediating multimedia traffic with strict delivery constraints. in Proc of IEEE International Symposium on Multimedia, December 2012.
- T. Polishchuk, M. Karl, T. Herfet, A. Gurtov, Scalable Architecture for Multimedia Multicast Internet Applications, in Proc. of IEEE WoWMoM'12, June 2012.
- J. Chuyko, T. Polishchuk, V. Mazalov, and A. Gurtov. Wardrop equilibria and price of anarchy in multipath routing games with elastic traffic. International Journal of Mathematics, Game Theory and Algebra, 20(4), 2012.

Network Society (NS) Programme

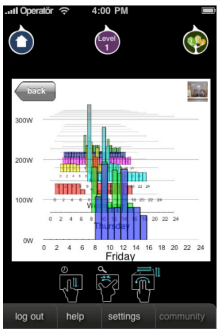


The mission of the Network Society research programme is to empower ubiquitous users with transparent and resourceful ICT. Example Challenges of the programme are listed in the sequel.

Maximum User Performance in Mobile Interaction: Numerous factors limit the mobile user's ability to interact efficiently with a mobile device. We are interested in exposing these factors through controlled experimentation and developing interaction techniques and methods that improve user performance.

Engaging and Embodied Interaction Across Modalities: Developing techniques and interactive systems utilising implicit and explicit interaction to foster co-experience, learning, physical and creative activity. This research is carried out by coupling in-depth user studies with design in the area of novel interfaces.

Service design, Media Experience and Media Crowdsourcing : theories and methodologies for user-centered service design, understanding of critical elements of media experience (e.g. playfulness and co-experience) through user research, laboratory experiments, and explorative designs. Media crowdsourcing: co-creativity in digital content creation, optimization of crowdsourcing tasks, creation of applications and games that utilise/produce metadata and media.



Analysing Actors and their Agencies in ICT environments, with special focus on infrastructures, use practices, stakeholders, business models, and other agencies shaping and affecting ICT technology. The goal of this analysis is to provide for socially sustainable design spaces, taking especially issues of privacy, societal tensions, values, and user involvement into account.

Virtual Economy and Empirical Law : to understand virtual value exchanges and to make a significant scientific and societal impact beyond virtual worlds on other services. Based on sound jurisprudential and empirical methods in co-operation with other disciplines make a significant scientific and societal impact on legal framework of ICT.

Emotional-Social Interaction: To understand emotional and cognitive processess during ICT-mediated social interaction. This research is carried out by using neurophysiological recordings, for example.

Digital Content Communities, Professor Marko Turpeinen

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 jurisprudencial 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:

Photo by Olli Pitkänen

- O. Pitkänen and H. Lehto. Legal aspects of living labs. International Journal of Product Development, 17(1):8–22, 2012.
- O. Pitkänen and V. Kristiina Tuunainen. Disclosing personal data socially - an empirical study on Facebook user's privacy awareness. Journal of Information Privacy & Security, 8(1):3–29, 2012.
- M. Salminen, N. Ravaja, K. Kallinen, and T. Saari. Mediated cues of group emotion during knowledge work tasks: Effects on subjective and physiological responses. Interacting with Computers, 2012.
- P. Virtanen. Claim clarity, full disclosure and software patents. Nordiskt Immateriellt Rättskydd, 2012(2):141–155, 2012.
- P. Virtanen. Football Dataco v Yahoo!: The ECJ interprets the database directive. SCRIPTed, 2012(2):258–267, 2012.
- H. Hietanen, A. Salovaara, K. Athukorala, and Y. Liu. <insertimage>: Helping the legal use of open images. In Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems (pp. 599-608). ACM, 2012.

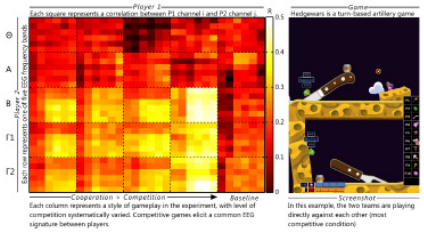
Social Interaction and Emotion (SIE), Professor Niklas Ravaja

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:

- G. Chanel, J. M. Kivikangas, and N. Ravaja. Physiological compliance for social gaming analysis: Cooperative versus competitive play. Interacting with Computers, 24(4):306–316, 2012.
- I. Ekman, G. Chanel, J. M. Kivikangas, S. Järvelä, M. Salminen, and N. Ravaja. Social interaction in games: Measuring physiological linkage and social presence. Simulation & Gaming, 43(2):321–338, 2012.
- J. Kätsyri, R. Hari, N. Ravaja, and L. Nummenmaa. The opponent matters: Elevated fMRI reward responses to winning against a human versus a computer opponent during interactive video game playing. Cerebral Cortex, 2012.
- J. Kätsyri, N. Ravaja, and M. Salminen. Aesthetic images modulate emotional responses to reading news messages on a small screen: A psychophysiological investigation. International Journal of Human-Computer Studies, 70(1):72–87, 2012.
- N. Ravaja, O.Somervuori, and M. Salminen. Predicting purchase decision: The role of hemispheric asymmetry over the frontal cortex. Journal of Neuroscience, Psychology, and Economics, 2012.
- S. Tamminen, E. Raita, V. Lehtinen, S. Silfverberg, and N. Ravaja. Teknologian sosiaalipsykologinen tutkimus. Psykologia, 47(5-6):410–422, 2012.
- S. Järvelä, I. Ekman, J. M. Kivikangas, and N. Ravaja. Digital games as experiment stimulus. In Proceedings of DiGRA Nordic 2012 Conference: Local and Global – Games in Culture and Society, 2012.
- P. Tikka, A. Väljamäe, A. de Borst, N. Ravaja, M. Kaipainen, and T. Takala. Enactive cinema paves way for understanding complex real-time social interaction in neuroimaging experiments. Frontiers in Human Neuroscience, 6, 1-6 (Article 298), 2012.



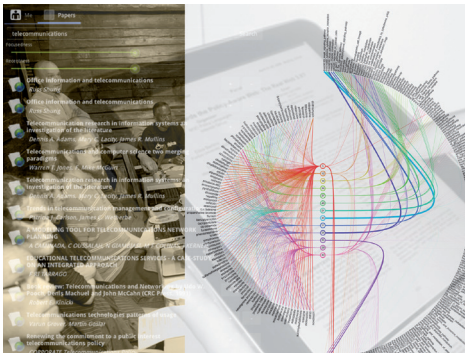
Ubiquitous Interaction (UiX), Professor Giulio Jacucci

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, adaptivite 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:

- E. Hoggan, C. Stewart, L. Haverinen, G. Jacucci, and V. Lant Pressages: Augmenting phone calls with non-verbal messages. In Proceedings of the 25th annual ACM symposium on User interface software and technology (UIST'12), 2012.
- A. Dünser, M. Billinghurst, J. Wen, V. Lehtinen, and A. Nurminen. Exploring the use of handheld ar for outdoor navigation Computers Graphics, 36(8):1095, 2012.
- A. Oulasvirta, A. Nurminen, and T. Suomalainen. How real is real enough? Optimal reality sampling for fast recognition of mobile imagery ACM Transactions on Applied Perception, 9(4):18, 2012.
- A. Oulasvirta, A. Pihlajamaa, J. P. Perkiö, D. Ray, T. Vähäkangas, T. Hasu, N. Vainio, and P. Myllymäki. Long-term effects of ubiquitous surveillance in the home. In Proceedings of the 14th International Conference on Ubiquitous Computing (Ubicomp 2012), pages 41–50, 2012.
- V. Lehtinen, A. Oulasvirta, A. Salovaara, and P. Nurmi. Dynamic tactile guidance for visual search tasks. In Proceedings of the ACM Symposium on User Interface Software and Technology (UIST 2012), pages 445–452, New York, NY, USA, 2012.
- L. A. Liikkanen. Musical activities predispose to involuntary musical imagery. Psychology of Music, 40(2):236–256, 2012.
- N. Corradi, K. Prifti, G. Jacucci, and L. Gamberini. Oops, I forgot the light on! The cognitive mechanisms supporting the execution of energy saving behaviors. Journal of Economic Psychology, Online first, http://dx.doi.org/10.1016/j.joep.2012.11.002.

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



In 2011 HIIT started a common spearhead endeavour with a significant investment of HIIT's internal resources, supported by external funding. We are developing next generation IT tools for research, especially data intensive research. The methods are based on HIIT's world-class research in computational modelling and data analysis as well as human-computer interaction, and on HIIT's wide interdisciplinary cooperation with application fields. The methods we are developing can be characterised as "informatics" of different sciences. The initial application areas are related to scientific documents and genomics data.

Augmented Research is an interdisciplinary research initiative in which researchers at HIIT combine their expertise in human-computer interaction, machine learning, information retrieval, and ubiquitous computing with a target to augment scientific information access using computation. We have collected one of the largest databases of scientific literature and related data in the world that can be utilized in research. We aim pushing research methods and results achieved in the focus area towards systems with real-world impact.

SciNet is an information access system for scientific literature that interactively directs users in their exploratory search tasks by predicting their search intents from interaction data. The system combines methods of optimized interactive data visualization, reinforcement learning based user modeling, and active relevance feedback for information retrieval.

To tackle the genomics challenge, we combine expertise from a number of HIIT groups working in algorithmics, machine learning, bioinformatics and computational systems biology. Together we are developing methods for accessing and utilising vast collections of molecular biology measurement data in biological and biomedical research. Thanks to new genome sequencing technologies, the sheer amount of molecular biology data is growing much faster than Moore's law rate of growth of computing power. In aggregate this data contains lots of useful information, but there is severe lack of tools to properly utilise it and hence make biology cumulative in terms of the collected data. We focus on developing methods for modelling and utilisation of gene expression and whole metagenome sequencing data sets.

Key publications:

- N. Välimäki and S. J. Puglisi. Distributed string mining for high-throughput sequencing data. In Proceedings of the 12th International Workshop on Algorithms in Bioinformatics (WABI 2012), Lecture Notes in Computer Science, pages 441–452. Springer Berlin Heidelberg, 2012.
- E. Georgii, J. Salojärvi, M. Brosche, J. Kangasjärvi, and S. Kaski. Targeted retrieval of gene expression measurements using regulatory models. Bioinformatics, 28(18):2349–2356, 2012.
- D. Głowacka, T. Ruotsalo, K. Konyushkova, K. Athukorala, S. Kaski, and G. Jacucci. Directing exploratory search: Reinforcement learning from user interactions with keywords. In Proceedings of IUI'13, International Conference on Intelligent User Interfaces, to appear. [Best Paper Award]

Administration

Personnel and funding

HIIT is a joint research institute of Aalto University and the University of Helsinki. 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, of which 37 per cent are of foreign origin) is much higher than the number of person-years.

In 2012 HIIT staff completed 144 person-years. The total funding of HIIT was 11.5 Million Eur, of which 83 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 2012 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. The members are appointed personal deputies. The staff of HIIT selects one board member and his or her deputy from among their colleagues. Board members are listed in the following, with their personal deputies in parentheses:

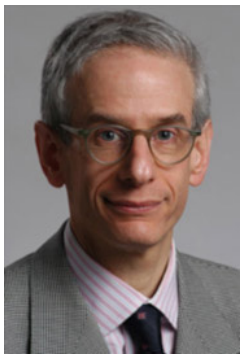
- Chairman: Until April 2012: Vice President, Professor Heikki Mannila, Aalto University. From May 2012 onwards: Vice President, Professor Ilkka Niemelä, Aalto University
  (Dean, Professor Outi Krause, Aalto University)
- Vice Chairman: Vice Rector, Professor Johanna Björkroth, University of Helsinki (Vice Rector, Professor Jukka Kola, University of Helsinki)
- Professor Olli Simula, Aalto University
  (Professor Heikki Saikkonen, Aalto University)
- Professor Hannu Toivonen, University of Helsinki
  (Professor Jukka Paakki, University of Helsinki)
- Vice President Henry Tirri, Nokia (Director Jyri Huopaniemi, Nokia)
- Docent Kari-Pekka Estola
  (Director Martin Mäklin, TeliaSonera Finland Oy)

- Director Petri Vasara, Pöyry Management Consulting Oy
  (Dr. Lars Gädda, Metsäklusteri Oy)
- MSc Kimmo Kiviluoto, Webmie Oy
  (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.

Scientific Advisory Board (SAB) meeting in June 2012

The SAB consists of internationally prominent scholars who are invited by the HIIT Board. The SAB convened 7-8 June 2012. The objective of the SAB meeting was to support and encourage HIIT in its activities, and to support the Board and the Director in developing HIIT and in revising and sharpening HIIT's strategic plans. SAB Members 2012:



- 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








An executive summary of the SAB's report is the following:

"The Scientific Advisory Board (SAB) finds that the Helsinki Institute for Information Technology (HIIT) remains a research institution of very high quality. The SAB is impressed with the energy and enthusiasm of its researchers, and the engagement of its postdocs and graduate students with their research. The research being undertaken is consistently rigorous, yet also exhibits excellent application relevance. New and effective international scientific collaborations are taking place in several fields. We applaud the Institute's success in addressing the recommendations made in the 2008 SAB report. For the mother universities, HIIT provides a high return on external funding from a surprisingly modest investment in the Institute, and the SAB strongly encourages the universities to increase the level of funding.

Our key summary recommendations are: (1) The mother universities should build on the successful institutional model of HIIT to encourage and expand collaborations both at HIIT and in other fields; (2) Raise the visibility of HIIT, and its sponsoring universities, within the world research community, leveraging on the world-class expertise on ICT issues at HIIT. (3) Build institutional initiatives to address the gender diversity issues in the field; (4) Make Augmented Science a driver for the integration of theory and applications with the Institute; (5) Exploit the Institute's competitive advantages at the intersection of large-scale data analysis and human-centered systems; and (6) Build an enhanced community among HIIT researchers."