



UNIVERSITY OF HELSINKI



HELSINKI UNIVERSITY OF TECHNOLOGY

HELSINKI INSTITUTE FOR INFORMATION TECHNOLOGY

---

# **Helsinki Institute for Information Technology**

**Overview material for the meeting of the**

**Scientific Advisory Board**

**November 2004**

**Martti Mäntylä and Esko Ukkonen (eds.)**

**October 26, 2004**



<b>1</b>	<b>INTRODUCTION</b> .....	<b>4</b>
<b>2</b>	<b>RESEARCH STRATEGY</b> .....	<b>5</b>
<b>3</b>	<b>RESEARCH GROUPS</b> .....	<b>6</b>
<b>3.1</b>	<b>MOBILE COMPUTING</b> .....	<b>6</b>
	Premises of the research .....	6
	Present status .....	7
	Highlights of recent achievements .....	7
	Future directions.....	8
	Selected publications .....	9
<b>3.2</b>	<b>COMPLEX SYSTEMS COMPUTATION</b> .....	<b>9</b>
	Premises of the research .....	9
	Present status .....	10
	Highlights of recent achievements .....	12
	Future directions.....	13
	Selected publications .....	15
<b>3.3</b>	<b>SEMANTIC COMPUTING</b> .....	<b>15</b>
	Premises of the research .....	15
	Present status .....	16
	Highlights of recent achievements .....	16
	Future directions.....	16
	Selected publications .....	17
<b>3.4</b>	<b>DIGITAL CONTENT COMMUNITIES</b> .....	<b>18</b>
	Premises of the research .....	18
	Present status .....	19
	Highlights of recent achievements .....	20
	Future directions.....	21
	Selected publications .....	22
<b>3.5</b>	<b>USER EXPERIENCE RESEARCH</b> .....	<b>22</b>
	Premises of the research .....	22
	Present status .....	24
	Highlights of recent achievements .....	24
	Future directions.....	25
	Selected publications .....	26
<b>3.6</b>	<b>DIGITAL ECONOMY</b> .....	<b>26</b>
	Premises of the research .....	26
	Present status .....	28
	Highlights of recent achievements .....	28
	Future directions.....	30
	Selected publications .....	32



<b>3.7 DATA MINING: THEORY AND APPLICATIONS .....</b>	<b>32</b>
Premises of the research .....	33
Present status .....	33
Highlights of recent achievements .....	35
Theory of data mining / future directions .....	35
Theory of data mining / selected publications .....	36
Applications of data mining / future directions.....	36
Applications of data mining / selected publications .....	37
<b>3.8 ADAPTIVE COMPUTING SYSTEMS .....</b>	<b>37</b>
Premises of the research .....	38
Present status .....	38
Highlights of recent achievements .....	40
Future directions.....	40
Selected publications .....	42
<b>3.9 NEUROINFORMATICS .....</b>	<b>42</b>
Premises of the research .....	42
Present status .....	43
Highlights of recent achievements .....	43
Future directions.....	44
Selected publications .....	45



# Helsinki Institute for Information Technology (HIIT)

## 1 Introduction

Helsinki Institute for Information Technology (HIIT) is a joint research institute of the University of Helsinki (UH) and the Helsinki University of Technology (HUT). HIIT conducts strategic research in information technology and related multi-disciplinary topics, aiming at the highest international level and impact. As an integral part of this, it also participates in graduate teaching at the parent universities. HIIT works in close co-operation with Finnish universities, research institutes, and industry. It also aims at creating a strong network of partnerships with leading foreign research universities and institutions.

HIIT's research is organised as four research programmes addressing selected major research challenges of the future: *Future Internet*, *Intelligent Systems*, *Network Society* and *Data Analysis*. The complexity and interaction of the challenges requires multidisciplinary research where computer science competence is combined with other disciplines. The present research groups in HIIT include *mobile computing*, *user experience*, *complex system computation*, *digital economy*, *digital content communities*, *semantic computing*, *adaptive computing systems*, *data mining*, and *computational neuroscience*. In addition, many other research groups from UH and HUT and elsewhere participate in its work.

Organisationally, HIIT at present consists of two units with a total of has some 110 research personnel (including part-time personnel). As discussed below, HIIT aims at merging the two units during 2005.

The Advanced Research Unit (ARU), established in 1999, mainly operates through two- to three-year long industry co-funded research projects. The ARU is located in the High Tech Center Helsinki in Ruoholahti, Helsinki.

The Basic Research Unit (BRU), established in 2002, focuses on long-term research issues in information technology in co-operation with other sciences and industry. The BRU is located at the premises of the Department of Computer Science of University of Helsinki in Kumpula, Helsinki, and the Department of Computer Science and Engineering of Helsinki University of Technology in Otaniemi, Espoo.



## 2 Research Strategy

HIIT was originally founded by agreement of UH and HUT for five years from 1 August 1999 to 31 July 2004. After this initial period, the agreement would continue for a year at a time unless specifically discontinued.

The Board of HIIT decided in February 2004 that the original agreement will be used for one additional year. During this time, the strategy of HIIT will be evaluated and reformulated if needed. Moreover, a new agreement reflecting the chosen strategy and more generally the experience of the first five-year period will be prepared and put to effect from 1 August 2005.

Work towards these goals has been underway since March 2004. By June 2004, a draft baseline strategy document was presented and accepted by the Board. A major conclusion of this document that focused in the interface between HIIT and its parent universities is that the present two units of HIIT and their research lines will be merged.

In parallel, complementary discussion amongst HIIT's senior researchers on HIIT's mission and the scope of its research domains has been underway. Early results of this discussion were presented in HIIT's Industrial Advisory Board meeting in September and also to HIIT Board in October.

One of the aims of the Scientific Advisory Board meeting in November 2004 will be to discuss the different alternatives in implementing the planned strategy. The HIIT strategy draft document is attached to the overview material as a separate appendix.



## 3 Research Groups

### 3.1 Mobile Computing

Director: Professor Kimmo Raatikainen. Senior researchers: Dr. Ken Rimey, Dr. Andrei Gurtoy, Ericsson Visiting Senior Research Scientist Dr. Pekka Nikander.

#### *Premises of the research*

The Mobile Computing Group focuses on the software infrastructure for future mobile applications.

Recent developments in mobile communication and small computing devices have had a tremendous impact on society. They have brought the dream of ubiquitous or invisible computing and communication closer to reality. Already in the near future communication and computing devices will be in a state that technologically enables mass market scale ubiquitous services and applications. Therefore, the main remaining challenge will be software that fulfils the needs of adaptive, context-aware, personalized services and solutions. In the longer term future, novel networked computing elements such as MEMS sensors and actuators equipped with short-range radio communications will however pose further challenges to mobile computing research. These may include novel architectures, programming models, and consideration of issues such as privacy and trust.

In this area various visions have been proposed and research challenges formulated. Mark Weiser spoke about invisible computing and ubiquitous computing. Leonard Kleinrock speaks about nomadic computing. Mahadev Satyanarayanan speaks about pervasive computing. The European Commission speaks about ambient intelligence. Wireless World Research Forum speaks about adaptable personalized ambient-aware services. When these (and similar proposals) are carefully examined, the conclusion is that they are — at least — very close to each other, if not different names for the same thing. There may be some minor differences in emphasis but the core and major challenges are similar.

A significant requirement has, already for several years, been ever-faster service development and deployment. The immediate implication has been the introduction of various service/application frameworks/platforms. Another significant factor, since mid/late 90s, is the requirement of world-wide access. This has made contributions to international standardization the most crucial impact factor.



### *Present status*

The activities of the HIIT's research theme Mobile Computing concentrate on development of mobile middleware. The main research questions studied are:

- to identify key service enablers (middleware functionality) for future mobile applications, and
- to specify selected service enablers and to implement their prototypes.

The current projects in progress are Fuego Core (Middleware for Mobile Wireless Internet), PDIS (Personal Distributed Information Store), Dynamos (Dynamic Composition and Sharing of Context-aware Mobile Services), and CONTEXT (Context Recognition by User Situation Data Analysis, joint research with Adaptive Computing Systems group).

The group is active in international standardization and co-operation. We have contributed to IETF (Host Identity Protocol, HIP), W3C (Binary XML), and Wireless World Research Forum (Reference Model for Service Architecture). We have co-operated with UC Berkeley for several years. Currently we are establishing co-operation with Waseda University (Tokyo, Japan), Tsinghua University (Beijing, China) and Petrozavodsk State University (Russia). In Europe the main co-operation forums are WWRF and European Science Foundation's MINEMA-project.

### *Highlights of recent achievements*

The Fuego Core project has selected SOAP/XML messaging [1], asynchronous events [2] and synchronization [3] as the key enablers that were also specified and implemented in the Fuego Core prototype. The prototype also includes the Host Identity Protocol (HIP) as a new way to support mobility, multicast and multihoming in a secure manner and Mobile Presence Service to demonstrate features of Fuego middleware services.

The PDIS project has created an engineering prototype of its state-transfer, update-anywhere replicated XML database concept. This has been ported to high-end phones running Symbian OS. The project is the most ambitious test user of Nokia's *Python for Series 60* platform.

The major achievements of the HIP activities include producing an open-source HIP implementation in the Linux kernel, developing the HIP protocol and standardizing it in the



IETF [4], and studying advanced issues related to a control plane for the Internet with establishing a research group in the IRTF.

### *Future directions*

Group's current results lay a sound basis for future developments. In particular, we will concentrate on the following topics:

- future developments in mobile middleware services (XML messaging, events, XML/RDF data access, run-time software configuration management) taking also into account implications of ad-hoc communities and privacy;
- end-user digital asset management for networked consumer electronic devices;
- separation of location and identity in Internet protocols.

The work on future middleware services, led by Prof. Raatikainen, is aimed at future applications expected to be context sensitive, reconfigurable, personalized, and available everywhere, anytime, and anyhow. The mission is to develop and demonstrate through functional prototypes fundamental enablers for future mobile computing. We plan to focus especially to ad-hoc communities and take security, trust and privacy into account.

In the digital asset management thread led by Dr. Rimey, we will leverage the PDIS software for maximum impact by building on it but shifting the research focus in follow-on projects. One direction will be to build a testbed home network to experiment with how consumers might manage their files on multiple handheld devices, computers, and home/car entertainment boxes. Another will focus on the underlying self-administering XML database technology for small devices, and particularly on automatic index management.

Future activities on the identifier/locator split led by Dr. Gurtoev will go in two main directions. First, an infrastructure required for wide scale deployment of HIP will be developed. This includes DNS extensions, NAT and firewall traversal, finalizing the HIP Linux implementation, and supporting multiple identities per a single host. In addition, more advanced concepts including the design of DoS-resilient control plane for HIP and process migration will be studied. Second, a more theoretical work on multipath communications using HIP will be commenced. This includes new algorithms for multicast and congestion control.



*Selected publications*

- 1) J. Kangasharju, S. Tarkoma, K. Raatikainen: Comparing SOAP Performance for Various Encodings, Protocols, and Connections. IFIP-TC6 8<sup>th</sup> International Conference, Personal Wireless Communications (PWC 2003), Venice, Italy, September 23-25, 2003. *Lecture Notes in Computer Science*, Vol. 2775.
- 2) S. Tarkoma: Event Dissemination Service for Pervasive Computing. Pervasive 2004 Doctoral Colloquium, 18-19.4. Linz/Vienna, Austria. Published in *Advances in Pervasive Computing*, Austrian Computer Society.
- 3) T. Lindholm: A Three-way Merge for XML Documents. Accepted for publication in *DocEng 2004, ACM Symposium on Document Engineering*, Wisconsin, October 28-30, 2004.
- 4) P. Nikander, J. Ylitalo, J. Wall: Integrating Security, Mobility, and Multi-Homing in a Hip Way. *Proceedings of Network and Distributed Systems Security Symposium (NDSS'03)*, February 6-7, 2003.
- 5) S. Tarkoma, R. Balu, J. Kangasharju, M. Komu, M. Kousa, T. Lindholm, M. Mäkelä, M. Saaresto, K. Slavov, K. Raatikainen: *State of the Art in Enablers for Applications in Future Mobile Wireless Internet*. Helsinki Institute for Information Technology HIIT (HIIT Publications 2004-2).

### **3.2 Complex Systems Computation**

Director: Professor Petri Myllymäki. Senior researchers: Dr. Wray Buntine, Dr. Jorma Rissanen, Nokia Research Fellow Henry Tirri.

*Premises of the research*

It has been said that we are living the century of information. Indeed, with the hundreds of billions of pages of information available in the Internet alone, it is clear that accessing or obtaining data is often not the problem, but a more and more important question is how to find the data we need, how to extract the information we need from the data, and how to communicate the essential parts of the extracted knowledge to others. Therefore, it is evident that methodologies for organizing and analyzing the available data are rapidly becoming the key technologies of the future, affecting all the areas of our society.



The goal of this research is to develop computationally efficient, general-purpose techniques for very large-scale data analysis. We view data analysis as a modelling problem, so that the objective is to develop methods for constructing models from sample data and possibly some domain-specific background information. A model is a compressed representation of the data, the compression being a result of exploiting regularities found in the analysis of the data. Our goal is to build models that can be used for both predictive tasks as well as explorative tasks. Examples of predictive tasks are classification and regression, whereas explorative tasks include data mining, data/model visualization and causal reasoning. Consequently, in summary we can say that the general goal in this research area is to develop computationally efficient methods for extracting useful information from massive amounts of data.

We strongly believe that theoretical work is not sufficient alone, but the developed theoretical frameworks must be implemented and validated in realistic real-world settings. Therefore, our work has both a strong basic research component, being at the intersection of computer science, information theory and mathematical statistics, and an applied component where the results are being used for solving real-world problems in various problem domains related for example to social sciences, criminology, ecology, medicine or industrial engineering. The research group benefits from a rare combination of theoretical competence with top-level programming skills, which has made it possible to perform the extensive empirical tests required for evaluating the theoretical methods developed.

### *Present status*

On the theoretical side, one of the key research areas is model selection. One research issue here is how to define theoretically justified objective criteria for constructing models from data, for both unsupervised and supervised settings. At the same time we are developing more efficient search algorithms that are required for finding good candidates from the set of all possible models. In this work, the emphasis has been on probabilistic graphical models, such as Bayesian networks, and the new results are being distributed to the research community through the unique B-Course data analysis server [10] maintained by the research group (see <http://b-course.hiit.fi>).

The model selection work is concerned both with Bayesian and non-Bayesian model selection criteria, and the work is done in close co-operation with Peter Grünwald from CWI. The work on non-Bayesian model selection is focused on the Minimum Description Length principle, with the developer of the MDL framework, Dr. Jorma Rissanen, participating



actively in the research. We are also working towards a theoretical framework unifying the Bayesian and the MDL approaches.

After Dr. Wray Buntine joined the CoSCo group in 2002, one of the focus research areas has been concentrated on next generation information retrieval techniques. Besides domain-specific theoretical work (model selection for text documents), the research involves also issues related to Internet crawling, distributed data processing, grid computing and so on. The most basic research part of the work is funded by the Academy of Finland, while the Finnish National Technology Agency Tekes and the EU fund the more applied research work. The Tekes project is performed in co-operation with the human-computer interaction group at University of Tampere, CERN, and the Finnish companies participating in the project. The EU funding comes through the STREP project Alvis where CoSCo works as the coordinating unit. The work in this area is done in co-operation with the researchers at the 11 European Alvis partner sites, with University of Tsinghua (China), and with Prof. Michael Jordan's group at Berkeley (USA). In addition, another research project funded by the PASCAL EU Network of Excellence was just started together with CWI (with Dr. Peter Grünwald and Prof. Paul Vitanyi), aiming at more ambitious information retrieval problems where the goal is to deal with documents containing not only text, but also other types of data, such as audio, images or video.

The probabilistic modelling techniques are currently being applied in several domains. One active domain is concerned with developing tools for adaptive, collaborative learning environments. Some of the tools developed are already in fielded use in higher education (for more details, see <http://cosco.hiit.fi/edutech/>). Another application area is medicine, where we are working with domain experts towards building new, intelligent tools for health management and analysis of treatment effectiveness. This work is done under contract from University of Kuopio, Department of Health Policy and Management. Together with the Center for Knowledge and Innovation Research of Helsinki School of Economics, and the Neural Networks Research Center at the Helsinki University of Technology, we are developing probabilistic models that learn from the actions of people to model their intentions and expectations, and use the models for disambiguating the users' vague commands and proactively anticipate their actions. The actions and interests are monitored by measuring eye fixations and movements that exhibit both voluntary and involuntary signs of both the cognitive state of the user and his intentions. In a project coordinated by the Laboratory of Computational Engineering of Helsinki University of Technology, we are developing distributed probabilistic modelling algorithms for handling micro array data.



### *Highlights of recent achievements*

Our continued work in the multinomial or discrete PCA (principle components analysis) has lead to the useful result that the technique turns out to be a discrete counterpart to ICA (see e.g.[7]). Our C and Linux based system for this task is able to handle datasets an order of magnitude larger than those addressed by researchers at Stanford, Berkeley and UC Irvine. We are using the technique as a basis for building topic hierarchies of large document collections. Here it is remarkably effective: the topic structures yield good insights in the content. This has spawned the auxiliary task of finding names for discrete text components. The results if this research area will be published as an open-source library forming a basis for next-generation search engines, the capabilities of which surpass those of Google and other existing systems.

The most recent version of the MDL criterion for is based on the normalized maximum likelihood criterion, which involves a normalization factor that is in principle intractable. However, we showed that using some combinatorial tricks, the normalizing factor can in many cases be computed in polynomial time [6]. Furthermore, the factor can also be approximated very accurately using computationally efficient techniques based on complex analysis. These results offer a theoretically elegant yet computationally feasible framework for model selection.

Recently it was proved that for a certain interesting class of Bayesian network model structures, the problem of finding the parameters maximizing the conditional likelihood (as opposed to the standard unsupervised joint likelihood) is a convex problem so that the optimal solution can be found easily [9]. This means, among other things, that the search methods developed for finding the conditionally optimized parameters of the Naïve Bayes classifier are insignificant, as the same result can be obtained by simple greedy search. The result also offers new possibilities for developing learning criteria for supervised domains.

An example of a commercially successful application of the methods developed in the group is given by the Ekahau Positioning Engine (EPE), a tool for locating devices in wireless devices. The EPE software was developed by Ekahau Inc, based on the innovations done by the group members [8], and to date, it has won the following prestigious awards:

- European Union: The European Information Society Technology Prize 2002.
- Technology Marketing Corporation (TMC): Best product of the year 2002.
- Planet PDA, the Global Summit on Enterprise & Custom Volume Handheld Computing: Best of show.



- Software Industry Summit: Best commercialized innovation in Finland in 2002.
- SearchNetworking.com: Bronze medal, best product of the year 2003 (enterprise wireless applications and systems).
- Wi-Fi Planet 2004: Best of Show.

### *Future directions*

Our work in next generation information retrieval has led us to the question of determining "relevance" for a query. How are matching documents ranked so that those with higher rank appear "more relevant" to the query author? This issue turns out to be a classical probability modelling question. Standard techniques in information retrieval such as PageRank and TF.IDF have good intuitive justifications but lack any formal framework, although their computational aspects have formal frameworks.

One of the challenges faced in the future is the fact that the amount of available information grows enormously fast, so that the computational techniques used must be made as efficient as possible. We plan to study the use of on-line modelling techniques where the constructed models can be efficiently updated incrementally as new data comes in.

Many of the current modelling techniques assume that the data is in some, predefined clean form. However, in reality the situation is more difficult: for instance, in information retrieval, the documents contain typically many types of information, not only clean text. On the other hand, in many problems domains the numerical data is typically quite noisy and requires pre-processing. The goal here is to develop intelligent pattern recognition and feature extraction methods so that the pre-processing stage can be incorporated in the overall theoretical framework.

As discussed above, we have studied how the standard unsupervised model selection criteria need to be modified in the supervised learning domains. More generally, one can ask how to modify the criteria for arbitrary loss functions. The work will be done both in the MDL setting and in the Bayesian framework. Simultaneously, we will be working towards a unifying theoretical framework by identifying links between MDL and Bayesian modelling.

Many theoretical frameworks assume that we are modelling a static environment with only one acting agent, while in reality we are faced with dynamical systems with several stochastically interacting agents. What is more, in many settings it is not sufficient to model the environment, but we may also want to make decisions based on the predictions obtained through the model, and possibly also optimize the environment being modelled.



One interesting manifestation of this type of an environment is offered by the sensor networks. We believe that many of our earlier results can be extended to this type of more complex settings: for example, while we have obtained good results in the problem of locating devices in infrastructure-based wireless networks, where the devices estimate their location with respect to transmitters at fixed locations, it would be a more challenging problem to estimate the locations of devices in an ad-hoc sensor network, where the devices can only sense their dynamically changing neighbourhoods.

With respect to national cooperation, joint efforts will continue both as formal research projects and as informal co-operation between the researchers. Now that Henry Tirri is working as a Nokia Research Fellow, we expect our co-operation with the Nokia Research Center to increase substantially. In particular, we expect to discover together new interesting research problems that are relevant to wireless networks of the future.

With respect to international co-operation, we feel that the group has reached a level where we can offer a challenging and rewarding working environment to top-level researchers, and our very ambitious goal is to recruit talented post-docs from abroad to stay in the group on a permanent basis. This effort was encouraged by the fact that we were able of getting Dr. Wray Buntine, one of the most appreciated researchers in our field, to join the group. More recently, it has been tentatively agreed with Ms. Huizhen Yu that she will move to Finland to join the group on a permanent basis, as soon as she will complete her Ph.D. at MIT under supervision of Professor Bertsekas, most probably during the first half of 2005. These recruitment efforts are supported by the joint research projects started with Professor Michael Jordan's group at UCB, and Professor Lizhu Zhou's group at University of Tsinghua. Two Ph.D. students from UCB have already visited the CoSCo group for a period of a month, and from Tsinghua one Ph.D. student and two more senior researchers will come for visits of 1-6 months, starting in 2004.

In Europe, much of the international cooperation with respect to the document retrieval work will naturally be carried out through the 11 partners of the Alvis project. The joint efforts with CWI in Amsterdam will continue even more actively now that we have a formal joint research project funded by the Pascal NoE. New potential European partners will be searched through the Pascal network, where Professor Myllymäki is working as a steering committee member and Helsinki core site manager.



### *Selected publications*

- 6) P. Kontkanen, P. Myllymäki, W. Buntine, J. Rissanen, H. Tirri: An MDL Framework for Data Clustering. In P. Grünwald, I.J. Myung, M. Pitt (eds.), *Advances in Minimum Description Length: Theory and Applications*. The MIT Press, 2004.
- 7) W. Buntine, A. Jakulin: Applying Discrete PCA in Data Analysis, in M. Chickering, J. Halpern (eds.). *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence (UAI'04)*, 59-66. AUAI Press, 2004.
- 8) P. Kontkanen, P. Myllymäki, T. Roos, H. Tirri, K. Valtonen, H. Wettig: Probabilistic Methods for Location Estimation in Wireless Networks. In R. Ganesh, S. Kota, K. Pahlavan, R. Agustí (eds.), *Emerging Location Aware Broadband Wireless Adhoc Networks*, chapter 11. Kluwer Academic Publishers, 2004.
- 9) H. Wettig, P. Grünwald, T. Roos, P. Myllymäki, H. Tirri: When Discriminative Learning of Bayesian Network Parameters Is Easy. In G. Gottlob, T. Walsh (eds.), *Proceedings of the 18th International Joint Conference on Artificial Intelligence*, 491-496. Morgan Kaufmann, 2003 (longer version to appear in the Machine Learning Journal).
- 10) P. Myllymäki, T. Silander, H. Tirri, P. Uronen: B-Course: A Web-Based Tool for Bayesian and Causal Data Analysis. *International Journal on Artificial Intelligence Tools*, Vol. 11(3): 369-387, 2002.

### **3.3 Semantic Computing**

Director: Professor Eero Hyvönen

#### *Premises of the research*

Our research focuses on Semantic Web and Web Services technologies and applications. Semantic web is the next generation of WWW based on machine-understandable data envisioned originally by Tim Berners-Lee. This new area has rapidly grown into a research field of its own with major conference series (ESWC, ISWC) and new journals (see e.g. <http://www.w3.org/2001/SW/> for the W3C efforts and <http://semanticweb.org> for pointers to academic research and development).





### *Present status*

SeCo (<http://www.cs.helsinki.fi/gourp/seco/>) is a leading European research group in this new emerging field. At the national level, the group directs the national “Semantic web ontologies in Finland” project. There are currently about 10 researchers in the research team (and more are joining the group) working together in multi-disciplinary research consortiums totaling in over 20 organizations and companies.

### *Highlights of recent achievements*

Our most important achievement until now is the semantic portal “MuseumFinland – Finnish Museums on the Semantic Web” [11, 12, 13] and a few other related applications to it (<http://www.cs.helsinki.fi/group/seco/museums/>) [14]. This system is not only of scientific interest: the pilot version of the portal was actually opened on March 8, 2004 at <http://museosuomi.cs.helsinki.fi>. Some 10 peer-reviewed international research papers of the system have been published in 2004. During the summer 2004, some 10,000 pages of the portal were loaded on the busiest day. Normally the figure is lower but shows that our ideas are of interest to end-users, too.

MuseumFinland was a Nominee of Nordic Digital Excellence in Museums (Nodem 2004) in May 2004. MuseumFinland will receive the international Semantic Web Challenge Award 2004 (<http://challenge.semanticweb.org/>) at the 3<sup>rd</sup> International Semantic Web Conference (ISWC 2004) in Hiroshima, Japan in November 2004, together with two other applications (the final ranking among the three candidates will be announced at the conference).

### *Future directions*

As a result of this work, a large national consortium of 19 major Finnish organizations and companies was formed to fund and start a follow-up project in 2003 (0.65Me). In this research project, focusing on semantic web ontology technologies (<http://www.cs.helsinki.fi/group/seco/ontologies>), MuseumFinland is developed further with new kinds of ontologies and semantic content (art, videos, narratives, etc.) into “CultureFinland”. Major research themes of this work include:

- Ontology libraries
- Ontology mapping
- Ontology versioning and management
- Uncertainty in ontologies





- Semantics of upper level ontologies
- Semantic interoperability of cross-media content
- Semantic recommendation systems

In another related major project “Intelligent Web Services”, semantic web technologies of MuseumFinland are developed further in the domain of intelligent yellow page catalog systems and web services. As one result, we have recently managed to scale the semantic RDF-based search engine of MuseumFinland up to 2,3 million web pages and 275,000 categories taken from the international <http://dmoz.org> repository [15].

The SeCo group, operating at the intersection of the HIIT-ARU and University of Helsinki, Department of Computer Science, has represented its hosting organizations in the European OntoWeb research network (<http://www.ontoweb.org>), and has become a national hub in its research field. For example, the group organized the “Semantic Web Kick-off in Finland” conference in 2001. In 2002, SeCo was a key organizer of the international “Towards the Semantic Web and Web Services” –conference, and in 2004 in the “Web Intelligence” –symposium held as a part of the 11<sup>th</sup> Finnish AI Conference

#### *Selected publications*

- 11) E. Hyvönen, S. Saarela, K. Viljanen, E. Mäkelä, A. Valo, M. Salminen, S. Kettula, M. Junnila: A Portal for Publishing Museum Collections on the Semantic Web. *Proceedings of ECAI/PAIS 2004*, Valencia, Spain, 2004.
- 12) E. Hyvönen, M. Junnila, S. Kettula, E. Mäkelä, S. Saarela, M. Salminen, A. Syreeni, A. Valo, K. Viljanen: Finnish Museums on the Semantic Web. User's Perspective on MuseumFinland. *Proceedings of Museums and the Web 2004 (MW2004)*, Arlington, Virginia, USA, March 31 - April 3, 2004. Selected Papers.
- 13) E. Hyvönen, M. Salminen, S. Kettula, M. Junnila: A Content Creation Process for the Semantic Web. *Proceedings of OntoLex 2004*, Lisbon, Portugal, 2004.
- 14) E. Hyvönen, S. Saarela, K. Viljanen: Application of Ontology Techniques to View-Based Semantic Search and Browsing. In C. Bussler, J. Davies, D. Fensel, R. Studer (eds.), *The Semantic Web: Research and Applications*. Proceedings of the First European Semantic Web Symposium (ESWS 2004), Springer-Verlag, LNCS 3053, 2004.



- 15) E. Makelä, E. Hyvönen, S. Saarela, K. Viljanen: OntoViews -- A Tool for Creating Semantic Web Portals. *Proceedings of the 3rd International Semantic Web Conference (ISWC 2004)*, Hiroshima, Japan, 2004, Springer-Verlag, forthcoming.

### 3.4 Digital Content Communities

Director: Dr. Marko Turpeinen. Senior researchers: Dr. Timo Saari.

#### *Premises of the research*

Information technology is social technology. In a networked society, it can reinforce existing physical communities and help build entirely new communities.

As information technology penetrates all realms of society – business, government, communities, and individuals – it is reconstructed by people to fit with their needs and priorities. Internet is also fostering a decentralized and community-oriented mode of media content creation and distribution. On-going technology adoption fuels radical consumer-as-producer (“prosumer”) innovations between the commercial actors and the community members. The scope of these phenomena is broader than “media convergence”, including digital convergence in governance, education, health care, security, life management, ...

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

The goal of our research is to enable and encourage people to belong to communities of content creators, to study how these activities promote a social change in their community setting, and to understand what new business opportunities and structures emerge around community-created content and technology.

The anticipated scientific impact of this research is both methodological and demonstrative. To understand the role and use of technology in people's lives, new inter-disciplinary methodologies are needed. The demonstrative impact is to construct and show new ways to take advantage of technology to support the community activities. The industrial impact is based on creating powerful demonstrators that help to communicate the ideas to the



company representatives, supporting open development communities that provide opportunities for new business, and having a continuous dialogue to communicate people's needs and habits to commercial product and service design

### *Present status*

Community content and societal computing forms the newest addition to the portfolio of HIIT research. This work is conducted in the *Digital Content Communities (DCC)* research group. The group was formed in Spring 2003 and it is led by Dr. Marko Turpeinen. He has six researchers in the group (three Ph.D. students and three M.Sc. students). The researchers have their background in information technology, media arts and sciences, industrial management and law. Dr. Turpeinen also acts as the VP of Research & Technology at Alma Media, which is a Finnish media group.

The current research work is focused on communities that create, share and consume digital media, and how these activities bring new opportunities for public and private services. We are also interested in studying how semantically rich metadata descriptions provide new opportunities for personal and professional media creation.

The *Mobile Content Communities (MC2)* project, which started in June 2003, studies the social meaning and impact of new communication technology for communities that are interested in mobile gaming. The expected results of the MC2 project include evaluated and tested scenarios of mobile community gaming, template-based design tools that allow people to create their own games and game-related content, new open source tools to empower the community activity, and company-specific case studies to help the industry partners to benefit from community-created content.

We also study how the integration of a camera into a mobile phone affects people and their photographing. The first phase of camera phone research included a mobile picture annotation system named *Mobile Media Metadata (MMM)* that was built and tested. The MMM system demonstrates a novel approach in combining contextual information, content-based image retrieval, and human-computer interaction to create media content metadata. The research question was how media content metadata could be created at the time of capture leveraging both computing resources and human interaction.

More recent work related to camera phones is focused on what kind of pictures people take and with whom they share them. The approach has been to study literature and people's picture-taking habits in traditional photography as well as digital photography. Based on



these studies, we constructed a mobile picture sharing application *MobShare*. The system is currently going through user tests where we try to find out how the system was adopted by the users as well as how it fits into the users' picture-taking and sharing habits.

The group has joint research activities with the User Experience Research Group (UERG) and Digital Economy (DE) research group at HIIT. Co-operation with other Finnish research groups are in the following areas: sociology and computing at HUT / SoberIT, media and games research at University of Tampere / HyperMediaLab, psycho-physiological effects of media at HSE / CKIR, and media production at VTT.

In international collaboration, the most important partners are University of California, Berkeley, the MIT Media Lab, and the Media Lab Europe in Dublin. Two of the group's Ph.D. candidates, Risto Sarvas and Matti Rantanen, have been working in the Garage Cinema Group at SIMS, UC Berkeley under the supervision of Prof. Marc Davis.

All DCC research is done in active collaboration with industry partners and 10 companies are currently involved. This collaboration has been fostered by Special Interest Group activities, which include workshops and events related to the research topics in question. The SIG has approximately 30 active participants coming both from the industry and the academia.

#### *Highlights of recent achievements*

The research in this field is still at an early stage, and therefore most of the work has been in the areas of literature studies, scenario building, and product and system design. However, there have been also more concrete results achieved during the first year of activities.

In the area of enabling tools we have been active in building a community around a mobile application platform called Multi-User Publishing Environment (MUPE). This is an open-source effort initiated by Nokia Research Center and this work is done in collaboration with Nokia's MUPE team. We have established a node for MUPE developer community at [mupedev.org](http://mupedev.org).

In the area of mobile pictures, the main achievements are the construction of the MMM system in co-operation with SIMS at UC Berkeley (September 2003 in Berkeley) and the construction of the MobShare system in co-operation with Futurice (July 2004 in Helsinki). The MobShare system demonstrates how mobile camera phones are especially suitable for sharing pictures and discussing them. The MobShare system takes advantage of the



contextual and social information on the mobile phone to support familiar requirements in photography, for example, privacy, organization into albums, storytelling, and social interaction. The construction of the system also demonstrates a multi-disciplinary approach into system design: many of the requirements and functionalities of the system were gathered from ethnographic interviews and literature in sociology and humanities.

We have also started new research activities in understanding the potential benefits and problems in combining professional media content with open user-created content. This work is based both on previous HIIT research efforts related to Creative Commons licensing, and a new research project done in collaboration with VTT.

#### *Future directions*

The research aims to produce models of societal computing including analyzing, profiling and inter-connecting communities. How do people produce and share digital media in groups? What is the structure of relationships in a group vs. wider network? What are the processes communities use to communicate and learn? How do these processes impact not just the creation of communities but also their long-term viability?

In our future work, we will build on what we have learned in our observations of these communities to understand how different communities can pool resources and learn from one another as new tools are developed to connect them.

The DCC group will implement enabling open source tools that support content creation, especially in the area of mobile gaming. The tools will help to create any communities organized around mobile content, to support their collaborative activities, and to help the development of related business activities. The goal is to make the source code available early in the implementation process to the participating communities, so that the project can leverage on the community member's input and contributions in the development of the tools.

The DCC group is expected to grow by 3-4 researchers within the next two years. There have also been active discussions to deepen the collaboration between DE, UERG and DCC research at HIIT. For example, the legal issues related to community-created content provide a range of open questions necessitating joint research with the DE group, and collaboration with UERG is planned in the areas of group-centered ubicomp and psychological customization.



To further develop the collaboration with MIT Media Lab and Media Lab Europe, a new “Finland – Media Lab” research collaboration program has been planned. This would include mostly HIIT researchers in the pilot stage and is to be launched during the Fall of 2004

#### *Selected publications*

- 16) R. Sarvas, M. Viikari, J. Pesonen, H. Nevanlinna: MobShare: Controlled and Immediate Sharing of Mobile Images. In *Proceedings of Multimedia (MM 2004)*, New York, NY, USA. ACM Press (Forthcoming 2004).
- 17) R. Sarvas, E. Herrarte, A. Wilhelm, M. Davis: Metadata Creation System for Mobile Images. In *Proceedings of MobiSys 2004*, Boston, USA. ACM Press 2004.
- 18) A. Wilhelm, Y. Takhteyev, R. Sarvas, N. Van House, M. Davis: Photo Annotation on a Camera Phone. In *Proceedings of CHI 2004*, Wien, Austria. ACM Press 2004
- 19) M. Davis, S. King, N. Good, R. Sarvas: From Context to Content: Leveraging Context to Infer Media Metadata. In *Proceedings of Multimedia (MM 2004)*, Brave New Topics Session, New York, NY, USA. ACM Press (Forthcoming 2004).
- 20) T. Saari, M. Turpeinen: Towards Psychological Customization of Information for Individuals and Social Groups. In J. Karat, M-C. Karat (eds.), *Personalization of User Experiences for eCommerce*. Kluwer, Germany, 2004.

### **3.5 User Experience Research**

Director: Professor Martti Mäntylä. Senior researchers: Dr. Timo Saari.

#### *Premises of the research*

In the next decade, information and communication technologies will have the possibility to exploit a variety of technological advances. To name a few: increasing availability of digital recording devices, novel connectivity techniques and interfaces, increasing wireless bandwidth and mobile storage, sensor technologies and image recognition applications. On another front, interactive systems are challenged to move beyond current imperatives such as the desktop computer, mobile devices, information appliances, as new relations are suggested between computing, multimedia, physical environment and social space.

Currently, research in ubiquitous computing is focusing on developing technologies, showing what is possible but lacking proper field studies. HCI research, on the other hand,



tends to focus on existing technologies and provides a single perspective account (alternatively the social, the psychological, or the behavioural). In this context, human centred research should aim at anticipating change as a complex co-evolution of technology, social practices, and cognition.

Anticipating change in the ways people communicate with and through technological artefacts requires a multi-perspective approach. We approach the complexity of *interaction* from three different angles.

From a human individual perspective:

- Examining adaptation of cognition to changing conditions (psychology of skills and expertise)
- Studying *experiential responses* to new digital-physical artefacts (psychophysiology).

From social and interaction perspectives:

- Understanding the social and cooperative dimension of technology use. Analyzing the emerging and negotiated nature of user experience (ethnomethodology and conversation analysis)
- Investigating changing practices of *configuring-perceiving expressions* in communicative events (anthropology of performance).

From technology and artefacts perspectives:

- *Tailoring and Appropriation*, how digital-physical artefacts, their use and roles in human-human interaction evolve over time
- Evaluating changes in relations between ethics and *information objects* (aesthetics and ethics).

Scientific research includes not only observation, analysis and synthesis (or construction), but, especially in our field, should include *multiperspective* reflection understood as the “concurrent or alternating use of several perspectives in the consideration of a phenomena”<sup>1</sup>. This comprises, for example, “the study of how changes, introduced according to one viewpoint, affect properties of the phenomena when regarded from another

---

<sup>1</sup> Nygaard, K., Foreword, In: Ciborra, C., (2002) *The Labyrinths of Information, Challenging the Wisdom of Systems*, Oxford University Press, Oxford.





viewpoint". Following this idea, we choose to apply the perspectives described above through an interventionist approach to provide an effective anticipation. In these interventions, novel technological artefacts are introduced in specific settings. Our distinctive program is characterised by understanding how the abovementioned perspectives *intertwine*, because the realms addressed by them permeate each other and can no longer be considered separate.

As part of our studies of technology production and use we also reflect on methodological issues:

- Concept development and design process
- Qualitative studies (e.g. finding new ways for uncovering motivations and goals)
- Quantitative strategies to inform design (e.g. quasi-experimentation)
- Introducing research knowledge to industry.

The group aims at an outstanding and international academic impact through visibility in the best journals and conferences and by initiating research activities with the world's leading research groups. More importantly, the group aims at impacting thinking and practices industry through research partnership.

#### *Present status*

The User Experience Research Group was formed in 2000 by Professor Martti Mäntylä, combining skills of researchers from various disciplines ranging from engineering and industrial design to cognitive science and psychology. At the present, the group consists of six full-time and two part-time members and is co-led by Dr. Timo Saari and Professor Mäntylä.

Currently, UERG researchers work in five projects, in Context, MC2, Wireless Woodstock Services, MobiLife and Drama. Its research partners include CKIR at the Helsinki School of Economics, IERG at the Helsinki University of Technology, SIMS at the University of California at Berkeley, Nokia Research, HCI Institute at the Carnegie Mellon University, and the Department of Computer Science at the University of Jyväskylä, Vienna University of Technology.

#### *Highlights of recent achievements*

Findings:

- Describing planful opportunism in human mobility [22]





- Explaining attentional resources in mobile HCI
- Applying the anthropology of performance to the design and understanding of mixed media spaces [24]
- Explanation of interruption tolerance [23]

#### Methods:

- Quasi-Experimentation, analysing generalised causal inference for mobile field experiments
- Humanistic research strategy, organizing research methodology according to humanistic values [25]
- Bodystorming, method for carrying out concept innovation "in the wild"
- UCPCD, User Centred Product Concept Design
- Combining devised theatre and concept design
- Software tool for scenario analysis

#### Systems:

- InfoRadar, user Interface and functionality for location-aware communication [21]
- CoffeeMug, tangible container interface for digital documents
- ContextPhone, context cues for availability reasoning for user of smart phones

#### Awards:

- Pro Oeconomia 2004, business book of the year, honourable mention
- Best Paper Award Design 2004 Dubrovnik

#### *Future directions*

In the near future, the group aims at initiating a research project focused on a multiperspective approach to anticipating change in communicative practices mediated by novel technology. This project will make it possible to refine analytic and procedural strategies to combine the perspectives and will inform academic research and industry of relevant future changes, grounded in a multidisciplinary understanding. As an example the group is currently developing a new idea of Living Lab, where interventions can be organised in larger organisational and temporal scales, and where concurrent and alternate use of perspectives is more appropriate. Emerging topics of research are mobile cognition, "rhythmanalysis" and event structures in interaction, and roles of ubiquitous multimedia to create and share experiences. Current pending proposals include a European project that will investigate, for urban renewal projects, high-quality portable AR environments in support of situated participation of multi-disciplinary planning teams and citizens.



### *Selected publications*

- 21) M. Rantanen, A. Oulasvirta, J. Blom, S. Tiitta, M. Mäntylä: InfoRadar: Group and public messaging in mobile context. A full paper accepted to *NordiCHI'04*, November 2004, Tampere, Finland.
- 22) S. Tamminen, A. Oulasvirta, K. Toiskallio, A. Kankainen: Understanding mobile contexts. *Personal and Ubiquitous Computing*, 8 (4), 2004.
- 23) A. Oulasvirta & P. Saariluoma. Long-term working memory and interrupting messages in human-computer interaction. *Behavior & Information Technology*, vol. 23 (1), 53-64, 2004.
- 24) G. Jacucci & I. Wagner: Performative Uses of Space in Mixed Media Environments, To be published in: E. Davenport, S. Turner, P. Turner P., *Spatiality, Spaces and Technologies*, Kluwer Academic Publishers (in press).
- 25) A. Oulasvirta: Finding meaningful uses for context-aware technologies: The humanistic research strategy. *Proceedings of the 2004 Conference on Human Factors in Computing Systems*, Vienna, Austria, ACM Press, 247-254, 2004.

### **3.6 Digital Economy**

Director: Professor Jukka Kempainen. Senior researchers: Dr. Pekka Himanen, Ericsson Visiting Research Scientist Dr. Pekka Nikander, Professor Martti Mäntylä.

#### *Premises of the research*

After the burst of the “dot-com bubble” in 2000-2002, some observers hastened to predict that “digital economy” had essentially failed, and that the drivers of development of economy and society should henceforth be looked elsewhere.

We do not think so. Indeed, rapid technological progress of information and communication technologies (ICT), although perhaps taking a different route than predicted some years ago, still prevails and drives what now often is called *digital convergence*, a term roughly denoting the penetration of the Internet, digital media, and digital transactions to further and further areas of business, everyday life, and underlying fabric of all developed societies of the world. ICT is still a major source of competitiveness – but not necessarily so much on



the level of individual companies (that find it difficult to gain sustainable edge from the employment of ICT within their organisations) as on the level of networks and entire economies.

Yet there are some lessons that the last few years have taught us (or should have). The brave new digital world will not just push away the old world, but builds on top of it – just like industrial economy and society were built on top of agricultural economy and society, and had to be consolidated with it in a slow and painful process that still is in some ways incomplete. For example, well-hyped ubiquitous computing or ubicomp is actually creeping slowly and nearly unnoticed into our everyday life already. During such a process, the otherwise constant friction between “new” and “old” economy will be punctuated by rapid bursts of “sore points”. We believe that such process is grinding away now as well, and also to have identified some prime candidates of “sore points” that eventually must be treated.

Those sore points include, for example:

- Difficulties in the management of digital content and rights in the content; the imbalance between the interests of different parties, e.g. content creators, distributors, and end-users.
- Challenges related to trust, privacy, and security from different viewpoints, such as communities, social networks, and economy.
- The lack of *soft law*, meaning self-regulatory, easily moderated and enforced rules for peer groups and content produced or purchased by groups.
- Mobile, P2P, and other technologies that disrupt existing structures. In the future these technologies may include also, for example, ever-spreading ubicomp, RFID, short range communication, or some biometric identification technology that once and for all can destroy privacy.

The above themes form the general backdrop of the research of the Digital Economy group. In addition, and importantly, it is also shaped and toned by specific issues relevant to the Finnish society and economy that showed remarkable dynamism and growth during the 90's [26] but has faced considerable challenges in the last few years. Indeed, a major theme of more recent Finnish societal and political debate has revolved around the issue of whether it is possible to preserve the traditional Nordic “welfare state” and its values such as inclusive democracy, egalitarianism, social security, and transparency in a world increasingly



characterized by globalization, fierce international competition, and war against terrorism. A consensus seems to have been reached on the need of considerable changes in the way the public, private, and “citizen society” sectors co-operate to keep the wheels of a network economy turning – but where to start?

The mission of the Digital Economy group at HIIT is to study the legal, societal, and business issues suggested by the above themes, especially the rapid development of information and communication technologies that challenge the traditional ways to structure, organize, analyze, and regulate the activities in a society. The research interests of HIIT’s DE group are related to solving these problems, especially from companies’ and policy makers’ viewpoint. We focus specifically on key “sore points” such as rights to digital information; unconventional agents (such as virtual communities and interest groups) acting in various roles in networked economy; trust and privacy; and also key technologies such as peer-to-peer systems that challenge prevailing business models and logic. We also study the overall logic and dynamics of the network society to maintain a coherent framework for all these phenomena.

#### *Present status*

The DE group at ARU was founded in 2000 although it has longer roots at Helsinki University of Technology. The group consists at present of 12 researchers, including software engineers and computer scientists, lawyers, a social philosopher, and an economist. Professor Jukka Kempainen is the Principal Scientist of the group, and Olli Pitkänen is the program coordinator. The group’s strengths include especially intellectual property rights; digital rights management; open source and content licensing; security, trust, and privacy.

#### *Highlights of recent achievements*

DE group has published a number of papers and reports including several journal articles. The researchers have also developed working software prototypes to study digital rights management (DRM) [29], peer-to-peer (P2P) news sharing, and multimedia distribution in P2P networks [28].

Digital Content Distribution Management System, *DiMaS*, demonstrates the concept that it is possible to make a system for multimedia producers to publish their work on highly popular P2P file sharing networks, enabling producers to insert content metadata, to manage intellectual property and usage rights, and to charge for the consumption. [28] DE



Core research project has also analyzed the legal implications of DiMaS used in a content creating community.

The legal protection of intellectual property is one of the most controversial issues in relation to peer-to-peer networks. Commercial content providers are extremely worried about the unauthorized copying and distribution of their products, and they insist on strong legal and technical protection of their intellectual property. On the other hand, in most cases authors do not actually need full protection for their works, but they wish to easily share their content on the Net and get credit for their creations. Stanford University based *Creative Commons* uses private rights to create public goods: creative works set free for certain uses. Like the free software and open-source movements, their ends are cooperative and community-minded, but their means are voluntary and libertarian. Creative Commons work to offer creators a best-of-both-worlds way to protect their works while encouraging certain uses of them — to declare "some rights reserved." Creative Commons is working with the HIIT's DE Group on porting the Creative Commons licenses to Finland. The Finland license has been integrated into the Creative Commons licensing process, so rights-holders are able to license their works under Finland's law. (<http://creativecommons.org/>)

Although Creative Commons can be applied even to commercial content distribution, it is especially suitable for communities' *open content* production that creates information products to be shared without monetary compensation. It emphasizes attribution and other moral rights of the author, which are central in open content communities.

In the software engineering world, open source communities resemble above mentioned open content communities. In the recent years, open source software development has challenged traditional software companies. In the wake of popular Linux operating system, a number of other open source software products have gained recognition. In addition to Linux, Finland is also the home country of many other important open source programs, like MySQL database engine. *RIPOS* project is studying, how open source software is utilized in the different sectors of the society, and which obstacles prevent wider usage of open source software.

*The Welfare of Nations* was a HIIT research project funded completely by TEKES, the National Technology Agency in Finland. The two-year project studied the interaction of technology, economy, and society through a comparison of three successful information societies: Silicon Valley / USA, Singapore / Greater China, and Finland / EU. The project



was carried out in close co-operation with the Berkeley Center for Information Society at the University of California, Berkeley. [26]

Recently, Dr. Pekka Himanen wrote for the Finnish Parliament, the Committee for the Future, a report titled “Challenges of the Global Information Society”. The widely acclaimed report looks at the big challenges that are going on in the information society especially from Finnish perspective and gives several proposals for policymakers.

#### *Future directions*

In the next few years, DE group is going to focus on the following topics:

- *The structures of the network society*, including value networks, transaction costs, relations between actors, pricing mechanisms, and business models.
- *Communities*; intellectual property that is created, improved, transferred, or consumed within a community; privacy and the distinction between private and public in communities; the legal essence of communities, legal actions by communities, and the rules that govern communities.
- *Managing metadata enriched multimedia distribution* for heterogeneous terminals – especially digital television and mobile devices – on unmanaged peer-to-peer networks.
- *Open source* and *auto-organization* as a modus operandi, development model, licensing models, economic analysis of copyright, incentives, and user rights; seeing the tendency into auto-organization as for example in peer-to-peer groups and non-economic endeavours (Gnu, Linux), the group participates in developing auto-regulative rules and mechanism of Creative Commons.
- *Society models* of information and welfare societies, digital sociology and politics, international comparison.
- *Privacy, trust, and economy* in P2P content distribution.

The co-operation with other research institutes, especially with University of California, Berkeley, School of Information Management and Systems (SIMS), and Lappeenranta University of Technology, will keep on going actively.



*WWSF* is the Finnish part of international Eureka/Celtic project called Wireless Festival (<http://projects.celtic-initiative.org/DB/>). The project is studying, prototyping, and evaluating solutions for large-scale events, and assessing the business, user, service and communication aspects. For Digital Economy group, these large events represent an interesting example of novel structures and communities, as well as service and communications needs.

An important feature of the network society is the emergence of various new types of informal or semi-informal organisations that represent horizontal networks of people sharing a common interest, participating in some process of information creation or sharing, or providing a service. Peer-to-peer networks, despite their sometimes disputed legal status, give an important example of such organisations. Therefore, a major line of future research of the group will be devoted to studying certain important aspects of peer-to-peer networks and the issues they raise.

*P2PdigiTV* project will study multimedia distribution for heterogeneous terminals (esp. digital TV and mobile devices) on peer-to-peer networks. It will develop a model to show how to distribute information products on a P2P network efficiently, securely, honouring rights, and in a commercially viable way. The system extensively exploits metadata. It enables content producers to manage intellectual property and usage rights, and to charge for the consumption in an easy way without introducing new file formats or dedicated client applications. The project will produce a working prototype and publications, as well as organize seminars. The wide cooperation network, including top-level universities both in Finland and abroad assures enough capabilities to complete the demanding and risky subject.

Digital Content Distribution Management System, *DiMaS*, has already proven a concept to enable multimedia producers to publish their works and metadata on P2P file sharing networks while still managing intellectual property rights. [28] Yet, *DiMaS* is to be further developed: currently DE group is studying e.g. the distribution of multimedia content to mobile devices and digital TV set-top boxes with the help of new versions of *DiMaS*.

A major issue limiting the potential of peer-to-peer systems is the level of trust that the members of the network feel towards each other. The recently launched Muppet project will study trust formation, expression, and use in peer-to-peer applications by looking into various important scenarios of peer-to-peer networks. In particular, we aim to build some form of economic incentives in the trust model so that it “pays” to users to behave in a trust-





enhancing way while using the peer-to-peer application. The first prototype application presently build will focus on a “yellow market” of children’s clothes and gear.

The planned project *Global Network Society* continues research on the interaction of technology, economy, and society by focusing in the dynamics of the information society development in China. In particular, the project will assume the viewpoint of foreign companies who are interconnected with the developing production and innovation networks in China. The second part of the project will study the impact of Chinese information society development at global scale. The project is planned to be implemented by Dr. Pekka Himanen in co-operation with Professor Manuel Castells, USC, Professor You-tien Hsing, Berkeley BCIS and Professor Martin Carnoy, Stanford University.

#### *Selected publications*

- 26) P. Himanen, M. Castells: *The Information Society and the Welfare State: The Finnish Model*. Oxford University Press, 2002.
- 27) O. Pitkänen, M. Mäntylä, M. Välimäki, J. Kempainen: Assessing Legal Challenges on the Mobile Internet. *International Journal of Electronic Commerce*, Fall 2003, Vol. 8 (1): 101-120, 2003.
- 28) T. Reti, R. Sarvas: The DiMaS System for Authoring Communities: Distributing Semantic Multimedia Content on Peer-to-Peer File Sharing Networks. In *Proceedings of Web Intelligence Symposium 2004*, Vantaa, Finland, September 2004.
- 29) A. Soininen (ed.), O. Pitkänen, M. Välimäki, V. Oksanen, T. Reti: *MobileIPR Final Report*. HIIT Publications 2003-3, 2003.
- 30) M. Välimäki: Dual Licensing in Open Source Software Industry. *Systemes d’Information et Management*, Vol. 8 (1): 63-75, 2003.

### **3.7 Data Mining: Theory and Applications**

Directors: Professor Heikki Mannila, Professor Hannu Toivonen, Dr. Jaakko Hollmén. Senior researchers: Dr. Ella Bingham, Dr. Aristides Gionis, Dr. Alexander Hinneburg, Dr. Saara Hyvönen, Dr. Mikko Koivisto, Dr. Marko Salmenkivi, Dr. Panyiotis Tsaparas





### *Premises of the research*

The importance of data analysis has grown in recent years, as the ability to collect and store scientific and industrial data has grown. Novel measurement methods yield large high-dimensional data sets, and the data is often in non-traditional forms (e.g., strings, trees, graphs). Data mining has emerged as a major research area in the interface of computer science and statistics, and data analysis questions are increasingly visible in database and algorithms research.

While the data analysis problems arise from applications, there are interesting general computational issues in the background. The goal of the research in the theory and applications of data mining in HIIT BRU is to develop novel data analysis techniques for the use of other sciences and industry. The research looks at data analysis problems arising in practice, abstracts new computational concepts from them, analyses the concepts and develops new computational methods, and takes the results into the practice. Theoretical work in algorithms and foundations of data analysis can have a large and fast impact in the application areas, while the applications feed interesting novel questions to the theoretical research. For example, in the work on molecular genetics we develop computational methods for genetic association analysis, haplotype analysis and reconstruction, utilisation of haplotype blocks and SNPs. This work has immediate, important applications: locating genes that predispose to diseases is essential for understanding the etiology of complex common diseases, such as heart disease or asthma. The research continues to build on our strong connections with top geneticists.

### *Present status*

The data mining research at HIIT BRU studies both the theory and applications of data mining methods. The applications are mostly from other sciences, while some work is done in industrial collaborations. Currently, the main themes in the research development of methods are pattern discovery and combining it with classification, methods for sequence decomposition, the study of the interplay of combinatorial and continuous methods in data mining, and techniques for the decomposition of large 0-1 data sets.

The main application theme of the research in data mining is bioinformatics, especially medical genetics; the applications span gene mapping, discovery of genome structure, and gene expression data analysis. Other application themes include palaeontology, linguistics, and ubiquitous computing. Especially in ubiquitous computing area the applications work



has lots of connections with other themes in HIIT (see the sections on adaptive computing systems and intelligent systems).

In the area of medical genetics, the main topics of study and their methodological core tasks are described below.

- 1) Gene mapping using association analysis. For a computer scientist, this problem can be described as a mixture of pattern discovery and classification
- 2) Haplotyping from population-based genotype data: a question of finding the highest probability strings (haplotypes) explaining sequences of pairs (genotypes).
- 3) Defining and utilising haplotype block structure of the human genome: describing and finding the possible mosaic-like structure of haplotypes (and genotypes).
- 4) Analysis of phenotypic data sets: finding robust clusters that have genetic explanations

In paleontology, one of the problems considered is

- finding good estimates of the ages of fossil sites: computationally this is equivalent to finding matrix reorderings that approximate the consecutive ones property as well as possible.

In linguistics, a typical problem is

- finding spatial structure of the distribution of place names and dialect words: a question of high-dimensional clustering.

In ubiquitous computing, an example problem is

- learning to recognize typical device contexts: on-line clustering of stream data.

The data mining area is investigated by three senior researchers (Prof. Heikki Mannila, Prof. Hannu Toivonen, and Dr. Jaakko Hollmén), seven postdocs (Ella Bingham, Aristides Gionis, Saara Hyvönen, Mikko Koivisto, Marko Salmenkivi, Panyiotis Tsaparas, and Alexander Hinneburg), and 9 Ph.D. students. The research done at HIIT BRU has a visible position in the international community, and the results have attracted a lot of attention. The work is organized loosely into three groups, one for each of the most senior researchers, but cross-group collaboration is very common.



### *Highlights of recent achievements*

Recent highlights of the results of the research include the following.

- A method for finding recurrent sources in sequences. The motivations for the problem is finding global structure in genomic sequences, and recognizing recurrent contexts in mobile device usage.
- Finding orderings of attributes from unordered binary data using combinatorial and spectral techniques. The problem stems from the question of dating fossil sites on the basis of taxonomic information, and from trying to find succinct representations of large 0-1 data sets.
- A new method for haplotyping, based on variable length Markov chains. The method is superior to existing methods in the difficult cases, i.e., for genetically large regions, both in terms of accuracy and computational efficiency, and on a par for the easier settings.
- Large simulation studies for the analysis of genotype vs. haplotype trade-off in gene mapping. The results are novel and unexpected.
- Preliminary results on pattern discovery and mixture modelling techniques for large onomastic data sets, especially on place names in Finland.
- Techniques for discovering uniform subsets from 0-1 data using spectral methods.

The current approach of pursuing both theory and applications of data mining has been very successful, and we intend to continue it. For clarity of presentation, however, we will in the following discuss these two aspects separately.

### *Theory of data mining / future directions*

On the theoretical side, some of the main issues are the following.

- The combination of continuous and combinatorial methods, for example in finding out what techniques could be used to infer distributions from collections of pattern frequency information.



- Concepts and algorithms for describing structure of sequences. Here techniques such as intensity modelling, combinatorial and probabilistic segmentation, and mixture modelling will play an important role.
- Methods for pattern discovery in and modelling of spatiotemporal data.
- Theoretical models for data mining (such as inductive databases).

Foundational issues in pattern discovery; e.g., investigating the relationship between the logical form of patterns and the difficulty in discovering them.

*Theory of data mining / selected publications*

- 31) D. Gunopulos, R. Khardon, H. Mannila, S. Saluja, H. Toivonen, R.S. Sarma: Discovering all most specific sentences. *ACM Transactions on Database Systems* 28, 2 (June 2003), 140 – 174.
- 32) A. Gionis, T. Kujala, H. Mannila: Fragments of order. In *Proceedings of ACM SIGKDD 2003*, P. Domingos, C. Faloutsos, T. Senator, H. Kargupta, L. Getoor (eds.), ACM 2003, p. 129-136.
- 33) F. Geerts, H. Mannila, E. Terzi: Relational link-based ranking. *The 30th International Conference on Very Large Data Bases (VLDB'04)*, 2004
- 34) F. Afrati, A. Gionis, H. Mannila: Approximating a collection of frequent sets. *Proceedings of ACM SIGKDD 2003*.
- 35) A. Gionis, H. Mannila, J. Seppänen: Geometric and combinatorial tiles in 0-1 data. *Proceedings of PKDD 2004*.

*Applications of data mining / future directions*

There are several interesting areas and problems in the applications.

- Genome structure and its relation to function will continue to be one of the main themes. Can we locate segments of genomic sequences stemming from different sources? What are the best ways of describing intra- and interspecies variation? More generally, systems biology will lead to lots of demand for computational techniques. The work on linguistics applications will be extended to the study of spatial and temporal variation in language.



- Work on ubiquitous computing and telecommunications applications will continue. One of the key questions will be the recognition of contexts in mobile applications.
- The paleontological and ecological work will also consider spatiotemporal issues: what types of methods are needed to detect patterns in the variation of, e.g., species abundances.
- Bio-database mining: gene mapping may result in a genomic area containing hundreds of genes, and the goal is to develop methods for mining public databases for information about those genes and for ranking the candidate genes by their interestingness. This information retrieval involves non-trivial similarity measures to find connections between concepts.

#### *Applications of data mining / selected publications*

- 36) A. Gionis, H. Mannila: Finding recurrent sources in sequences. In W. Miller, M. Vingron, S. Istrail, P. Pevzner, M. Waterman (eds.), *Proceedings of ACM ReCOMB 2003*, p. 123-130.
- 37) L. Eronen, F. Geerts, H. Toivonen. A Markov chain approach to reconstruction of long haplotypes. Pacific Symposium on Biocomputing (PSB 2004), 104-115, Hawaii, USA, January 2004. World Scientific.
- 38) P. Sevon, H. Toivonen, V. Ollikainen: Tree pattern mining for gene mapping. Accepted for publication in *Information Sciences*.
- 39) P. Sevon, H. Toivonen, P. Onkamo: Gene Mapping by Pattern Discovery. In J. Wang et al. (eds.), *Data Mining in Bioinformatics*, Springer 2004.
- 40) H. Wikman, J.K. Seppänen, V. K. Sarhadi, E. Kettunen, K. Salmenkivi, E. Kuosma, K. Vainio-Siukola, B. Nagy, A. Karjalainen, T. Sioris, J. Salo, J. Hollmén, S. Knuutila, S. Anttila. Caveolins as tumor markers in lung cancer detected by combined use of cDNA and tissue microarrays. *Journal of Pathology*, Vol. 203, pp. 584-593, 2004.

### **3.8 Adaptive Computing Systems**

Directors: Dr. Patrik Floréen, Prof. Hannu Toivonen. Senior researchers: Dr. Greger Lindén.



### *Premises of the research*

Adaptive computing focuses on the methodology and implementation of systems that adjust to different situations relating to the user(s), the device(s) and the environment. Adaptive computing plays an important role in ubiquitous and pervasive computing, as well as in intelligent and user-friendly applications, and is thus also of high practical and industrial importance.

The activities of HIIT's research theme Adaptive Computing (AC) can be divided into two areas: context-aware computing and ad hoc networks.

In context-aware computing, the (present, past and predicted future) situation is used to make applications more useful and easier to use for the user. Context data can be made available to other people, making communication between peers more efficient. Context-aware computing is challenging, since practically any information can be context, i.e. characterize the situation of an entity. The context data may be incomplete, inconsistent, erroneous, and privacy-sensitive.

Ad hoc networks are data communication networks with no predefined infrastructure, typically requiring multiple hops for connecting all the nodes to each other. A special case of wireless ad hoc networks are sensor networks, where a large number of low-cost, low-power sensor nodes collect information. Sensor networks have many applications in logistics, health telemonitoring, material fatigue detection etc. The communication and computation in ad hoc and sensor networks are difficult to model, as the network nodes typically have limited computing, memory and communication capacities, as well as a limited amount of energy available. Also, forwarding the data using multi-hop wireless communication is complex, especially when the network changes due to node failure and mobility.

The research will contribute to how context-aware systems are built and reasoning about contexts performed. Furthermore, it will provide fundamental theoretical results for ad hoc and sensor networks.

### *Present status*

The work combines theoretical and application-oriented work, including software implementation. The main research questions studied are:



- Characterization and analysis of context information on mobile devices and the use of context in proactive adaptation:
  - How to recognise user-dependent important locations from GSM cell data and predict the target location?
  - How to analyse context streams under limited resource requirements?
  - How to build a real context-triggered presence service for standard mobile phones?
- Context-aware selection of software components on wireless devices:
  - How to adapt existing component-based software architectures to cope with context-awareness?
- Topology control and routing in ad hoc networks, as well as data gathering in sensor networks:
  - How to maximize the multicast and broadcast connectivity lifetime of ad hoc networks under energy constraints?
  - How to maximize the amount of gathered data in sensor networks, while at the same time ensuring a minimum amount of received data from all parts of the covered area?
- Self-organisation of ad hoc networks:
  - How to model cooperative behaviour using game theory?

There are two groups participating: one lead by Dr. Floréen on context-aware computing and ad hoc networking with four PhD students, and one lead by Prof. Toivonen on context-aware computing with three PhD students. The expertise represented includes combinatorics, optimization, data analysis, machine learning, language technology and component architectures.

In context-aware computing, Prof. Toivonen's group collaborates with HIIT/ARU. Dr. Floréen's group participates in a large EU IST integrated project (MobiLife) and is subcontractor to Nokia in an ITEA-project (Space4U). In the field of ad hoc networks, Dr. Floréen's group presently collaborates with two groups at HUT, most closely with the Laboratory for Theoretical Computer Science.

From a science policy point of view, HIIT/BRU has a prominent position in the area of ubiquitous computing in Finland, as coordinator of the Academy of Finland's research programme on Proactive Computing (PROACT, 2002-2005).



### *Highlights of recent achievements*

In the area of context-aware computing, the main results are concepts and data analysis algorithms for the discovery and recognition of important locations for the user, based on the user's movement through the GSM cells, and prediction of the target location [41]. This is used for a presence service and for automatic annotation (e.g. "at home", "in Kumpula") of photographs taken with a mobile phone. This fully functional system for automatic presence service has been a major software production effort. A field study of the prototype system on Nokia Series 60 mobile phones has just been completed, and we are in the process of analysing the results. The software has been published as open source, and is used by a number of research groups (MIT Media Lab; SIMS/U. Berkeley; Oslo National College of the Arts; University of Art and Design in Helsinki). Furthermore, a blackboard architecture has been designed for the context-aware re-configuration framework for ROBOCOP terminals [42].

In the area of ad hoc networks, one main achievement is NP-hardness results and algorithms for multicast lifetime maximization under energy constraints, by dynamically choosing transmission power levels for the network nodes [43]. Another achievement is a notion of balanced data gathering in sensor networks, with algorithms for optimal data routing, also when introducing additional auxiliary relay nodes to the network [44]. Recent work includes also modelling of the cooperation of nodes for routing, using dynamic Bayesian games [45].

### *Future directions*

The AC research theme lies at the intersection of many of the research groups of HIIT and the Department of Computer Science at UH, since many other groups (of Professors Kaski, Myllymäki/Tirri and Raatikainen) also deal with context-awareness, personalisation and adaptation. We therefore aim to take better advantage of the existing in-house expertise in data analysis and probabilistic reasoning, through joint projects with these groups. Such joint work can be in the form of TEKES projects, which also include companies.

Research in context-aware computing typically requires system implementations. Both theoretical and practical work will thus be pursued, including software development. An obstacle for work in context-aware computing is the lack of real context data, which is needed for development and evaluation of context reasoning methods.





The main research issues that we plan to study over the next years are:

- Context reasoning and the use of context:
  - How to deduce higher-level context from lower level context?
  - How to infer and present meaningful contexts to the user?
  - How to predict future contexts and to detect changes in context?
  - How to detect the need for changes to the model used for the context reasoning, and to update it using machine learning methods?
  - How to use the context for adaptation and proactive services?
- Trust and privacy issues of the users in context-aware applications:
  - What practical and theoretical requirements does privacy preservation impose on data analysis methods and systems design?
  - How do people actually perceive the risks and benefits of systems that are aware of personal data?
- Combining context-awareness and component architectures:
  - How to use a component-based approach for the development of context reasoning mechanisms?
  - How to extend the notion of component-based software with context-awareness?
- Modelling and algorithms for topology control and routing in ad hoc networks and data gathering in sensor networks:
  - How to choose well-founded model parameters for real networks?
  - What effect does the choice of model parameters have on the construction of algorithms?
  - How to implement topology control and routing using distributed algorithms?
  - How to perform routing using on-line algorithms?
- Self-organisation of ad hoc networks:
  - How to model cooperative behaviour using game theory?

The knowledge and skills within these topics will be enhanced through internal seminars and normal research training of the PhD students, through collaboration with other groups inside and outside of HIIT and through the recruitment of complementing expertise, with particular attention to the recruitment of postdocs.



### *Selected publications*

- 41) K. Laasonen, M. Raento, H. Toivonen: Adaptive On-Device Location Recognition. In Proc. 2<sup>nd</sup> Intl. Conf. on Pervasive Computing (PERVASIVE, Vienna, Austria, April 2004), *Lecture Notes in Computer Science 3001*, 287-304. Springer, Berlin, 2004.
- 42) J. Muskens, O. Virtanen, M. Chaudron, R. MacLavery: Maintaining Terminal Integrity and Context-Aware Reconfiguration. In workshop *Component-Oriented Approaches to Context-Aware Computing at the European Conference on Object-Oriented Programming* (ECOOP, Oslo, Norway, June 2004).
- 43) P. Floréen, P. Kaski, J. Kohonen, P. Orponen: Multicast Time Maximization in Energy Constrained Wireless Networks. In Proc. *DIALM-POMC Joint Workshop on Foundations of Mobile Computing* (San Diego, Calif., Sept 2003) at MobiCom 2003, ACM, 50-58, 2003.
- 44) E. Falck, P. Floréen, P. Kaski, J. Kohonen, P. Orponen: Balanced Data Gathering in Energy-Constrained Sensor Networks. In Proc. 1<sup>st</sup> Intl. Workshop on Algorithmic Aspects of Wireless Sensor Networks (Algosensors, Turku, Finland, July 2004). *Lecture Notes in Computer Science 3121*, 59-70. Springer, Berlin, 2004.
- 45) P. Nurmi: Modelling Routing in Wireless Ad Hoc Networks with Dynamic Bayesian Games. In Proc. *1st IEEE Intl. Conf. on Sensor and Ad Hoc Communications and Networks* (IEEE SECON, Santa Clara, Calif., Oct 2004), IEEE, 2004.

### **3.9 Neuroinformatics**

Director: Dr. Aapo Hyvärinen. Senior researchers: Dr. Patrik Hoyer, Dr. Jarmo Hurri.

#### *Premises of the research*

Neuroinformatics is broadly defined as the intersection of Information technology and neuroscience. The two-fold goal in neuroinformatics is to import concepts and technologies from IT to benefit the neurosciences, and to take ideas and concepts from the neurosciences to advance information technology. Both computer science and the neurosciences are rapidly developing disciplines, so opportunities for great scientific and technological discoveries are remarkable. While neuroscience has become more and more dependent on computational models, we have seen that computational models developed in the neurosciences can be applied to a wide variety of information processing tasks.



### *Present status*

Our current research goals are:

1. To build statistical models of sensory processing in the brain, in particular in the visual cortex. Our basic approach is to consider how sensory processing is adapted to the statistical properties of the environment, i.e., the natural input that the sensory system receives.
2. To develop new multivariate statistical models. A fundamental mathematical method that we use is independent component analysis (ICA) and some of its extensions. New models are needed for achieving the other goals, and they are often interesting in their own right.
3. To apply advanced statistical methods on neuroscientific data. Usually these models are the same or very similar to the models used for modelling sensory processing in point one above, so obvious synergy benefits can be found. Further synergy can be found if the neuroscientific data is measured from the sensory processing areas of the brain.

The group consists of the leader, Dr. Aapo Hyvärinen, currently with two post-docs, four PhD students, and a variable number of international visitors. This group was established at HIIT BRU in 2003, and most of the members joined the group in 2004.

Collaborative partners include (for each goal, respectively):

1. Computational Vision Lab at New York University and Riken Brain Science Institute (Japan)
2. Neural Networks Research Centre at HUT, Division of Mathematical Science at Osaka University, CoSCo group at HIIT
3. Department of Psychology at UH, Department of Neurological Sciences at the University of Naples, Department of Neurocognition at Maastricht University

### *Highlights of recent achievements*

In collaboration with the Neural Network Research Centre of HUT and the Department of Neurological Sciences of the University of Naples, we developed a method for analyzing the reliability (stability) of independent components. So far, algorithms for estimating independent components have simply given estimates of the components with no



information on how much statistical or computational error (local minima) the results might contain. Analysis of reliability is obviously of great importance in the application of ICA [46].

New models for the statistical structure of natural images have been developed. While previous research has concentrated on neural networks with one or two layers to analyze images, we have extended this analysis to three layers. This analysis shows how to compute highly nonlinear features of images so that the obtained features are statistically optimal. Also, a first model of the statistical basis of Gestalt phenomena, i.e., segmentation and perceptual grouping, has been developed. This is the very first model to attack this difficult problem (two manuscripts just submitted).

We have also investigated new, related applications for these statistical methods. First, we applied the same statistical models to the analysis of natural language. This leads to the emergence of syntactic and semantic categories [47]. Second, we have attempted to characterize whole images instead of small image patches using these models. While these results are still very preliminary, they are quite promising [48] and may be useful for application related to image retrieval.

#### *Future directions*

Theory of ICA and related topics is a direction that we certainly cannot overlook due to its importance in our main applications (image modelling, neuroscientific data analysis). While the estimation of basic ICA is a mature topic, different kinds of “post-processing” techniques may offer completely new vistas for data analysis. This means methods that take the ICA decomposition as input and either analyze the properties (such as stability) of the decomposition, or use it as input for further analysis. Also, different violations of the basic model assumptions need to be considered.

For example, we have recently started a new direction which is based on the analysis of causality [49], in collaboration with researchers from the Division of Mathematical Science of Osaka University. Using the higher-order structure of data analyzed by ICA, it is sometimes possible to infer the causal direction, i.e., whether the value of the variable  $X$  is caused by  $Y$  or vice versa. The stability analysis of the components [46] is another post-processing method that still needs to be developed further.



Other important theoretical directions are to incorporate the time structure of the components in the model, and to model the dependencies of components that turn out not to be independent as in the basic model specification [50].

We are also continuing the longer-term subproject of estimating multi-layer statistical models for natural images. This is an important topic for neuroscience, where the statistically optimal representation gives an interesting hypothesis for experimental neuroscience: Lack of good hypotheses is one of the main problems in current visual neuroscience. On the other hand, from the viewpoint of image processing, the models show what kind of nonlinear features would be useful to improve the performance of current methods for compressions, denoising, and image retrieval.

We have also started a new collaborative project with the department of psychology where we attempt to develop further the method called “classification images” that has recently attracted a lot of attention in vision science.

Ultimately, the research will fan out to meet many other principles of data analysis and machine learning. The topographic version of ICA, which we developed a couple of years ago, is one example, consisting of a combination of ICA and Kohonen maps. Our recent research on Gestalt (segmentation) phenomena is another example. Techniques related to nonlinear projections are another group of methods that, when combined with ICA would lead to most interesting new methods.

#### *Selected publications*

- 46) A. Himberg, A. Hyvärinen, F. Esposito: Validating the independent components of neuroimaging time-series via clustering and visualization. *NeuroImage* 22(3):1214-1222, 2004.
- 47) T. Honkela, A. Hyvärinen: Linguistic Feature Extraction using Independent Component Analysis. *Proc. Int. Joint Conf. on Neural Networks (IJCNN2004)*, Budapest, Hungary, July 2004.
- 48) J. L. Lindgren, A. Hyvärinen: Learning high-level independent components of images through a spectral representation. *Proc. Int. Conf. on Pattern Recognition (ICPR2004)*, Cambridge, UK, August 2004.



- 49) S. Shimizu, A. Hyvärinen, Y. Kano: Finding a causal order using independent component analysis. *Proc. Factor Analysis centennial Symposium*, Osaka, Japan, October 2004.
- 50) A. Hyvärinen, J. Hurri: Blind separation of sources that have spatiotemporal dependencies. *Signal Processing*, 84(2): 247-254, 2004.