

Helsinki Institute for Information Technology HIIT



Helsinki Institute for Information Technology HIIT

Annual Report 2011

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:
Helsinki Institute for Information Technology HIIT
PO Box 19215, FI-00076 Aalto, Finland

Street addresses:
Innopoly 2, Tekniikantie 14, Espoo
Computer Science Building, Konemiehentie 2, Espoo
Telephone: +358 9 47001
Fax: +348 9 694 9768

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 9 1911
Fax: +358 9 191 51120

Table of Contents

Review of Year 2011	5
What's new	5
Other key achievements in 2011	6
Collaboration	7
Societal impact	8
Doctoral studies	8
Views for 2012	9
Research	10
Algorithmic Data Analysis (ADA)	10
Algorithmic Systems (AS)	19
Future Internet (FI)	24
Network Society (NS)	29
HIIT-Wide Focus Area	33
Administration	34
Personnel and funding	34
Board	34
Scientific Advisory Board (SAB)	35

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 in 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 Centres of Excellence and co-operate with the information industry and with sciences applying information technology. HIIT is now in its 12th year of operation with a budget of about 10 million euros, realising in total about 135 person-years of work; a major part of these operations are done in direct collaboration within the departments of the mother universities. The research by HIIT has been assessed and deemed excellent by its Scientific Advisory Board in 2008, 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	2008	2009	2010	2011
Funding M€	8.6	9.0	12.1	10.5
External/competitive funding	76 %	71 %	83 %	78 %
Person-years	165	185	175	135
Refereed publications	209	184	223	182

Review of Year 2011

What's new

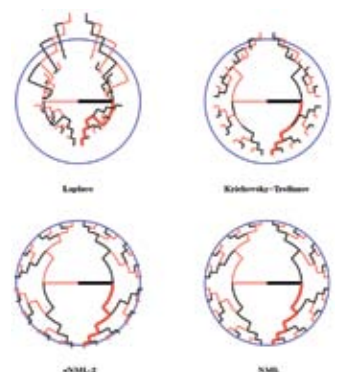
Research evaluation at UH

Research and doctoral training were evaluated in University of Helsinki in 2011. This time, researchers were allowed to volunteer for evaluation as bottom-up organized research communities. The research community of Algorithms and Machine Learning, which to a large degree overlaps with HIIT's ADA and CI programmes on the University of Helsinki side, scored absolute top of the whole university. The community participated in the "international cutting edge" category and received highest possible marks in every single aspect: *"The community constitutes one of the world-leading centres in the area of algorithmic and computational data analysis". "The community is excellently positioned, with an outstanding scientific record in which foundational and application-oriented research are both excellently represented."* The research community of Networks and Distributed Systems, also containing several HIIT groups, received the best scores in another category: *"The community is in an excellent stage of development and the overall focus is very strong. The subject area is highly relevant and well-positioned in the foundational and technical research for modern ICT systems."* *"The community has an excellent record of industrial cooperation and actively contributes to leading standardization bodies like the IETF, ISO, the WWRF forum and the W3C consortium."*



New Centres of Excellence starting

HIIT has been particularly successful regarding Centres of Excellence of the Academy of Finland. During the period 2012-2017 HIIT will contribute 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 Finnish Centre of Excellence in Algorithmic Data Analysis Research (Algodan, 2008-2013) forms the majority of HIIT's Algorithmic Data Analysis programme.



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 Science or Computationally-Supported Collective Science (CS)², 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 will focus on two case studies: general augmented research, and computational cumulative biology.

A new research group in social psychology

A new group named Social Interaction and Emotion, led by professor Niklas Ravaja from University of Helsinki, Department of Social Research, was established in 2011. Ravaja was nominated to hold the first professorship of social psychology of ICT in Finland as a joint professorship of University of Helsinki Department of Social Research and HIIT – a major achievement for HIIT in its own right. The group will study emotional and cognitive processes during mediated social interaction, in privacy management, and in entertainment and serious digital games.



Strategy documents

The year 2011 has involved heavy strategic planning at both mother universities, and writing of both long-term and short-term implementation plans for HIIT. The plans will be presented to HIIT's SAB in 2012.

Other key achievements in 2011



Several HIIT researchers again received Best Paper Awards in high-level scientific conferences: ACM CHI Best Paper Award 2011 and ACM Mobile-CHI Best Note Award to Antti Oulasvirta's group, IEEE Globecom 2011 Best Paper Award to Andrei Gurtov's group, ECML 2011 Best Paper Award to Samuel Kaski's group.

Professor Petteri Kaski, a group leader at HIIT, was awarded the 2011 Young Investigator's Award of the Finnish Foundation for the Advancement of Technology (Tekniikan edistämissäätiö). The award is granted annually to one young researcher who has in his or her research accomplished notable scientific results and/or created significant technological or technological innovations. The award covers all areas of technology.

HIIT researchers organized several international conferences in 2011: 22nd International Conference on Algorithmic Learning Theory (ALT), 14th International Conference on Discovery Science (DS), International Workshop on Innovative Database Research (IDAR), International Conference on Artificial Neural Networks (ICANN) and 10th International Symposium on Intelligent Data Analysis (IDA).

HIIT has traditionally been successful in turning research results into new start-up companies and commercial products. Autumn 2011 saw the establishment of Screen.io which provides platforms and services for participative events.

Collaboration

Internationality is an integrated part of all operations. Cooperation with key international research institutes and universities is active: Berkeley (ICSI and UC Berkeley), Tsinghua University, Waseda 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.

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. For example Dr Antti Oulasvirta, a junior group leader at HIIT, got a senior group leader position in Max Planck Institute for Informatics in Saarbrücken, Germany in 2011. 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.

Henry Tirri, former group leader at HIIT, started as the CTO of Nokia in 2011.

Projects supported by the Academy of Finland and Tekes are important for HIIT. HIIT groups belong to four Centres of Excellence of the Academy of Finland. HIIT is a strategic partner of Tekes and participates in Tekes's discussions on the renewal of funding instruments.

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. To give an example of impact from the Future Internet Programme, several of the techniques developed on malware detection and classification with industrial partners F-Secure (main partner), Stonesoft and Nokia have since then been integrated into the malware analysis tools in production use at F-Secure. 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 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. One of the main HIIT activities at EIT ICT Labs is PI3: Pervasive Information, Interaction, & Interfaces.

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. Professor Giulio Jacucci coordinates a EU STREP Project on Energy Awareness with 8 international partners. HIIT collaborates in the MARCUS Exchange (Mobile Augmented Reality and Context in Urban Settings) project through the IRSES, part of FP7 Marie Curie Actions, with TU Graz, Fraunhofer FIT, University of Otago, University of Canterbury, and HIT Lab NZ.

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. It 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 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 now led by HIIT programme director, professor Petri Myllymäki, and was previously led by professor Hannu Toivonen of HIIT.

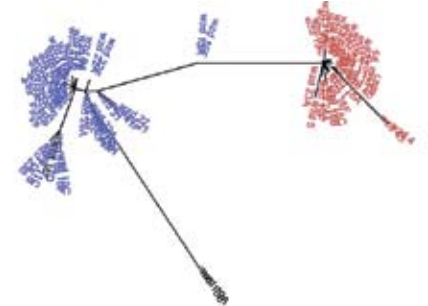
HIIT also contributed strongly in establishing the EIT ICT Labs Doctoral School, and will remain a partner in the Helsinki node of the school.

7 HIIT researchers got their PhD during the year.

Views for 2012

Programme renewal

The research programmes of HIIT will be slightly rearranged in early 2012, to better align with Centres of Excellence of the Academy of Finland, and to reflect new areas of investigation. A new programme Scalable Networks will have two spearhead activities: a holistic view on energy efficiency, and new programming paradigms for massively scalable networks.



SAB

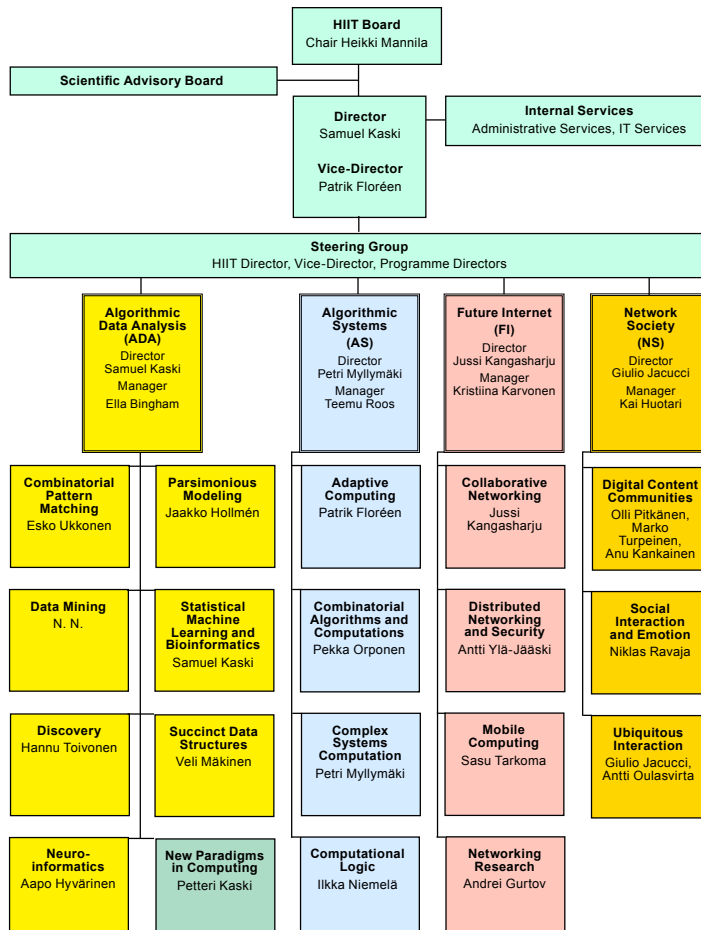
The Scientific Advisory Board of HIIT will convene in June 2012 to assess recent changes and future directions of HIIT.

Open Innovation House

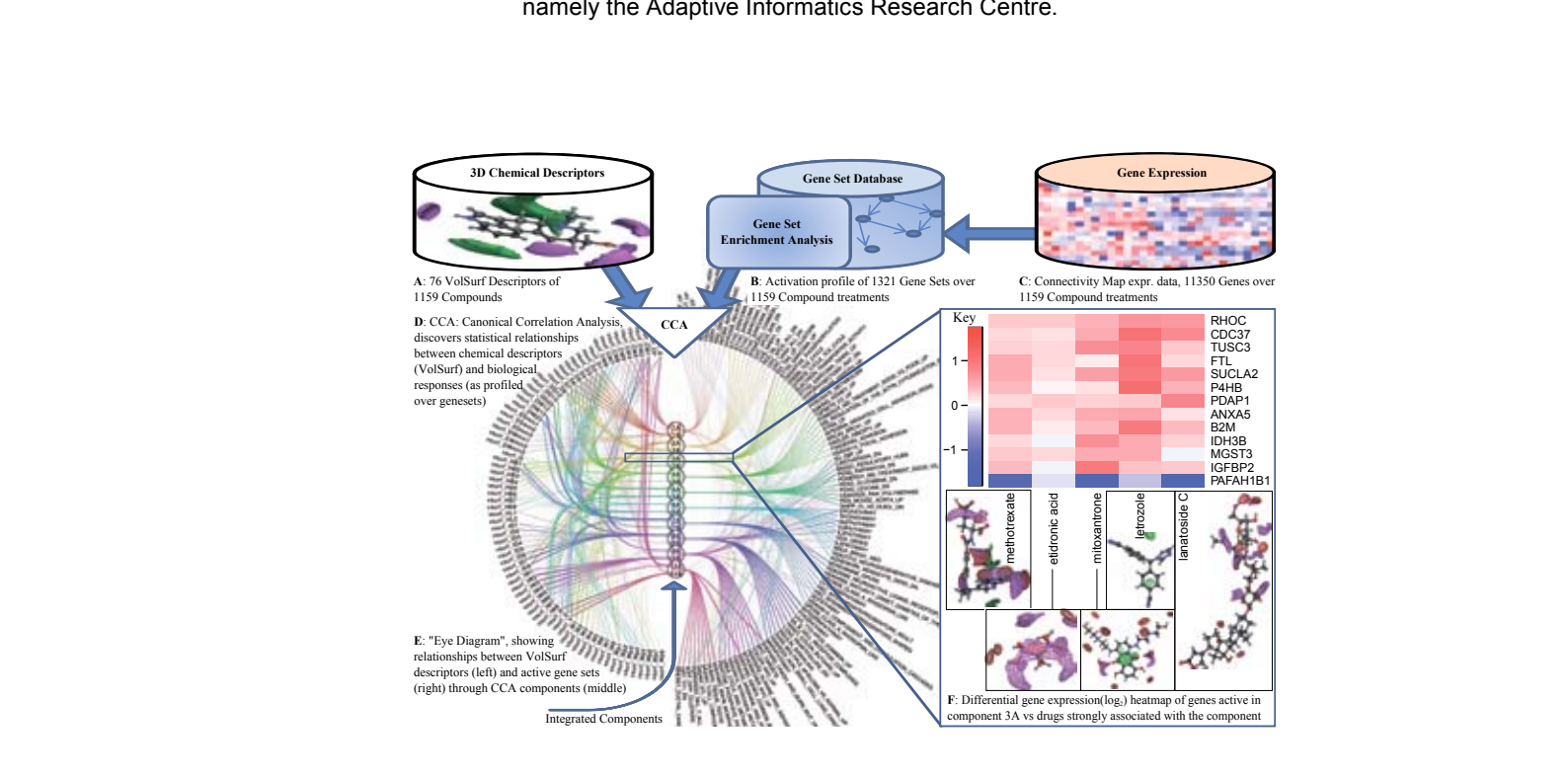
In summer 2012 a welcomed change will take place as HIIT researchers located at Innopoli business park will move to new premises more centrally located at Aalto University's Otaniemi campus, shared with EIT ICT Labs Helsinki node's premises.



Helsinki Institute for Information Technology HIIT



Research

[illegible]

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.

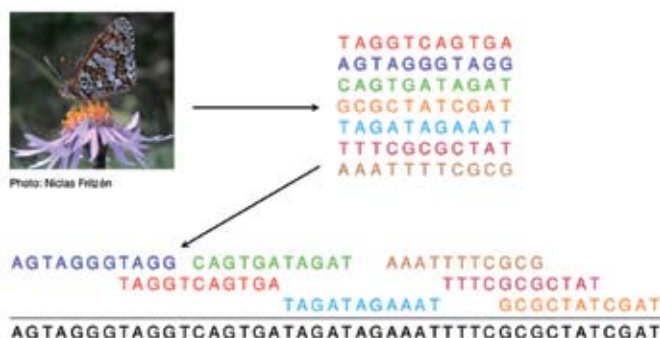
Recently we have focused in developing algorithms for assembling genomes from short read data. We have proposed new algorithms for correcting sequencing errors in the data. We have also developed new methods for the scaffolding phase where contiguous sequences assembled from the short read data are arranged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

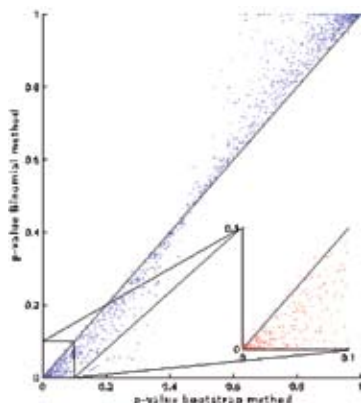
Another recent research focus is regulatory regions in genomes. We have developed new models based high throughput data for describing regulatory elements and we have proposed new algorithms for recognizing such elements in genomic sequences.



Key publications:

- L. Salmela, V. Mäkinen, N. Välimäki, J. Ylinen and E. Ukkonen. Fast scaffolding with small independent mixed integer programs. *Bioinformatics*, 27(23):3259–3265, 2011.
- L. Salmela and J. Schröder. Correcting errors in short reads by multiple alignments. *Bioinformatics*, 27(11):1455–1461, 2011.
- C. Pizzi, P. Rastas, and E. Ukkonen. Finding significant matches of position weight matrices in linear time. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 8(1):69–79, 2011.
- G.-H. Wei, G. Badis, M. F. Berger, T. Kivioja, K. Palin, M. Enge, M. Bonke, A. Jolma, M. Varjosalo, A. R. Gehrke, J. Yan, S. Talukder, M. Turunen, M. Taipale, H. G. Stunnenberg, E. Ukkonen, T. R. Hughes, M. L. Bulyk and J. Taipale. Genome-wide analysis of ETS-family DNA-binding in vitro and in vivo. *The EMBO Journal*, 29:2147–2160, 2010.





Data Mining, Docent Kai Puolamäki (Acting group leader)

The Data Mining: Theory and Applications group conducts research on finding local patterns and global models in discrete high-dimensional data. Techniques for this task include both algorithmics in the traditional computer science sense and probabilistic methods. The group was founded by Professor Heikki Mannila who was later appointed Vice-President of Aalto and then President of the Academy of Finland, and who still contributes to doctoral student supervision in the group.

The highlights of the group's research activities in 2011 include our continuing work on randomization methods and statistical significance testing in data analysis. We have applied the randomization and statistical testing methodology especially in the analysis of sequences.

As an example of the analysis of (word) sequences, we have shown that even the simplest statistics such as the word frequency in the text may be uninformative, if typical assumptions such as bag of words are used. This is part of our focus on computational linguistics in collaboration with the group of Professor Terttu Nevalainen.

In collaboration with the group of Professor Mikael Fortelius we have continued studying computational methodologies in ecology and paleontology.

Key publications:

- A. Kallio, N. Vuokko, M. Ojala, N. Haiminen, and H. Mannila. Randomization techniques for assessing the significance of gene periodicity results. *BMC Bioinformatics*, 12:330, 2011.
- P. Papapetrou, V. Athitsos, M. Potamias, G. Kollios, and D. Gunopulos. Embedding-based subsequence matching in time series databases. *ACM Transactions on Database Systems (TODS)*, 36(3):39, 2011.
- P. Papapetrou, G. Benson, and G. Kollios. Mining Poly-regions in DNA Sequences. *International Journal of Data Mining and Bioinformatics*, 2012. To appear.
- T. Nevalainen, H. Raumolin-Brunberg, and H. Mannila. The diffusion of language change in real time: Progressive and conservative individuals and the time depth of change. *Language Variation and Change*, 23:1–43, 2011.
- J. Lijffijt, P. Papapetrou, K. Puolamäki, and H. Mannila. Analyzing word frequencies in large text corpora using inter-arrival times and bootstrapping. In *Proceedings of the European Conference of Machine Learning and Knowledge Discovery in Databases – Part II*, pages 341–357. Springer Berlin, 2011.
- G. C. Garriga, E. Junttila, and H. Mannila. Banded structure in binary matrices. *Knowledge and Information Systems*, 28(1):197–226, 2011.
- A. Kallio, K. Puolamäki, M. Fortelius, and H. Mannila. Correlations and co-occurrences of taxa: the role of temporal, geographic, and taxonomic restrictions. *Palaeontologia Electronica*, 14(1):4A, 2011.

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

The group's focus is on algorithmic methods for discovering links and patterns in data, and recently also on their use in creative systems.

A methodological focus area is analysis and exploration methods for weighted (biological) graphs. 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.

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

We have also continued international collaboration in data mining research, on bioinformatics and database integrity constraints.



Key publications

- H. Toivonen, F. Zhou, A. Hartikainen, and A. Hinkka. Compression of weighted graphs. In *Proceedings of the 17th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, pages 965–973, 2011.
- V. Podpecan, N. Lavrac, I. Mozetic, P. Kralj Novak, I. Trajkovski, L. Langohr, K. Kulovesi, H. Toivonen, M. Petek, H. Motani, and K. Gruden. Seg-Mine workflows for semantic microarray data analysis in Orange4WS. *BMC Bioinformatics*, 12(416), 2011.
- A. Valitutti. Ambiguous lexical resources for computational humour generation. In *International Conference on Agents and Artificial Intelligence (ICAART)*, 2012.
- L. Langohr and H. Toivonen. A model for mining relevant and non-redundant information, In *ACM Symposium on Applied Computing (SAC)*, 2012.
- L. Eronen and H. Toivonen. Biomine: Predicting links between biological entities using network models of heterogeneous database. *BMC Bioinformatics*, accepted for publication.
- J. Li, J. Liu, H. Toivonen, J. Yong. Effective pruning for the discovery of conditional functionaldependencies. *The Computer Journal*, accepted for publication.

*Elot sai karkelojen teitä,
lumi ajan kotia,
hiljaa soi kodit autiot,
hiljaa sai armaat karkelot -
laiho sai lumien riemut.*

*Lives got the frolic ways,
snow the home of time,
softly chimed abandoned homes,
softly got frolics beloved -
ripening crop got the snows' joys.*



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 seminal method for testing independent components (Hyvärinen, 2011). Independent component analysis is widely used for scientific data analysis, for example, in neuroimaging. However, methods for testing the statistical significance of the components have been missing so far.

We have also 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:

- A. Hyvärinen. Testing the ICA mixing matrix based on inter-subject or inter-session consistency. *NeuroImage*, 58:122–136, 2011.
- S. Shimizu, T. Inazumi, Y. Sogawa, A. Hyvärinen, Y. Kawahara, T. Washio, P.O. Hoyer, and K. Bollen. DirectLiNGAM: A direct method for learning a linear non-gaussian structural equation model. *Journal of Machine Learning Research*, 12:1225–1248, 2011.
- M. 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.
- 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*, in press.

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:

- M. A. Prada, J. Toivola, J. Kullaa, and J. Hollmén. Three-way analysis of structural health monitoring data, *Neurocomputing*, (80):119–128, March, 2012.
- A. Kotsifakos, P. Papapetrou, J. Hollmén, and D. Gunopulos. A subsequence matching with gaps-range-tolerances framework: A query-by-humming application. In *Proceedings of the Very Large Database Endowment (PVLDB)*, (4)11:761–771, August 2011.
- M. Korpela, P. Nöjd, J. Hollmén, H. Mäkinen, M. Sulkava, P. Hari. Photosynthesis, temperature and radial growth of Scots Pine in northern Finland: identifying the influential time intervals. *Trees – Structure and Function*, 25(2):323–332, April, 2011.
- O. Kostakis, P. Papapetrou, and J. Hollmén. ARTEMIS: Assessing the similarity of event-interval sequences. In *Proceedings of The European Conference of Machine Learning and Principles and Practices of Knowledge Discovery in Databases (ECML/PKDD)*, Volume 6912 of *Lecture Notes in Computer Science*, pages 229–244, Springer, September 2011.
- A. Usvasalo, R. Rätty, A. Harila-Saari, P. Koistinen, E.-R. Savolainen, S. Knuutila, E. Elonen, U. M. Saarinen-Pihkala, J. Hollmén. Prognostic classification of patients with acute lymphoblastic leukemia by using gene copy number profiles identified from array-based comparative genomic hybridization data. *Leukemia Research*, 34(11):1476–1482, November, 2010.



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.

The group has recently developed Bayesian methodology for modeling dependencies between multiple data sets with co-occurring samples, in the form of latent variable models that factorize the variation in a data set collection into common and data set-specific components. These models have enabled for example decoding natural continuous speech from single-trial MEG recordings (Koskinen et al., Human Brain Mapping, accepted for publication), translating metabolomics findings between model organisms and men (Systi-Aho et al, PLoS Computational Biology, 2011), and analyzing the impact of chemoinformatic structure of drugs on the molecular response profiles of cells (Khan et al, BMC Bioinformatics, accepted for publication).

Furthermore, we work on information visualization and retrieval. We have recently formulated data visualization as a rigorous information retrieval problem, introduced a data glasses and eye tracking-based real-world contextual information access prototype, and developed model-driven methods for retrieving earlier experimental data. A main application goal is to help make molecular biology more cumulative by making possible retrieval of relevant experiments (REx) from the large public gene expression dataset repositories.



Key publications:

- A. Ajanki, M. Billinghamurst, H. Gamper, T. Järvenpää, M. Kandemir, S. Kaski, M. Koskela, M. Kurimo, J. Laaksonen, K. Puolamäki, T. Ruokolainen and T. Tossavainen. An augmented reality interface to contextual information. *Virtual Reality*, 15:161–173, 2011.
- 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):i246–i253, 2012.
- P. Glaus, A. Honkela, M. Rattray. Identifying differentially expressed transcripts from RNA-seq data with biological variation. doi: 10.1093/bioinformatics/bts260. *Bioinformatics*, 2012.
- S. Kaski and J. Peltonen. Dimensionality reduction for data visualization. *IEEE Signal Processing Magazine*, 28(2):100–104, 2011.
- S. Virtanen, A. Klami, S. Kaski. Bayesian CCA via group sparsity. In *Proceedings of the 28th International Conference on Machine Learning (ICML)*, pages 457–464, ACM, 2011.

Succinct Data Structures, Professor Veli Mäkinen

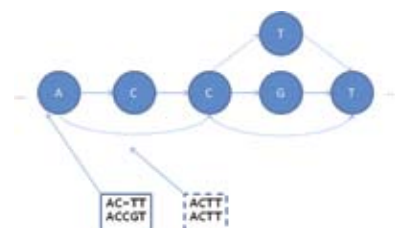
The study of succinct data structures extends traditional data compression with the functionality preserving property: data structure functions need to be efficiently computable directly from the compressed representation. Applications are sought mainly in bioinformatics, especially to the new challenges set by next-generation sequencing technologies.

Our recent developments include an extension of Burrows-Wheeler transform to finite automaton representing reference genome together with its common variations among the population (WABI 2011). This enables a space-efficient index structure to be constructed to support efficient DNA sequencing read alignment to a rich model of the population (see attached image: Read ACTT needs two edits to align to reference (backbone), while it matches exactly a variant represented by the automaton).

Optimized sampling schemes for compressed suffix arrays were presented (ESA 2011). We also worked on a space-efficient index structure for computing approximate overlaps between DNA reads to form overlap graphs for de novo fragment assembly. Additionally we collaborated to implement a fully functional compressed representation for XML documents. We have recently added the support for XML documents representing a genome annotation database, enabling queries by annotation restrictions (e.g. organism type, gene function, promoter, etc.) and sequence content (PWM matrix and approximate search support).

Key publications:

- J. Sirén, N. Välimäki, and V. Mäkinen. Indexing finite language representation of population genotypes. In *Proceedings of the 11th International Workshop on Algorithms in Bioinformatics (WABI)*, pages 270–281, Springer, 2011.
- P. Ferragina, J. Sirén, and R. Venturini. Distribution-aware compressed full-text indexes. In *Algorithms - ESA 2011 - 19th Annual European Symposium*, pages 760–771, Springer, 2011.
- N. Välimäki, S. Ladra, and V. Mäkinen. Approximate all-pairs suffix/prefix overlaps. In *Proc. 21st Annual Symposium on Combinatorial Pattern Matching (CPM 2010)*, pages 76–87, Springer, 2010.
- D. Arroyuelo, F. Claude, S. Maneth, V. Mäkinen, G. Navarro, K. Nguyen, J. Sirén, and N. Välimäki. Fast in-memory XPath search using compressed indexes. In *Proc. 26th IEEE International Conference on Data Engineering (ICDE 2010)*, pages 417–428, IEEE, 2010.
- V. Mäkinen, G. Navarro, J. Sirén, and N. Välimäki. Storage and retrieval of highly repetitive sequence collections. *Journal of Computational Biology*, 17(3):281–308, 2010.





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.

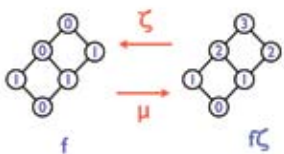
Among key highlights in 2011, our theme of pursuing interactions of algebraic, combinatorial and geometric techniques in algorithm design yielded a breakthrough on the algorithmics of Möbius inversion on finite lattices. We showed that every lattice with v elements, n of which are nonzero and join-irreducible, has FFT-like arithmetic circuits of size $O(vn)$ for computing the zeta transform and its inverse, thus enabling fast multiplication in the Möbius algebra. A second highlight is the nonconstructive enumeration of the main classes of Latin squares of order 11.

Under the theme of restricted models of computation and local algorithms, we introduced the notion of locally checkable proofs to study decision problems from the perspective of nondeterministic distributed algorithms. We classified graph problems according to their local proof complexity, whereby a natural hierarchy emerges: for many classical graph problems, the proof complexity is either 0, $\Theta(1)$, $\Theta(\log n)$, or $\text{poly}(n)$ bits per node.

As a novel technique for algorithmic data analysis, we introduced a new partial-order based MCMC method for Bayesian network discovery, improving both the mixing and convergence. For parameterized variants of so-called polytrees, we have shown tractability and intractability results. A further highlight considers localization, clustering, and activity scheduling in wireless sensor networks: we developed a combined approach to the three problems, applying the "Occam's razor principle" to data collected by the sensors in both centralized and distributed settings.

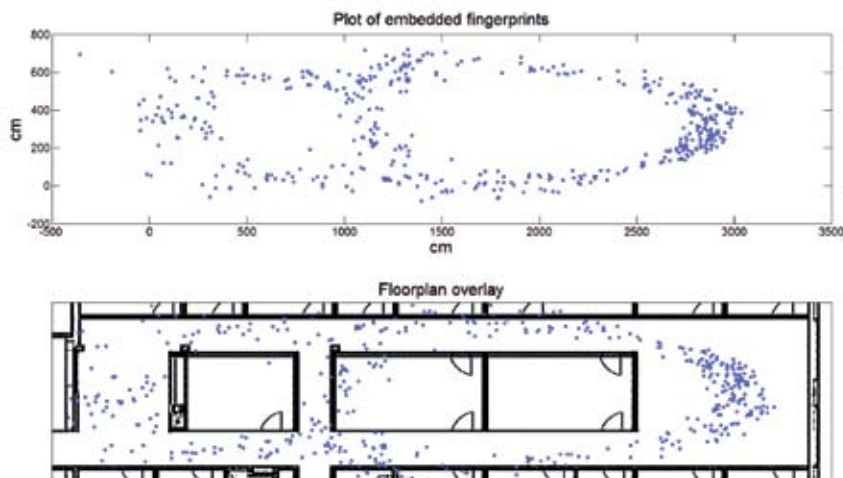
Key publications:

- P. Agarwal, A. Efrat, C. Gniady, J. S. B. Mitchell, V. Polishchuk, and G. Sabhnani. Distributed localization and clustering using data correlation and the Occam's razor principle. In *Proc. DCOSS*, 2011.
- A. Björklund, T. Husfeldt, P. Kaski, M. Koivisto, J. Nederlof, and P. Parviainen. Fast zeta transforms for lattices with few irreducibles. In *Proc. SODA*, 2012.
- S. Gaspers, M. Koivisto, M. Liedloff, S. Ordyniak, and S. Szeider. On finding optimal polytrees. In *Proc. AAAI*, 2012.
- M. Göös and J. Suomela. Locally checkable proofs. In *Proc. PODC*, pages 159–168, 2011.
- A. Hulpke, P. Kaski, and P. R. J. Östergård. The number of Latin squares of order 11. *Mathematics of Computation*, 80:1197–1219, 2011.
- T. Niinimäki, P. Parviainen, and M. Koivisto. Partial order MCMC for structure discovery in Bayesian networks. In *Proc. UAI*, pages 557–565, 2011.
- J. Suomela. Survey of local algorithms. *ACM Computing Surveys*, to appear.



Algorithmic Systems (AS) Programme

The Algorithmic Systems research programme aims at finding algorithmic solutions for science and society, and with this target in mind, the goal is to study the theory and practice of modelling, designing and managing complex systems. The work has a strong basic research component that intersects artificial intelligence, machine learning, theoretical computer science, information theory and mathematical statistics, and the results of this methodological work are applied to both scientific and industrial applications. An example of the work motivated by industrial applications concerns adaptive interfaces for consumer applications, and we have, together with our industrial partners, set up in a large store in Helsinki a test environment involving indoor positioning, user tracking and personalised feedback. In this real-world environment we have studied for example how location-aware and personalised advertisements can be produced and how people react to them, how customers move around in the shop, and how haptic feedback can be used for finding products from a shelf.



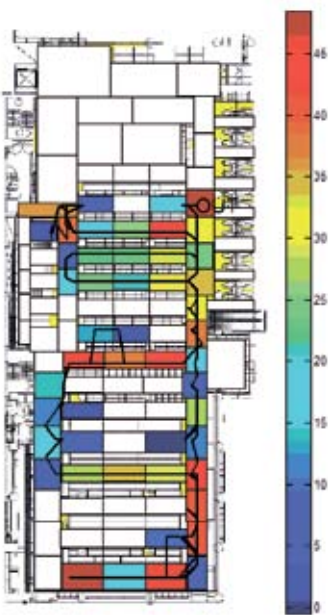


Adaptive Computing, Docent Patrik Floréen

The group works on ubiquitous computing, user modelling and interaction with a view of making the life of ordinary people easier through easy-to-use digital services. A central theme is positioning and location-awareness. Location is one of the most important dimensions of context and hence information about the user's location is essential for ubiquitous computing solutions. We have studied both indoor positioning using WiFi as well as outdoor positioning and trajectory tracking. Location-awareness is also central in the group's work on adaptive interfaces for consumer applications. The goal of this project with several companies is to enable adaptive interfaces to information by combining context-awareness with personalisation and to perform user acceptance studies on a number of prototypes for mobile consumer applications.

The multidisciplinary VISCI consortium, consisting of a global user company, innovative service developer and tester organizations, and four research teams with different scientific backgrounds, has developed prototypes for novel ICT-enabled tools and accompanying processes that increase the efficiency of virtual collaborative innovation.

The group is active in the EIT ICT Labs in the Action Line Smart Spaces. In collaboration with the German Research Center for Artificial Intelligence (DFKI) in Saarbrücken, our group has performed a large real-world deployment of a novel WiFi positioning system into a section of a hypermarket in Saarbrücken.



Key publications:

- M. B. Kjaergaard, S. Bhattacharya, H. Blunck, P. Nurmi. Energy-efficient trajectory tracking for mobile devices. In *Proc. 9th International Conference on Mobile Systems, Applications and Services (MobiSys)*, pages 307–320, 2011.
- P. Nurmi, S. Bhattacharya, J. Kukkonen. A grid-based algorithm for on-device GSM positioning. In *Proc. 12th ACM International Conference on Ubiquitous Computing (Ubicomp)*, pages 227–236, ACM, 2010.
- P. Nurmi, A. Salovaara, S. Bhattacharya, T. Pulkkinen, G. Kahl. Influence of landmark-based navigation instructions on user attention in indoor smart spaces. In *Proc. 2011 Conference on Intelligent User Interfaces (IUI)*, pages 33–42, 2011.
- T. Pulkkinen, P. Nurmi. AWESOM: Automatic discrete partitioning of indoor spaces for WiFi fingerprinting. In *Proc. 10th International Conference on Pervasive Computing (Pervasive)*, 2012.
- T. Pulkkinen, T. Roos, P. Myllymäki. Semi-supervised learning for WLAN positioning. In *Proc. International Conference on Artificial Neural Networks (ICANN)*, pages 355–362, LNCS 6792, Springer, 2011.
- T. Vähäkangas, J. Pyykkö. VisciPad: Peeking into a collaborative creative writing project in elementary school. In *Proc. 10th International Conference on Creating, Connecting and Collaborating through Computing (C5)*, 2012.

Combinatorial Algorithms and Computation, Professor Pekka Orponen

The mission of the Combinatorial Algorithms and Computation group is broadly speaking the development of efficient algorithmic methods for the interests arising from novel application areas, in close collaboration with domain experts. Here the group has had a long-standing collaboration with Doc. Patrik Floréen and his team in the area of distributed algorithmics, especially in the context of sensor networks. This work has over the past few years been extended in the direction of online optimisation, where research has been pursued since 2009 jointly with academic and industrial partners from the area of industrial production planning. Another domain of industrial collaboration, also since 2009, have been the algorithmic and data-analytic problems arising in data security applications. However as the latter work shifted in 2011 more clearly to the data-analytic side, it was handed over to another group where the match to research focus was better.

The group has also been performing, for a number of years, fundamental algorithmic research in the area of local search methods for complex problems, in collaboration with a team of statistical physicists. A recent fascinating application area, with far-reaching potential, is the synthesis of tile sets for patterned DNA self-assembly. Here the group has been considering the abstract problem of finding, for a given 2D pattern of colored “tiles”, a minimal set of tile types self-assembling to this pattern in the so called Abstract Tile Assembly Model of Winfree et al. This problem has applications in e.g. the design of DNA-origami based assembly scaffolds for functional materials. In addition to the simple minimal tile set synthesis problem, also the issues of reliability of the synthesized tile sets and their size/reliability tradeoffs, have been considered.



Key publications:

- M. Alava, J. Ardelius, E. Aurell, P. Kaski, S. Krishnamurthy, P. Orponen, and S. Seitz. Circumspect descent prevails in solving random constraint satisfaction problems. In *Proceedings of the National Academy of Sciences (USA)* 105:15253-15257, 2008.
- T. Lempiäinen, E. Czeizler, P. Orponen. Synthesizing small and reliable tile sets for patterned DNA self-assembly. In *Proc. DNA17: 17th International Conference on DNA Computing and Molecular Programming* (Pasadena), 145–159. Springer, Berlin, 2011.
- T. Erlebach, S. E. Nikolettseas, P. Orponen. *Algorithms for Sensor Systems. 7th International Symposium on Algorithms for Sensor Systems, Wireless Ad Hoc Networks and Autonomous Mobile Entities, ALGOSENSORS 2011*, Saarbrücken, Germany, September 8-9, 2011, Revised Selected Papers, Springer, 2012.
- H. Haanpää, A. Schumacher, P. Orponen. Distributed algorithms for lifetime maximization in sensor networks via Min-Max spanning subgraphs. *Wireless Networks*, 16(3): 875–887, 2010.

Complex Systems Computation (CoSCo), Professor Petri Myllymäki



The CoSCo research group studies computational problems in complex systems, especially regarding issues related to machine learning, model selection and statistical inference. The main research areas of the group include Bayesian networks and other probabilistic graphical models, the Minimum Description Length (MDL) principle and other information-theoretical approaches to inference, and clustering and visualization methods for multi-dimensional data.

A recent area of interest is stemmatology research: given a collection of imperfect copies of a textual document, is to reconstruct the history of the text, indicating for each variant the source text from which it was copied. In addition to our cross-disciplinary research activities in this area, we have been running *Studia Stematologica*, a series of six international workshops on stemmatology.

In our basic research work on probabilistic modeling, we have continued our pioneering work on the Minimum Description Length (MDL) principle together with Jorma Rissanen, the originator of MDL. Recently we have focused on finding more efficient forms of MDL for problems related to model selection and on-line inference. We have also developed similar methodologies for discriminative learning of probabilistic models.

Another recent highlight concerns intelligent information retrieval and tools for virtual collaborative innovation (joint work with the Adaptive Computing research group).



Key publications

- A. Carvalho, T. Roos, A. Oliveira, and P. Myllymäki. Discriminative learning of Bayesian networks via factorized conditional log-likelihood. *Journal of Machine Learning Research*, 12:2181–2210, 2011.
- T. Roos and Y. Zou. Analysis of textual variation by latent tree structures, In *Proceedings of the 2011 ICDM IEEE International Conference on Data Mining*, pages 567–576, IEEE, 2011.
- J. Rissanen. *Optimal Estimation of Parameters*. Cambridge University Press, 2012 (to appear).
- T. Silander, T. Roos, and P. Myllymäki. Learning locally minimax optimal Bayesian networks. *International Journal of Approximate Reasoning*, 51(5):544–557, June 2010.
- J. Rissanen, T. Roos, and P. Myllymäki. Model selection by sequentially normalized least squares. *Journal of Multivariate Analysis*, 101(4):839–849, April 2010.

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 (SMT), and rule-based constraints. 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 system. In 2011, new model checking techniques were developed for IEEE Property Specification Language.

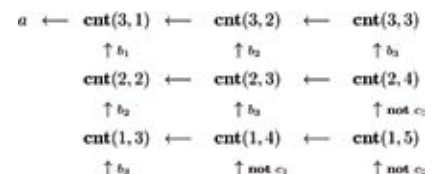
In 2011 techniques were developed for solving challenging SAT problems using computational grids and for the purpose has studied methods for partitioning search spaces. As regards solver development, the group developed novel techniques for parity reasoning based on equivalence classes.

In 2011 the group also studied the interconnections of different logical formalisms, enabling the transfer of concepts and methods. Concerning strong equivalence, originally proposed for answer-set programs, it was shown how an analogous notion can be applied in the context of abstract argumentation frameworks. Moreover, series transformations were developed to effectuate the reduction from non-disjunctive answer-set programs to propositional satisfiability and allow the use of SAT solvers for answer set computation.



Key publications:

- T. Launiainen, K. Heljanko, and T. Junttila. Efficient model checking of PSL safety properties. *IET Computers & Digital Techniques*, 5(6):479–492, November 2011.
- T. Janhunen, I. Niemelä, J. Oetsch, J. Pührer, and H. Tompits. Random vs. structure-based testing of answer-set programs: An experimental comparison. In *Proceedings of the 11th International Conference on Logic Programming and Nonmonotonic Reasoning*, pages 242–247, Springer, LNCS 6645, 2011.
- A. E. J. Hyvärinen, T. Junttila, and I. Niemelä. Partitioning search spaces of a randomized search. *Fundamenta Informaticae*, 107(2–3):289–311, 2011.
- T. Laitinen, T. Junttila, and I. Niemelä. Equivalence class based parity reasoning with DPLL(XOR). In *Proceedings of the 23rd IEEE International Conference on Tools with Artificial Intelligence*, pages 649–658 IEEE Computer Society Press, 2011.
- E. Oikarinen and S. Woltran. Characterizing strong equivalence for argumentation frameworks. *Artificial Intelligence*, 175(14–15):1985–2009, 2011.
- T. Janhunen and I. Niemelä. Compact translations of non-disjunctive answer set programs to propositional clauses. In *Proceedings of Symposium on Constructive Mathematics and Computer Science*, pages 111–130, Springer, LNCS 6565, 2011.



Future Internet (FI) Programme

The vision of the Future Internet research programme is that the Internet will become a mission critical backbone of global information society with billions of mobile and wire line users instantly connected to information and each other, and using the Internet to communicate, conduct business, manage their everyday lives, express themselves, and enjoy entertainment.

The objective is to develop concepts, technologies, and supporting theories and methodologies needed to design and implement future computing services and products for the Internet of the future.



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:

- J. Ott, E. Hyytiä, P. Lassila, T. Vaegs, J. Kangasharju, S. Santra. Floating content for probabilistic information sharing. *Pervasive and Mobile Computing*, 7(6):671–689, December 2011.
- M. Pervilä, J. Kangasharju. Cold air containment. In *Proceedings of ACM SIGCOMM Workshop on Green Networking and Smart Systems*, pages 7–12, 2011.
- W. Wong, M. V. L. Giraldo, M. F. Magalhães, J. Kangasharju. Content routers. Fetching data on network path. In *Proceedings of 2011 IEEE International Conference on Communications (ICC)*, 2011.
- E. Hyytiä, J. Virtamo, P. Lassila, J. Kangasharju, J. Ott. When does content float? Characterizing availability of anchored information in opportunistic content sharing. In *Proceedings of IEEE Infocom*, pages 3137–3145, 2011.
- J. Ott, E. Hyytiä, P. Lassila, T. Vaegs, J. Kangasharju. Floating content: Information sharing in urban areas. In *Proceedings of IEEE International Conference on Pervasive Computing and Communications*, pages 136–146, 2011.
- M. Pervilä, J. Kangasharju. Running Servers around Zero Degrees. *Computer Communications Review*, 41(1):96–101, January 2011.



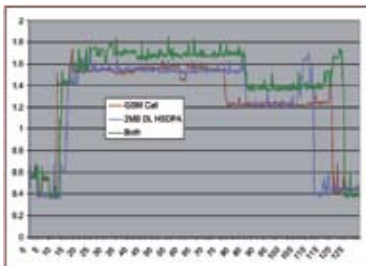
Distributed Networking and Security, Professor Antti Ylä-Jääski



Energy-aware computing and communications is a very timely and important topic. The capabilities of mobile devices have improved rapidly over the last ten years transforming also the way these devices are being utilized – creating 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. 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.

In distributed systems and services research we 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. 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:



- S. Saha, A. Lukyanenko, and A. Ylä-Jääski. CombiHeader: Minimizing the number of shim headers in redundancy elimination systems. In *IEEE INFOCOM Global Internet*, pages 809–814, IEEE, 2011.
- Y. Xiao, P. Hui, P. Savolainen, and A. Ylä-Jääski. CasCap: cloud-assisted context-aware power management for mobile devices. In *Proceedings of the 2nd International Workshop on Mobile Cloud Computing and Services (MCS'11)*, pages 13–18, 2011.
- Y. Xiao, P. Savolainen, A. Karppanen, M. Siekkinen, and A. Ylä-Jääski, Practical power modeling of data transmission over 802.11g for wireless applications. In *Proceedings of the 1st International Conference on Energy-Efficient Computing and Networking (e-Energy '10)*, pages 75–84, ACM, 2011. Best paper award.
- R. S. Kalyanaraman, Y. Xiao, and A. Ylä-Jääski. Network prediction for energy-aware transmission in mobile applications. *International Journal on Advances in Telecommunications*, 3(1-2):72-82, 2010.
- Y. Xiao, R. Bhaumik, Z. Yang, M. Siekkinen, P. Savolainen, and A. Ylä-Jääski. A system-level model for runtime power estimation on mobile devices. In *Proceedings of the 2010 IEEE/ACM International Conference on Green Computing and Communications (GREENCOM'10)*, 2010.

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. Recently the group has focused on energy modeling and optimization and mobile traffic and computation offloading.

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.

Recent research results include e.g. a new experimental networking and mobile computing laboratory at the Kumpula campus (NODES lab), and joint research activity through ICSI at University of California Berkeley.

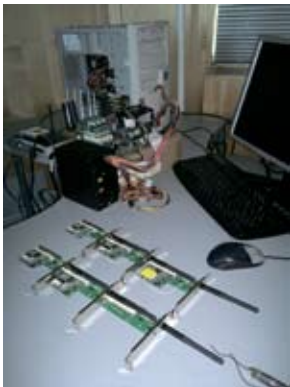


Key publications

- S. Tarkoma. *Publish/Subscribe Systems: Design and Principles*. Textbook. Wiley, 2012. To appear in August.
- S. Tarkoma, C. Rothenberg, and E. Lagerspetz. Theory and practice of Bloom filters for distributed systems. *IEEE Communications Surveys and Tutorials*, 14(1):131–155, 2012.
- S. Aalto, P. Lassila, P. Savolainen, and S. Tarkoma. How impatience affects the performance and scalability of P2P video-on-demand systems. *ACM SIGMETRICS Performance Evaluation Review*, 39(3), 2011.
- S. Tarkoma, and E. Lagerspetz. Arching over the mobile computing chasm: Platforms and runtimes. *Computer* (New York), 44(4):22–28, 2011.
- Y. Xiao, P. Savolainen, M. Siekkinen, A. Ylä-Jääski, and A. Karppanen. Practical power modelling of data transmissions over 802.11g for wireless applications. *ACM e-Energy 2010*, 2010. Best paper award.
- W. Rao, R. Vitenberg, and S. Tarkoma. Towards optimal keyword-based content dissemination in DHT-based P2P networks. In *IEEE Peer-to-Peer Conference (P2P 2011)*, pages 102–111, 2011.



Networking Research, Professor Andrei Gurtov



The Networking Research group investigates the architecture of the future Internet, especially locator/identifier split, performance of wireless protocols, enterprise traffic analysis, multipath routing and multicast, sensor network security, and peer-to-peer systems.

The group collaborates tightly with top-notch scientists from the International Computer Science Institute (ICSI) in Berkeley. We study a wide range of communication protocols, starting from the medium access control up to hierarchical P2P network architectures. 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. In future, we plan to focus on Internet-of-Things security, medical ICT, evolving Internet architecture and flexible routing.

Key publications:

- S. Varjonen, T. Heer, K. Rimey, A. Gurtov. Secure resolution of end-host identifiers for mobile clients. In *Proc. of IEEE Globecom'11*, December 2011. Best paper award.
- B. Nechaev, D. Korzun, A. Gurtov. CR-Chord: Improving lookup availability in the presence of malicious DHT nodes. *Computer Networks*, 55(13):2914–2928, September 2011.
- D. Korzun, A. Gurtov. Survey on hierarchical routing schemes in "flat" distributed hash tables. *Peer-to-Peer Networking and Applications*, 4(4):346–375, 2011.
- T. Koponen, S. Shenker, H. Balakrishnan, N. Feamster, I. Ganichev, A. Ghodsi, P. B. Godfrey, N. McKeown, G. Parulkar, B. Raghava, J. Rexford, S. Arianfar, D. Kuptsov. Architecting for innovation. *ACM Computer Communications Review*, July 2011.
- B. Nechaev, V. Paxson, M. Allman, A. Gurtov. On calibrating enterprise switch measurements. In *Proc. of ACM SIGCOMM Internet Measurement Conference*, November 2009.
- P. Nikander, A. Gurtov, T. Henderson. HIP: Host identity protocol. In B. Ramamurthy, G. Rouskas, K. Sivalingam (eds), *Next-Generation Internet Architectures and Protocols*, Cambridge University Press, 2011.

Network Society (NS) Programme

The mission of the Network Society research programme is to empower ubiquitous users with transparent and resourceful ICT. This includes investigating ubiquitous interaction, virtual behavior, online practices, mediated experiences through explorative design and user research. The Network Society program has consolidated work in public displays and co-located interaction resulting in both publications and startups (screen.io and multitaction.com)

Legacy work in mobile interaction has continued with work on mobile manual usability. In these studies, facilitative design features were identified, whose absence explains, for example, why the mouse and stylus function poorly in multi-object performance. One emergent area of work has been privacy and self-expression in social network services.

Our senior and principal researchers have had an important role in the establishment of the EIT (European Institute of Innovation and Technology) ICT Labs. Currently Professor Marko Turpeinen is the director of the Helsinki node. The outreach of Network Society and international collaboration included research visits of PhD students to UC Berkeley iSchool lasting from 6 to 12 months and visits to Stanford University. A major achievement has been to establish the international Sizzle Virtual Center of Excellence (Sizzle VCE) with UC Berkeley iSchool, BUPT, University of Nairobi, and EIT ICT Labs Helsinki Node. Sizzle VCE is an international consortium studying digital services for end-users. The studies are relatively long and they involve large populations in real world experiments



Digital Content Communities, Docent Olli Pitkänen, Professor Marko Turpeinen, Docent Anu Kankainen (on leave)



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 equires 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 viewpoints 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.

Key publications

- O. Pitkänen, R. Sarvas, A. Lehmuskallio, M. Simanainen, V. Kantola, M. Rautila, A. Juhola, H. Pentikäinen, O. Kuittinen. *Future Information Security Trends*. Helsinki: Liikenne- ja viestintäministeriö, 40 (Arjen tietoyhteiskunta), 2011.
- V. Lehtinen, A. Kankainen. Creative personal projects of the elderly as active engagements with interactive media technology. In *Creativity & Cognition (C&C'11)*, 2011.
- M. Wahlström, A. Kankainen. Digital TV transition and the hard disk drive revolution in television watching. *International Journal of Communication*, 5:1606–102, 2011.
- A. Lampinen, V. Lehtinen, A. Lehmuskallio and S. Tamminen. We're in it together: interpersonal management of disclosure in social network services. In *CHI '11 Proceedings of the 2011 Annual Conference on Human factors in computing systems*, 2011.

Social Interaction and Emotion (SIE), Professor Niklas Ravaja

Social Interaction and Emotion (SIE) is a new research group the mission of which 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. Recently we have started to examine, for example, neurophysiological processes associated with mediated haptic communication during social decision making.

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:

- J. King, A. Lampinen and A. Smolen. Privacy: is there an app for that? In *SOUPS '11 Proceedings of the Seventh Symposium on Usable Privacy and Security*, 2011.
- S. Silfverberg, A.L. Liikkanen and A. Lampinen, "I'll press play, but I won't listen": profile work in a music-focused social network service. In *CSCW '11 Proceedings of the ACM 2011 conference on Computer supported cooperative work*, 2011.
- J. Kätsyri, N. Ravaja, M. Salminen. Aesthetic images modulate emotional responses to reading news messages on a small screen: A psychophysiological investigation. *International Journal of Human-Computer Studies*, 2011.
- I. Ekman, G. Chanel, S. Järvelä, M. Kivikangas, M. Salminen, N. Ravaja. Social interaction in games: Measuring physiological linkage and social presence. *Simulation & Gaming*, 2011.
- M. Kivikangas, I. Ekman, G. Chanel, S. Järvelä, B. Cowley, M. Salminen, P. Henttonen, N. Ravaja. Review on psychophysiological methods in game research. *Journal of Gaming & Virtual Worlds*, 3:181–199, 2011.



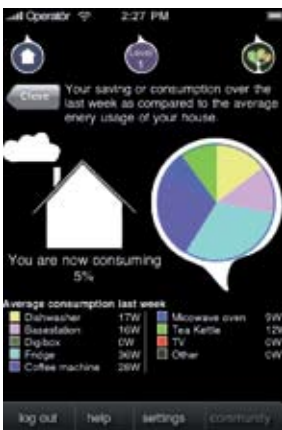
Ubiquitous Interaction, Professor Giulio Jacucci



Ubiquitous Interaction (UiX) studies opportunities in ubiquitous and mobile computing, 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 perspectives. The group has recently published visions of mobile and ubiquitous computing on mobile augmented reality, environmental awareness for the home and future development of public screens.

Several ACM CHI papers and journal papers addressed co-located interaction in ubiquitous settings. Studies of mobile augmented reality with smartphones on a physical map detailed how such technology invites collaboration through providing common ground and place making as compared to purely digital and mobile applications (Morrison et al 2009, 2011). Large public display studies detailed performing, turn taking and configuring of collective interaction on a multitouch walk-up-and-use display (Jacucci et al 2010, Peltonen et al. 2008). Another interesting recent result concerns studies of mobile interaction including manual multitasking, ergonomics of hand use, and walk speed in mobile interaction. For example a work varied the walking speed and measured the effects on discrete aiming movements on a touchscreen.

Developing and evaluating a persuasive game for energy awareness including the development of smart advice as contextually triggered and personalized advices has been another recent focus (IEEE Computer 2011, Persuasive 2012, Advances in Computer Entertainment 2012).



Key publications:

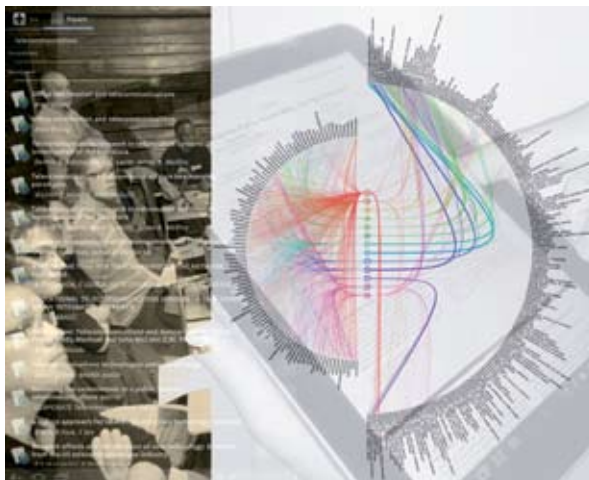
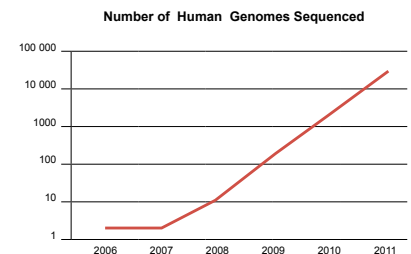
- K. Kuikkaniemi, G. Jacucci, M. Turpeinen, E. E. Hoggan, J. Müller. From Space to Stage: How interactive screens will change urban life. *IEEE Computer* 44(6):40–47, 2011.
- A. Spagnolli, N. Corradi, L. Gamberini, E. E. Hoggan, G. Jacucci et al. Eco-feedback on the go: Motivating energy awareness. *IEEE Computer* 44(5):38–45, 2011.
- P. Froehlich, A. Oulasvirta, M. Baldauf, and A. Nurminen. On the move, wirelessly connected to the world. *Communications of the ACM*, 54(1):132–138, 2011.
- A. Oulasvirta, J. Bergstrom-Lehtovirta. Ease of juggling: studying the effects of manual multitasking. In *CHI'2011*, pages 3103–3112, 2011. Best paper.
- J. Bergström-Lehtovirta, A. Oulasvirta, S. A. Brewster. The effects of walking speed on target acquisition on a touchscreen interface. In *Mobile HCI'2011*, pages 143–146, 2011. Best paper.

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

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

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.

In genomics 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 are currently no tools to properly utilise it and hence make biology cumulative in terms of the collected data. We are working on two areas of methods development for improving utilisation of this data: modelling methods for personalised medicine and content-based retrieval of molecular biology data sets.



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 (almost 300) is much higher than the number of person-years. 29 per cent of HIIT staff are of foreign origin.

In 2011 HIIT staff completed 135 person-years. The total funding of HIIT was 10.5 Million Eur, of which 78 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.

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.

In 2011 the Board convened four times. Board Members and their personal deputies:

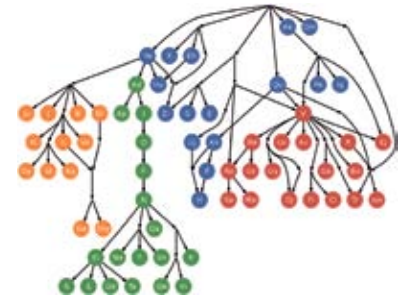
- Vice President, Professor Heikki Mannila, Aalto
(Vice Dean, Professor Outi Krause, Aalto)
- Professor Olli Simula, Aalto (Professor Heikki Saikkonen, Aalto)
- Vice Rector, Professor Johanna Björkroth, UH
(Vice Rector, Professor Jukka Kola, UH)
- Professor Hannu Toivonen, UH (Professor Jukka Paakki, UH)
- Vice President Henry Tirri, Nokia (Director Jyri Huopaniemi, Nokia)
- Docent Kari-Pekka Estola
(Director Martin Mäklin, TeliaSonera Finland Oy)

- Director Petri Vasara, Pöyry Management Consulting Oy (Dr. Lars Gädda, Metsäklusteri Oy)
- MSc Kimmo Kiviluoto, Webmie Oy (Research Fellow Aimo Maanavilja, Elisa Communications)
- Until August 2011: Dr Risto Sarvas, HIIT (Docent Ella Bingham, HIIT).
From August 2011 onwards: 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 Secretaries were Research Coordinator, Dr. Greger Lindén (until June 2011) and Research Coordinator, Docent Ella Bingham (from July 2011 onwards).

Scientific Advisory Board (SAB)

The objective of the Scientific Advisory Board (SAB) is to provide critical guidance about HIIT's research activities and to advise the HIIT Board on strategic planning for future research directions of HIIT. The SAB consists of internationally prominent scholars who are invited by the HIIT Board. The SAB did not convene in 2011, but a SAB meeting is planned for June 2012.



SAB Members 2008-2011:

- Professor Ross Anderson, University of Cambridge
- Professor Alberto Apostolico, Georgia Tech
- Professor Richard Buxbaum, University of California at Berkeley
- Professor Christos Faloutsos, Carnegie Mellon University
- Professor Bengt Jonsson, Uppsala University
- Professor Randy Katz, University of California at Berkeley
- Professor Martin Kersten, University of Amsterdam and CWI
- Professor Kari-Jouko Räihä, University of Tampere
- Professor Mart Saarma, University of Helsinki
- Professor Angela Sasse, University College London
- Professor John Shawe-Taylor, University of Southampton
- Professor Hal Varian, University of California at Berkeley
- Professor Martin Vingron, Max Planck Institute for Molecular Genetics

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 in 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 Centres of Excellence and co-operate with the information industry and with sciences applying information technology. HIIT is now in its 12th year of operation with a budget of about 10 million euros, realising in total about 135 person-years of work; a major part of these operations are done in direct collaboration within the departments of the mother universities. The research by HIIT has been assessed and deemed excellent by its Scientific Advisory Board in 2008, 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.