Department of Computer Science
Series of Publications A
Report A-1998-1

Computer Science at the University of Helsinki 1998

Greger Lindén and Martti Tienari (eds.)

University of Helsinki
Finland

DEPARTMENT OF COMPUTER SCIENCE
SERIES OF PUBLICATIONS A
REPORT A-1998-1

Computer Science at the University of Helsinki 1998

Greger Lindén and Martti Tienari (eds.)

UNIVERSITY OF HELSINKI
FINLAND

Contact information

Postal address:
Department of Computer Science
P.O. Box 26 (Teollisuuskatu 23)
FIN-00014 University of Helsinki
Finland

Email address: postmaster@cs.Helsinki.FI (Internet)

URL: http://www.cs.Helsinki.FI/

Telephone: +358 9 708 51

Telefax: +358 9 708 44441

ISSN 1238-8645
ISBN 951-45-8138-5
Computing Reviews (1991) Classification: A.1, A.2, K.3.2
Helsinki 1998
Helsinki University Printing House

Computer Science at the University of Helsinki 1998

Greger Lindén and Martti Tienari (eds.)

Department of Computer Science
P.O. Box 26, FIN-00014 University of Helsinki, Finland
Greger.Linden, Martti.Tienari}@cs.helsinki.fi

Report, Series of Publications A, Report A-1998-1
Helsinki, May 1998, 112 pages
ISSN 1238-8645, ISBN 951-45-8138-5

Abstract

The Department of Computer Science at the University of Helsinki is the largest such department in Finland. In this report we present the activities at the department. We describe ongoing research and research projects as well as give a list of recent publications of the faculty according to their classification. We also present the faculty members and note their research interests as well as publications and contact information. We describe the educational program, graduate schools and courses taught at the department and also give a list of accepted theses and abstracts of the most recent Ph.D. theses. We give short descriptions of the department library and the computing facilities, and finally present an overview of our international relations.

Computing Reviews (1991) Categories and Subject Descriptors:

A.1	Introductory and Survey
A.2	Reference
K.3.2	Computers and Education: Computer and Information Science Education

General Terms:

Additional Key Words and Phrases:
computer science education, research information

Contents

Chapter 1

Overview

The Department of Computer Science was founded in 1967 when the first full professorship in computer science was established at the University of Helsinki. It is within the Faculty of Science, along with the departments of Mathematics, Physics, Chemistry, and others.

Our current teaching faculty (May 1998) comprises 46 full time teachers. It can be categorized using American academic terminology: 5 full professors (Mannila, Tienari, Ukkonen, two vacant positions), 4 associate professors (Mäkelä, Paakki, Raatikainen, Sippu), 23 assistant professors (9 research oriented senior assistants and 14 teaching oriented lecturers) and 14 teaching assistants. We also employ approximately 40 of our students in teaching on a part-time basis. Fourteen senior experts are also associated with the department. These so-called docents work mainly outside the university but occasionally give courses or supervise theses in the area of their speciality. Approximately 60 research positions are financed from outside sources. We also have a staff of 12 persons.

The department annually admits 240 students to major in computer science. The students are selected according to their standing in a national student examination or in a special entrance examination (or both). The number of completed M.Sc. degrees (5 year degree) was 44 in 1995, 51 in 1996, and 62 in 1997. The study time for a M.Sc. degree ranges from five to eight years. Many of our students work in industry, which slows down or stops the progress of their studies. Fairly frequently our students, after having acquired the basic skills in computer science, redirect their studies by transferring to the Helsinki University of Technology, the Helsinki School of Economics, Faculty of Social Sciences, or some other educational institution. Many students study computer science as a minor while pursuing a major in another subject, such as mathematics, physics, economics, psychology, or social sciences. We offer two curricula for students minor-ing in computer science. In 1997 our "approbatur curriculum in computer science" (15–34 credit units) was completed by 141 students and our "cum laude curriculum in computer science" (35–69 credit units) was completed by 117 students.

There are two graduate degrees in Finland: the Ph.Lic. (3 years) and the Ph.D. degree (4 years). The latter has higher quality requirements. Both build upon the M.Sc. degree (5 years). The high demand for our M.Sc. graduates to fill well-paid jobs in industry is a fact which has hampered our Ph.D. education. Our department granted five Ph.D. degrees and five Ph.Lic. degrees in the two year period 1996–97. In postgraduate education we cooperate with the Helsinki University of Technology within the Helsinki Graduate School in Computer Science and Engineering (HeCSE) and with the University of Turku and the Center for Scientific Computing of Finland within the Graduate School in Computational Biology, Bioinformatics, and Biometry (ComBi).

Three principal sources provide funds for computer science research in Finland. The Academy of Finland under the Ministry of Education and Science provides funding for basic research. The second important research financier is the Technology Development Centre (TEKES) under the Ministry of Trade and Industry. We also enjoy financing of the European Commission in the research and development programs ESPRIT and ACTS.

The department maintains jointly with the University Computing Centre a good computer science library. It subscribes to most major international journals in computer science and related fields and acquires a majority of the most important computer science books and conference publications. The library is run by a librarian and a secretary.

The htmladdnormallinkUniversity Computing Centrehttp://www.helsinki.fi/atkk/ maintains a communication backbone network and offers UNIX and PC services. In addition, the department maintains its own workstation network of approximately 280 Linux PCs and about 10 servers. Windows 3.1, Windows 95 or Windows NT can be used as an alternative for Linux. About 30 of the Linux workstations are mobile laptops which can join and leave the network dynamically. Classrooms with Linux PCs are available. Each office of the department has a PC or a workstation.

The department has three informal sections that are used in the planning of the curricula and in administration. The division is not strict, and several research projects span two sections. The sections cover roughly the following subject areas:

1. General Computer Science (Prof. Esko Ukkonen, Assoc. Prof. Matti Mäkelä): algorithms and data structures, computational complexity, computational geometry, machine learning, Bayesian networks, neural networks, computer graphics, numerical and symbolic computation, computational biology, geoinformatics, computationally intensive tasks, computer-aided instruction, computers in education,

2. Computer Software (Prof. Martti Tienari, Assoc. Prof. Jukka Paakki, Assoc. Prof. Kimmo Raatikainen): programming languages, compilers, formal specification and verification, software engineering, distributed systems, computer networks, operating systems, performance evaluation, and

3. Information Systems (Prof. Heikki Mannila, Assoc. Prof. Seppo Sippu): databases, human-computer interfaces, computer supported co-operative work, information system design methodology, design of databases, text databases, object-oriented databases, logic databases, database structures and algorithms, document management, data mining and knowledge discovery, management of spatial data (GIS).

The University of Helsinki has many diverse teaching and research offerings related to computer applications. At the Department of Mathematics there is an active group in mathematical logic, numerical analysis and symbolic computation as well as some interest in theoretical computer science. The Rolf Nevanlinna Institute under the Faculty of Science is a research Institute of mathematics, computer science, and statistics with the main tasks of research and doctoral education. Our students also benefit from the hardware-oriented teaching (e.g. electronics, digital electronics, microcomputers, interface electronics) given at the Department of Physics. In the Faculty of Social Sciences some teaching and research is devoted to computational statistics, administrative information systems, and the social effects of data processing. In the Faculty of Arts there is a research unit in computational linguistics and a degree program in linguistic theory and cognitive science.

Chapter 2

Research

2.1 Review

The research at the department has evolved over the years similarly to the international research trends in computer science. Early work in numerical analysis in the 1960's made room for work in programming languages and compilers in the 1970's. Since then the research has diversified and its volume has increased. In the following, the research activities of each section of the department are reviewed.

The main research areas in the section of General Computer Science are algorithms and data structures, machine learning, probabilistic reasoning, computations by complex dynamic systems (cellular automata and genetic algorithms) and computational biology. Algorithms and data structures is the area with the longest tradition. The work on string matching algorithms (Ukkonen, Tarhio) has been particularly successful. Theoretical work has often been conducted within the framework of systems research providing practical motivation for the problems studied. Currently, special emphasis is given to the research on algorithmic problems in computational biology and bioinformatics.

Machine learning and probabilistic modeling are active research directions related to artificial intelligence. The complex systems computation group (CoSCo) has studied prediction and model selection issues in the probabilistic and information-theoretic frameworks, focusing on probabilistic model families, such as Bayesian networks and finite mixture models. The group has also performed extensive empirical and theoretical work on stochastic optimization methods, including simulated annealing and genetic algorithms. The machine learning group (Kivinen, Mannila, Ukkonen) has studied different learning models and the complexity of learning tasks within these models. One of the results is the first

MDL learning algorithm that has a proven performance guarantee. The more practice-oriented work has developed, for example, new Occam algorithms for learning decision trees and decision lists, and software tools for testing and comparing various machine learning algorithms.

Our computer software research can be subdivided into two main areas: distributed systems and telecommunication software (Raatikainen, Tienari) and programming languages and software engineering (Paakki). In some projects these two areas are intertwined.

In telecommunication software we have both industrially and academically oriented research projects. In the former category one could classify the European ACTS-project DOLMEN (Service Machine Development for an Open Long-Term Mobile and Fixed Network Environment, Raatikainen) as well as the project Mowgli (Tienari, Raatikainen, Alanko) concentrating on mobile computing in general and also in developing a software architecture for implementing mobile-aware applications. Also the project RODAIN (Real-Time Object Based Database Architecture for Intelligent Networks, Raatikainen) has strong ties to the Finnish telecommunication and software industries. A new project on agents in mobile computing has just started.

The projects in distributed systems are OCDE (Tienari), investigating open distributed computing environments, and MOCO (Modeling Concurrency, Tienari), concentrating on formal specification and verification of distributed systems. In the latter area, in particular, we have achieved notable results, well-received by the international research community.

The department has long traditions in the research of programming languages and compilers. This research still continues but is mainly directed now to object-oriented languages and their implementation problems. From this work a new research line in software engineering (Paakki) has evolved. There are two projects in this area, A Channel into Object-Oriented Protocol Design and A Framework Editor for Java (FRED).

In information systems the largest research project concentrates on data mining (Mannila, Toivonen, Verkamo), also known as knowledge discovery in databases. The work on data mining has several subprojects including industrial projects as well as an ESPRIT project. The research is done in cooperation with the machine learning group, with statisticians, and with the appliers of the work. Recent research results include efficient data mining methods for database re-engineering, methods for finding recurrent episodes within event sequences, and development of automatic tools for the simulation of complex statistical models.

Other research projects include document management (Kilpeläinen, Mannila), computer-supported cooperative work (Erkiö), and animation of algorithms (Tarhio). Research results include new efficiently evaluable query languages for text databases and methods for assembling structured documents.

Active projects and research areas as well as individual research of some researchers and graduate students are presented in more detail in Section 2.2. Section 2.3 lists a selection of recent publications. More information on research projects can be found on the WWW pages of the department.

2.2 Projects

a) General Computer Science

Algorithms on Strings

Research on strings algorithms started at the department quite early, in 1981. The initial impulse came from computer applications in molecular genetics where there is a lot of demand for efficient and sophisticated string-manipulation algorithms. The group is now one of the leaders in its special field in the world and has obtained several basic results that have been included in recent international text books.

The reputation of the group is based on the work done in the 1980's on developing new algorithms for the following problems: edit distance, approximate string matching, DNA sequence assembly, and shortest common superstring. In the 1990's, various sublinear approximate string-matching algorithms and new measures for string similarity have been developed. For the important problem of how a text should be preprocessed to speed-up subsequent approximate string searches, several solutions have been discovered based on the suffix-tree of the text or the so-called q-grams. For constructing suffix-trees, a natural on-line algorithm has been developed. The two-dimensional string matching problem has also been studied. We were able to obtain the first optimal expected time solutions using simple, practical algorithms.

The current research topics include feature-based algorithms for approximate string matching, two and higher dimensional string matching with applications in computational biology (modeling of viruses, protein folding), sequence databases with applications in bioinformatics, the indexing and clustering of strings and documents, and the search by content in image and music databases.

Current members of the group are Prof. Esko Ukkonen (group leader), Doc. Jorma Tarhio, M.Sc. Kimmo Fredriksson, M.Sc. Raul Hakli, M.Sc. Juha Kärkkäinen, Teemu Kivioja, M.Sc. Kai Korpimies, M.Sc. Kjell Lemström, Dr. Matti Nykänen, M.Sc. Janne Ravantti, Ph.Lic. Erkki Sutinen, and M.Sc. Hellis Tamm. The group gets support from the Academy of Finland.

Publications: [84, 88, 90, 92–95, 114–116, 136, 137, 140, 141].

Home Page: http://www.cs.helsinki.fi/research/pmdm/cpm/

Machine Learning

Being able to build computer systems that can learn in some sense is one of the very central problems in artificial intelligence. Inspired by certain theoretical advances in the field such as Valiant's PAC model and Rissanen's MDL principle, the project generally aims to apply to machine learning the approach of theoretical computer science and algorithmics. This international trend is called computational learning theory. Our theoretical work is supported by experimenting with the algorithms.

Results on decision tree learning include analyses of the requirements for efficient multisplitting on numerical attributes and comparison of the most commonly used attribute evaluation functions with respect to these requirement. A practical method that results in optimal splitting has also been developed. Testing environments for learning algorithms have been developed.

Some promising results have been obtained about MDL based learning algorithms in particular for the case where the instances are strings of arbitrary length. The method has been applied to a clustering problem of biological sequences.

Research has also been done on on-line learning algorithms that make no assumptions about the distribution of noise in the data. This is known as agnostic on-line learning. Emphasis has been on learning simple statistical models, such as generalized linear models, with algorithms that learn fast even if there is a large number of irrelevant variables present.

The current research themes of the group are:

- development of MDL-based and other learning algorithms with a particular emphasis on potential applications in computational biology,
- development of decision tree learning algorithms that have a sound theoretical basis and perform well in practice,

- generalizing the agnostic on-line learning methods further and deepening the understanding of their relationship with more classical statistical approaches,
- applications of symbolic learning approaches to problems of mobile robotics, and
- (lazy) learning for planning in continuous domains, with applications to an industrial case-based bioprocess planner (joint work with VTT and Finnish breweries).

The group has good international reputation and is together with nine other European groups a member of the ESPRIT Working Group NeuroCOLT II and a site of European Machine Learning Network. The group also works in close co-operation with Prof. Heikki Mannila's data mining group.

The members of the group are Prof. Esko Ukkonen (group leader), Doc. Jyrki Kivinen, Dr. Tapio Elomaa, M.Sc. Tibor Hegedüs, M.Sc. Markus Huttunen, M.Sc. Juho Rousu (VTT Biotechnology and Food Research) and M.Sc. Jaak Vilo. The group gets funding from the Academy of Finland and from the European ESPRIT Programme.

Publications: [189, 195–197, 203–206, 208–215, 218, 220–222, 259–262].

Home Page: http://www.cs.helsinki.fi/research/pmdm/ml/

Probabilistic Modeling and Complex Systems Computation (CoSCo)

A complex system is a collection of simple interacting agents, elements or processes, whose collective behavior exhibits interesting large scale phenomena. Such systems can be found in various disciplines, including computer science, economics, mathematical biology and physics. The Complex Systems Computation (CoSCo) research group studies computational issues related to complex systems focusing on prediction and model selection issues. Current work of the CoSCo group is concentrated on theory and applications of *Bayesian (belief) networks*, and related probabilistic model families, such as *finite mixture models*. Other research areas addressed include *case-based reasoning*, *artificial neural networks*, and stochastic optimization methods, such as *simulated annealing* and *genetic algorithms*. The results achieved include extensive theoretical and empirical studies concerning

- efficient methods for learning probabilistic models from sample data;

- a novel, computationally efficient Bayesian criterion for selecting the most relevant model features (attributes);
- the accuracy of different marginal likelihood approximation techniques for Bayesian networks with hidden variables;
- similarities and differences between Bayesian and information-theoretic (MDL, MML) modeling approaches;
- a general Bayesian framework for case-based reasoning;
- techniques for mapping Bayesian networks to neural network architectures, thus allowing massively parallel implementations of Bayesian reasoning;
- computationally efficient stochastic optimization methods, including a novel variant of simulated annealing with an adaptive cooling schedule.

The theoretical results obtained have been empirically tested by using several real-world data sets. Some of the JAVA software used in these experiments can be downloaded from the group home page (`http://www.cs.helsinki.fi/research/cosco/`). Examples include

- BAYDA: Bayesian Predictive Discriminant Analysis with feature selection.
- D-SIDE: a Bayesian decision support system with JAVA interface.

In the empirical tests with various public domain classification datasets, D-SIDE consistently outperforms the results obtained by alternative approaches, such as decision trees and neural networks.

Basic research work by the group has been supported by grants from the Academy of Finland, University of Helsinki, and various foundations. More applied work has been performed with support from TEKES and the domestic and foreign industrial partners which include, e.g., Kone, Nokia, ABB, Enso and the AT&T corporations. Some of the resulting software has been adopted in the industry.

Current members of the CoSCo group are Doc. Henry Tirri, Dr. Petri Myllymäki, M.Sc. Tomi Silander, Petri Kontkanen, Jussi Lahtinen and Kimmo Valtonen. The group has also hosted several foreign academic visitors and graduate students, and participates in two EC-funded research networks (the NeuroCOLT working group on neural and computational learning theory, and the Network of Excellence in Neural Computing, NEURONET).

Publications: [118, 120–128, 192, 193, 217, 223–233, 236, 237].

Home Page: http://www.cs.helsinki.fi/research/cosco/

Animation Aided Problem Solving (AAPS)

Animation is a standard technique in computer-aided instruction. The project aims at applying the methods of algorithm animation in problem solving. The idea is based on the internal similarity of algorithm design and problem solving.

Traditionally animations were coded by hand demanding much time. We have developed two systems for fast generation of animations for algorithms. Eliot works on Linux workstations and Jeliot runs on the World-Wide Web. The speed of generation is one of the key factors of the applicability of the approach.

Besides generation of algorithm animations we have studied how to prepare animations of simple algorithms with the Microsoft Excel spreadsheet program. We are also developing ways to use Excel animations in problem solving. This approach is important because of the wide availability of Excel. In addition to animation we are making tools for creating and and managing concept maps in education and problem solving.

Our topics lie in the borderland of computer science and education, and the research is carried out as a joint project with the Department of Teacher Education. Current members of the AAPS group are Doc. Jorma Tarhio (group leader), Assoc. Prof. Veijo Meisalo (Department of Teacher Education), Ph.Lic. Erkki Sutinen, Erkki Rautama, and Tommi Teräsvirta. The group has got funding from the Ministry of Education and University of Helsinki in 1996-98.

Publications: [69, 273, 281, 284–286, 289, 290].

Home Page: http://www.cs.helsinki.fi/research/aaps/

Individual Research

Visualization in Science and Education (Assoc. Prof. Matti Mäkelä): Advanced computer graphics provides means of visualizing scientific data and complicated formal structures and interactions which have earlier been difficult to manage and conceive by human brains. Scientific visualization offers a method to make pictures from the abstract material produced by scientific models and measurements. Thus, the natural human ability to see and think visually is utilized besides the traditional scientific thinking. The same applies in education, too. One essential problem is to select a proper type of pictures for representing the abstract data and to formalize the transformation from data to pictures. Another problem is the visual literacy, that is the ability to express thoughts by pictures, and the ability to interpret pictures. The scope of this research is to find out practical recommendations to attack these problems. The present activities will include animation tools, and visualization and hypermedia applications for education. As the problem area is highly interdisciplinary there are several contacts within the university, and with the Helsinki University of Technology, the Center for Scientific Computing, and the University of Industrial Arts (Helsinki), among others.

Publications: [269–271, 274–276].

b) Computer Software

Modelling of Concurrency (MOCO)

Distributed systems and parallel programs are notoriously prone to errors caused by subtle differences in the relative execution orderings of the components of the system. In order to avoid such errors, systems should be carefully specified and analyzed. The international research in this area, often called simply concurrency, has been active in the 1980's and has resulted in numerous specification methods and models for this purpose, e.g. state transition systems, Petri nets, process algebras and temporal logics. Elegant and rich theories supporting these models have made the models even more useful. Results in these theories can be employed in computer based tools for checking and verifying concurrent real life software.

An important class of distributed algorithms consists of computer communication protocols. We have worked with the problems of protocol analysis since 1984. We begun by constructing a reachability analysis tool PROTAN88 designed for protocols specified with an extended state transition model (ESTELLE specification language). In the 1980's this tool has been tried out by us in analyzing several well-known real-life protocols like X.25, FTAM and X.411/P1. Some of our improvement suggestions were included in the FTAM state tables published by CCITT and ISO.

Lately, our work has been more conceptual and theoretical. Various equivalence concepts have been studied in process algebra contexts (Milners CCS and ISO LOTOS) and the equivalence preserving minimization of labeled state transition systems has been studied. We have been especially interested in divergence preserving behavioural equivalences and preorders.

The analysis of liveness properties is especially important in the context of specification and verification of temporal logic properties of concurrent systems. Temporal logics are a class of logic-based formal languages containing temporal operators, such as 'sometime' or 'always'. Their applications to verification of concurrent systems have been internationally one of the main research directions in the concurrency research for the last twenty years. In our work on temporal logic, we have examined the preservation of temporal logic properties in the construction of reduced models based on divergence-sensitive equivalences and preorders. We have also examined the relations of temporal logics, particularly the so-called mu-calculi or fixed-point temporal logics, with automata on infinite objects.

A recurrent theme in our research is to make analysis and verification feasible for concurrent systems of realistic size. That is why a goal in our research is to understand and support compositional (or modular) specification and verification of concurrent systems. We have been e.g. analyzing and applying the weakest equivalence (CFFD-equivalence) preserving deadlocks and linear next-timeless linear temporal formulae. Our research employs case analyses (often computer communication protocols) to ensure the practical usefulness of the theoretical concepts.

At the university of Helsinki we have developed a computer-based verification tool BIDMIN allowing us to minimize labeled transition systems with respect to divergence preserving bisimilarity and branching bisimilarity as well as the related congruences. Vesa Hellgren has been developing parallel verification algorithms. The algorithms have been implemented in PAVEL (Parallel Verification of LOTOS/LTS) which is a software running in the Cray T3E supercomputer and a network of Linux workstations. We also often use a LOTOS oriented verification tool ARA TOOLS developed at the htmladdnormallinkTechnical Research Centre of Finlandhttp://www.tekes.fi/english/.

Current members of the MOCO group are Prof. Martti Tienari (group leader), Doc. Roope Kaivola, M.Sc. Vesa Hellgren, Ph.Lic. Timo Karvi, M.Sc. Päivi Kuuppelomäki and M.Sc. Matti Luukkainen (Nokia Research Centre). As an associate member of the group we have Prof. Antti Valmari (Tampere University of Technology, Software Systems Laboratory) who is the chief designer of ARA TOOLS. Our concurrency group has been participating (1994–97) in the European COST247-project Verification and Validation Methods for Formal Descriptions.

Publications: [18, 98, 100, 103, 106–111].

Home Page: http://www.cs.helsinki.fi/research/moco/

Open Distributed Computing Environments (ODCE)

The rapid advances in data communication networks have greatly increased the possibilities to make use of the various computational and information services available within reach of a computer network. These services may have different origins, they are implemented on different types of computational platforms, they can be located on sites far apart, and they can be administered independently. As different services become reachable, interoperability of services must be considered. Distributed computation needs an enhanced functionality of the infrastructure, including the emerging technologies.

The research of the ODCE group contributes to open architecture models, such as the Open Distributed Processing reference model (RM-ODP) standardised by ISO and ITU, and open platforms, such as OMG CORBA. The research is also in close relationship with telecommunication related architectures, like TINA.

The department started research activity in the area of open distributed processing at the end of 1980's. In our first project (AHTO) we designed a distributed software environment (middleware) offering for a user a homogeneous interface to the computing services available in a heterogeneous computer network. Since that time the department has participated in the development of the Reference Model for Open Distributed Processing, under the auspices of ISO and ITU-T (former CCITT).

The DRYAD project, 1992-1996, concentrated on heterogeneous environments where autonomously administered systems federate with each other. Federation problems arise from the asynchrony of member system evolution, and from the independent technology choices at member systems (hardware, operating system, middleware, languages, applications). The federation solutions are built on meta-information services, like trading services (global and dynamic repository of service providers), and type repository services (dynamic repository of service types, federation contract schemata, and mappings to local technology solutions). The DRYAD project developed a prototype software package for some of the middleware services - especially for the trading function.

The CORBA-FORTE project (CORBA-Based Framework for Telecommunications) is starting in 1998. The project will concentrate on CORBA technology. It will seek general patterns for improving CORBA system performance. Prospects for improved quality of service and performance are found on two areas. First, additional services can be implemented into the platform itself. Second, applications can exploit the platform services more efficiently when tailored object design patterns are used.

The members of the ODCE group are Prof. Martti Tienari, Dr. Timo Alanko, Ph.Lic. Lea Kutvonen, Chief Systems Analyst Petri Kutvonen, M.Sc. Pekka Kähkipuro and M.Sc. Liisa Marttinen.

Publications: [25–37].

Home Page: http://www.cs.helsinki.fi/research/odce/

Mobile Computing (MOWGLI)

Recent developments in wireless communication and personal computer technology have laid a new foundation for mobile computing. Modern portable computers and wireless connections have created a new platform for distributed information processing where a mobile user have access to various computing services and information stores any time and independently of his or her present location.

The goal of the MOWGLI project is to study, design and test a data communication architecture for a pan-European GSM-based mobile data service, and alike. The environment of an application consists of mobile PC's which can be connected over a wireless WAN to the Internet. The work in the project is concentrated on the architectural aspects supporting mobility-aware computing. The main issues have been data transport service for constrained wireless links, control of computing in weakly connected and disconnected states, support for legacy software in a mobile environment and design of mobility-aware application software for nomadic users (file transfer, e-mail, WWW), and communication performance in this environment. The experimental research is based on a prototype implementation of the architecture.

The members of the MOWGLI project group are Prof. Martti Tienari, Dr. Timo Alanko, Assoc. Prof. Kimmo Raatikainen, M.Sc. Heikki Helin, M.Sc. Markku Kojo, and M.Sc. Heimo Laamanen. The industrial partners of the project are Nokia Mobile Phones, Nokia Research Centre, Nokia Telecommunications Systems, SSH Communications Security, and Telecom Finland.

Publications: [6–9, 11–14, 19, 38, 45].

Home Page: http://www.cs.helsinki.fi/research/mowgli/

Service Machine Development for an Open Long-term Mobile and Fixed Network Environment (DOLMEN)

The EC/ACTS project DOLMEN (ACTS Ref. AC036), 1995–98, demonstrates, assesses and promotes a Service Architecture (called Open Service Architecture for a Mixed fixed and mobile environment – OSAM) that meets the requirements of open provision of communication services over both fixed and mobile heterogeneous and multi-provider telecommunications networks.

The project's approach is based upon the RACE Open Service Architecture (OSA) and also progresses results from RACE projects on mobility. The TINA-C current and future developments have been taken into account. In developing OSAM, the parallel work of related ACTS projects are constantly monitored and utilized as appropriate. As in the case of OSA and TINA-C, the DOLMEN approach subscribes to a vision, beyond that currently offered by IN and TMN, of a telecommunications service infrastructure as a large scale, distributed processing environment.

Participants in DOLMEN are Fondazione Ugo Bordoni (Italy), University of Catania (Italy), FINTRACOM S.A. (Greece), National Technical University of Athens (Greece), Personal Communications Services (UK), Research and Consultancy (UK), Sema Group (France), University of Helsinki (Finland), VTT Information Technology (Finland), Koniglijke PTT Nederland N.V. Research (The Netherlands), AT&T Network Systems Nederland B. V. (The Netherlands), and Telecom Finland (Finland),

The current members of the DOLMEN research group are Assoc. Prof. Kimmo Raatikainen (group leader), M.Sc. Stefano Campadello, M.Sc. Oskari Koskimies, M.Sc. Mika Liljeberg and Jarkko Sevanto.

Publications: [15, 16]

Home Page: http://www.cs.helsinki.fi/~kraatika/DOLMEN/summary.html

Real-Time Object-Based Database Architecture for Intelligent Networks (RODAIN)

Database technologies will, already in the near future, have a dominant role in telecommunication and data communication networks. The information and knowledge needed in network operations and management will be organized as a logical entity. Due to the world-wide nature of communication the only way to achieve the logical uniformity is the interoperability of autonomous databases.

The research project Darfin, Database ARchitecture For Intelligent Networks, 1993-95, examined database architectures that can fulfill the requirements of Intelligent Networks (IN) and Telecommunication Management Networks (TMN). The current ITU-T Recommendations for TMN are based on object-oriented modelling. In addition, the long-term architecture of IN heavily advocates OO-modelling. Therefore, the project focused on object-oriented real-time databases.

The research project RODAIN continues the work done in the Darfin project. In the project we have designed and specified a real-time object-oriented database architecture for Intelligent Networks and implemented a simple prototype. The current objective of the RODAIN project is to extend the prototype to a distributed real-time object-oriented database that can be used in telecommunication.

Research topics:

- real-time transaction processing,
- main memory databases,
- distribution,
- recovery, and
- fault-tolerance.

The industrial partners and financial supporters of the project are Nokia Telecommunications, Solid Information Technology, and Telecom Finland. The RODAIN project is also funded by the Finnish Technology Development Centre (TEKES).

Current members of the research group are Assoc. Prof. Kimmo Raatikainen (group leader), Jan Lindström, M.Sc. Tiina Niklander, B.Sc. Pasi Porkka and Ph.Lic. Juha Taina.

Publications: [5, 39–44, 46–54, 65, 131, 142, 143, 145–147, 148, 156].

Home Page: http://www.cs.helsinki.fi/research/rodain/

A Channel into Object-Oriented Protocol Design (Kannel)

The constantly growing complexity of distributed and telecommunications applications has made it necessary to develop high-level application-oriented languages and tools. There exists a number of standardized languages and their implementations in the area, but usually they focus just on some rather narrow aspect of the problem. This makes it necessary to use several unrelated languages and environments when developing a nontrivial distributed application.

This research project develops Kannel, an integrated language for the design and implementation of communication protocols. The main features of Kannel are high-level application support, object-orientation, sound protocol refinement, and visual notations. In contrast to conventional languages in the area, Kannel and its programming environment provide facilities for all the main tasks in a typical protocol development effort.

The first version of Kannel has been implemented. The environment includes a visual Kannel editor, a translator into C++, and a graphical animator/debugger.

The members of the research group are Assoc. Prof. Jukka Paakki (group leader), M.Sc. Kari Granö, and M.Sc. Antti-Pekka Tuovinen.

Publications: [10, 17, 24, 99].

Home Page: http://www.cs.helsinki.fi/research/kannel/

A Framework Editor for Java (FRED)

Reuse is one of the most promising ways to increase the productivity and quality of software development. Unfortunately, reusability has turned out to be rather hard to be reached both in theory and in practice.

Recently, special object-oriented methods have been developed to solve the reusability problem. In this research project we study two of the most promising ones, design patterns and application frameworks. The former captures general and mature design decisions into a reusable form, whereas the latter category provides large-scale code reuse in the form of an extensible implementation of the core functionality of a family of related applications.

The project develops a prototype CASE tool for the development of reusable application frameworks founded on design patterns. The tool makes it possible to instantiate integrated design patterns as code into the framework and to systematically develop a final application from it. The implementation language of the tool, the frameworks, and the applications is Java.

The project is carried out as co-operation between our department (UH) and the Department of Computer Science at the University of Tampere (UT). The project members are Prof. Kai Koskimies (group leader, UT), Assoc. Prof. Jukka Paakki (UH), M.Sc. Markku Hakala (UT), M.Sc. Juha Hautamäki (UT), M.Sc. Jyrki Tuomi (UT), M.Sc. Antti-Pekka Tuovinen (UH), M.Sc. Antti Viljamaa (UH), and M.Sc. Jukka Viljamaa (UH). The main sponsor of the project is TEKES, and the industrial partners are TT Tieto, Stonesoft, Nokia Telecommunications, Novo Group, Sun Microsystems, Major Blue Company, ICL Data, Dycom, Nokia Research Centre, Telecom Finland, Valmet Automation, and Profit.

Publications: [55–57, 62, 63, 67, 68, 70, 73–75].

Home Page: http://www.cs.helsinki.fi/research/fred/

c) Information Systems

Data Mining

Data mining (or knowledge discovery in databases) is a new research area developing methods and systems for extracting interesting and useful information from large sets of data. Data mining methods can be used in a variety of application areas, such as commercial databases, telecommunications, epidemiological data, etc. The area combines techniques from databases, statistics, and machine learning.

The Data Mining research group has developed data mining methods and studied the theory of data mining. The research started in late 1980's in the context of developing tools for inferring integrity constraints from databases.

We have developed methods for finding recurrent episodes in event sequences, and used these to locate strong rules about the occurrences of events. Clustering methods have also been applied to locate regularities in sequential data. For numerical time-series data we have developed methods that are able to discover similarities in various aspects of potentially related time-series.

Data mining can produce large amounts of new information. We are working on the data mining process as a whole and on the selection of the interesting regularities in particular. In connection with the Document Management group, we have considered these issues in the analysis of text and structure in marked documents.

The group has studied the theory of data mining, e.g., by looking at the relationship of the logical complexity of the discovered sentences and the sample size needed for discovery, and by investigating various frameworks for data mining.

A growing research topic has been the use of Markov chain Monte Carlo methods in data analysis, in particular in the analysis of event data. We develop tools for the automatic analysis of complex statistical models (Bayesian or full probability models), and we model and analyze data with other scientists, e.g., in epidemiology, paleoecology, and archaeology.

The research is done in several projects funded by the Academy of Finland, TEKES, and the European ESPRIT Programme. The group has close cooperation with the Document Management and Machine Learning groups.

The members of the Data Mining group are Prof. Heikki Mannila (group leader), Dr. Helena Ahonen, M.Sc. Oskari Heinonen, Ykä Huhtala, M.Sc. Mika Klemettinen, Karri-Pekka Laakso, Tommi Mononen, M.Sc.

Vesa Ollikainen, M.Sc. Pirjo Ronkainen, M.Sc., M.Th. Marko Salmenkivi, Jouni Seppänen, Dr. Hannu Toivonen, and Doc. Inkeri Verkamo.

Publications: [20–23, 91, 144, 150, 151, 161–172, 175, 176, 180–186, 207, 219, 234, 235, 238, 240–242, 244–246, 250, 253, 257, 258, 263, 264].

Home Page: http://www.cs.helsinki.fi/research/fdk/datamining/

Document Management (DocMan)

Text with a structure is quite common: dictionaries, reference manuals, and annual reports are typical examples. In recent years, research on systems for processing structured documents has flourished. The SGML and ODA standards have further increased the interest in the area. The Document Management (DocMan) Research Group studies the theory and application of such structured documents.

Structured and Intelligent Documents (SID) is an on-going project within the DocMan group that studies and develops methods for attaching intelligent features to structured documents. The purpose of these features is to make the manipulation (storage, retrieval, and assembly) of documents easier. The project started in 1995. SID is part of the Electronic Printing and Publishing program started by the Technology Development Centre of Finland (TEKES). Funding for SID is provided by TEKES and a group of supporting companies.

One of the basic problems in document management is to provide on-demand generation of individualized documents through dynamic document assembly. Document assembly composes new documents from an existing collection of documents. Naturally, document markup and structure contribute to the retrieval and reuse of document fragments.

An intelligent document contains knowledge about itself and its environment. It supports assembly of documents based on inputs given by the user. It is no longer a passive, linear representation of text, but is able to construct itself dynamically. Document assembly is intelligent when it uses application-domain-specific information about the document in addition to the contents and their structure.

The goals of the SID project include (1) defining the information and the knowledge a structured document must contain so that it can work in an active and intelligent way, (2) developing prototype tools for intelligent assembly, and (3) defining a methodology for incorporating intelligence into document material. As a basis for the project we consider structured documents marked up with SGML. The project combines methods and tools from, e.g., structured-document management, information retrieval, pattern matching, data mining, distributed systems, and machine learning. When dealing with documents in morphologically rich languages like Finnish, also natural language processing is vital to the success of document assembly.

Other ongoing research within the DocMan group includes creating automatically meaningful fragments of long documents, and classifying roles of structured document elements. Former research projects include the RATI project (1988–91) for building a prototype document manipulation system which provides multiple views of a document and the sgrep project (1995) which designed and implemented a search tool for structured documents. Also some results of the VITAL project (1990–95) are usable in this context: one of the tools built in the VITAL project was a general purpose text transformation generator suitable also for structured document transformations.

Researchers of the group are Prof. Heikki Mannila, Doc. Pekka Kilpeläinen, Dr. Helena Ahonen, M.Sc. Barbara Heikkinen, M.Sc. Oskari Heinonen, Jani Jaakkola, Dr. Greger Lindén, Jyrki Niemi and Kimmo Paasiala.

Publications: [71, 117, 133, 139, 167–169, 198–202, 247–256].

Home Page: http://www.cs.helsinki.fi/research/rati/

Transaction Management Support for Cooperative Applications (TRANSCOOP)

Today's software tools mostly aim at supporting a single user only. For example, consider text editors for writing documents, CAD-tools or software development environments. There are excellent text editors available, but they do not appropriately support multiple authors to work on the same document concurrently.

A major conceptual problem in this framework is to ensure consistency criteria for the data concurrently processed by multiple users. Conventional database technology already provides mechanisms to absolutely guarantee consistency constraints by controlling the concurrent access of different users to shared data. Unfortunately, existing transaction management concepts are not suitable for supporting and controlling cooperation between users, because they are designed to fully isolate users from each other.

The design of cooperative systems includes the description and formal specification of cooperative activities. Such specifications have to be mapped to a cooperative transaction management model in order to enable appropriate database management support at run time. Cooperative transactions are thus intrinsically complex operations that are difficult to understand completely. A theory is needed for distinct notions of correctness and for correctness-preserving transformations that allow the designer to map specifications onto implementation platforms in a guided way.

The project finished in 1997. The partners of the project were GMD (D), University of Twente (NL) and Technical Research Centre of Finland (VTT) having the University of Helsinki as a subcontractor.

Members of the research group are Doc. Jari Veijalainen, Doc. Hannu Erkiö, Doc. Henry Tirri, and Ph.Lic. Juha Puustjärvi.

Publications: [149, 152–155, 158].

2.3 Publications

Recent publications of the department are listed below according to the ACM Computing Reviews (CR) classification system. The list contains selected publications in 1994–95 and new publications since January 1996.

General Literature – Introductory and Survey (A.1)

1. K. Korpimies: *The Basics of Computer Science: A Student's Guide* (in Finnish). Lahti Research and Training Centre, University of Helsinki, 1995 and 1996.

2. K. Korpimies and J. Vanha-Eskola: *The Basics of Computer Science: A Tutor's Guide* (in Finnish). Lahti Research and Training Centre, University of Helsinki, 1995 and 1996.

3. J. Tarhio and M. Tienari (eds.): Computer science at the University of Helsinki 1996. Report A-1996-3, Department of Computer Science, University of Helsinki, 1997.

Computer Systems Organization – Computer-Communication Networks – General (C.2.0)

4. K. Korpimies: *Internet: A User's Guide* (in Finnish). Suomen Atkkustannus Oy, 1994 and 1995.

5. J. Taina: *Database Architecture for Intelligent Networks*. Ph.Lic. Thesis, Report C-1997-50, Department of Computer Science, University of Helsinki, 1997.

Network Architecture and Design (C.2.1)

6. T. Alanko, L. Hippeläinen, M. Kojo, H. Laamanen, M. Liljeberg and K. Raatikainen: Mowgli: Enhanced communication services for mobile computing using GSM. In *Proc. ITU Telecom Interactive 97 Forum*, 1997.

7. T. Alanko, H. H. Kari, M. Kojo, H. Laamanen, M. Liljeberg, K. Raatikainen and M. Tienari: Communication services for mobile office in wireless WAN environments. In *Global Communications: Interactive '97* (ed. R. Struzak), Hanson Cooke Ltd, 1997, 219–225.

8. T. Alanko, M. Kojo, H. Laamanen, K. Raatikainen and M. Tienari: Mobile computing based on GSM: The Mowgli approach. In *Proc. IFIP World Conference on Mobile Communications: Technology, Tools, Applications, Authentication and Security* (eds. J. L. Encarnaçao and J. M. Rabaey), Chapman & Hall, 1996, 151–158.

9. T. Alanko, M. Kojo, M. Liljeberg and K. Raatikainen: Mowgli: Improvements for Internet applications using slow wireless links. In *Proc. 8th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, 1997, 1038–1042.

10. K. Granö, J. Paakki, A. Viljamaa and J. Viljamaa: Protocol engineering & Kannel. *Dr. Dobb's Journal* **23**, 4 (1998), 46-53.

11. J. Kiiskinen, M. Kojo, M. Liljeberg and K. Raatikainen: Data channel service for wireless telephone links. In *Proc. 2nd International Mobile Computing Conference*, ACM, 1996, 60–69.

12. M. Kojo, K. Raatikainen and T. Alanko: Connecting mobile workstations to the Internet over a digital cellular telephone network. In *Mobile Computing* (ed. T. Imielinski and H. Korth), Kluwer Academic Publishers, 1996, 253–270.

13. H. Laamanen, T. Alanko, M. Kojo and K. Raatikainen: The Mowgli project for pan-European GSM data. In *Mobile Internet*, IBC Technical Services, 1996.

14. M. Liljeberg, H. Helin, M. Kojo and K. Raatikainen: Mowgli WWW software: Improved usability of WWW in mobile WAN environments. In *Proc. IEEE Global Internet 1996 Conference*, 1996, 33–37.

15. M. Liljeberg, S. Furnell, K. Raatikainen and P. Reynolds: Internet information browsing using GSM data communications: A benchmarking study. In *ACTS Mobile Communications Summit '97: "Towards a wireless communication society"*, vol. 2, Center for PersonKommunikation, 1997, 876–881.

16. K. Raatikainen and S. Trigila: Software radio and service architectures. In *ACTS Mobile Communications Summit '97: Invited and round table papers*, Center for PersonKommunikation, 1997.

Network Protocols (C.2.2)

17. A. Arvonen, K. Granö, J. Harju and J. Paakki: Experiences with the integration of protocol software tools. *Computer Communications*, **19**, 2 (1996), 141–151.

18. R. Kaivola: Using compositional preorders in the verification of sliding window protocol. In *Proc. 9th International Conference on Computer-Aided Verification, CAV '97* (ed. O. Grumberg), Lecture Notes in Computer Science 1254, Springer, 1997, 48–59.

19. M. Kojo, K. Raatikainen, M. Liljeberg, J. Kiiskinen and T. Alanko: An efficient transport service for slow wireless telephone links. *IEEE Journal on Selected Areas in Communications* **15**, 7 (1997), 1337–1348.

Network Operations (C.2.3)

20. K. Hätönen, M. Klemettinen, H. Mannila, P. Ronkainen and H. Toivonen: TASA: Telecommunications alarm sequence analyzer, or "How to enjoy faults in your network". In *Proc. IEEE Network Operations and Management Symposium, NOMS '96*, IEEE, 1996, 520–529.

21. M. Klemettinen, H. Mannila and H. Toivonen: A data mining methodology and its application to semi-automatic knowledge acquisition. In *Proc. International Workshop on Database and Expert Systems Applications, DEXA '97* (eds. A. Hameurlain and A. M. Tjoa), IEEE Computer Society Press, 1997, 670–677.

22. M. Klemettinen, H. Mannila and H. Toivonen: Rule discovery in telecommunication alarm data. To appear in *Journal of Network and Systems Management*, 1998.

23. M. Klemettinen, H. Mannila and H. Toivonen: Interactive exploration of interesting findings in TASA. To appear in *Information and Software Technology*, 1998.

Distributed Systems (C.2.4)

24. K. Granö and J. Paakki: From peer to service, from service to peer — object-oriented protocol refinement in Kannel. Report TUM-I9623, Institut für Informatik, Technische Universität München, 1996.

25. P. Kähkipuro: Basic patterns of distributed object-oriented computing. In *Proc. 2nd International Conference "The White Object-Oriented Nights", WOON' 96* (eds. A. Smolyaninov and A. Shestialtynov), St. Petersburg Electrotechnical University, 1996, 115–125.

26. P. Kähkipuro: Distributed software localization. In *Industrial and Short Paper Proc. International Workshop on Trends in Distributed Systems, TreDS '96* (eds. O. Spaniol, C. Linnhoff-Popien and B. Meyer), Band 17, Aachener Beiträge zur Informatik, Verlag der Augustinus Buchhandlung, 1996, 104-111.

27. P. Kähkipuro: An object model framework for middleware-based distributed systems. In *Proc. IFIP TC6 WG6.1 International Working Conference on Distributed Applications and Interoperable Systems, DAIS '97* (eds. H. König, K. Geihs and T. Preuss), Chapman & Hall, 1997, 201–207.

28. P. Kähkipuro, L. Kutvonen and L. Marttinen: Federated naming in an ODP environment. In *Proc. IFIP/IEEE International Conference on Open Distributed Processing and Distributed Platforms* (eds. J. Rolia, J. Slonim and J. Botsford), Chapman & Hall, 1997, 314–325.

29. P. Kähkipuro, L. Marttinen and L. Kutvonen: Reaching interoperability through ODP type framework. Report C-1996-96, Department of Computer Science, University of Helsinki, 1996.

30. P. Kähkipuro, L. Marttinen and L. Kutvonen: Reaching interoperability through ODP type framework. In *Proc. Conference on Convergence of Telecommunications and Distributed Computing Technologies, TINA '96*, VDE-Verlag, 1996, 283–284.

31. L. Kutvonen: Overview of the DRYAD trading system implementation. In *Proc. IFIP/IEEE International Conference on Distributed Platforms: Client/Server and Beyond: DCE, CORBA, ODP and Advanced*

Distributed Applications (eds. A. Schill, C. Mittasch, O. Spaniol and C. Popien), Chapman & Hall, 1996, 314–326.

32. L. Kutvonen: *The Role of Trading Function in Open Distributed Processing Infrastructure*. Ph.Lic. Thesis, Report C-1996-84, Department of Computer Science, University of Helsinki, 1996.

33. L. Kutvonen: Management of application federations. In *IFIP TC6 WG6.1 International Working Conference on Distributed Applications and Interoperable Systems* (eds. H. König et al.), Chapman & Hall, 1997, 33–46.

34. L. Kutvonen: Why CORBA systems cannot federate? In *Proc. OMA/ODP Workshop*, Object Management Group, 1997.

35. L. Kutvonen: Architectures for distributed systems: Open Distributed Processing reference model. In *HeCSE Workshop on Emerging Technologies in Distributed Systems*, Helsinki University of Technology, 1998.

36. L. Kutvonen: Supporting global electronic commerce with ODP tools. In *Trends in Distributed Systems '98: Electronic Commerce*, Morgan Kaufmann, 1998.

37. P. Kutvonen and L. Kutvonen: Measured performance of a special-purpose database engine for distributed system software. *Computer Communications Journal* **19**, 1 (1996), 21–29.

38. M. Liljeberg, H. Helin, M. Kojo and K. Raatikainen: Enhanced services for World-Wide Web in mobile WAN environment. In *Proc. 3rd International Conference on Communicating by Image and Multimedia, IMAGE'COM 96*, 1996.

39. P. Porkka and K. Raatikainen: CORBA access to telecommunications databases. In *Proc. IFIP TC6 WG6.7 International Conference on Intelligent Networks and Intelligence in Networks* (ed. D. Gaiti, Chapman & Hall, 1997, 281–300.

Special-Purpose and Application-Based Systems (C.3)

40. T. Niklander, J. Kiviniemi and K. Raatikainen: A real-time database for future telecommunication services. In *Proc. IFIP TC6 WG6.7 International Conference on Intelligent Networks and Intelligence in Networks* (ed. D. Gaiti), Chapman & Hall, 1997, 413–430.

41. T. Niklander and K. Raatikainen: Real-time platform for the RODAIN database. Report C-1996-53, Department of Computer Science, University of Helsinki, 1996.

42. T. Niklander and K. Raatikainen: Interworking requirements in the RODAIN database management system. Report C-1996-77, Department of Computer Science, University of Helsinki, 1996.

43. K. Raatikainen, J. Kiviniemi, T. Niklander and P. Porkka: Specification of RODAIN database architecture and interfaces. Report C-1996-103, Department of Computer Science, University of Helsinki, 1996.

44. K. Raatikainen, J. Kiviniemi, T. Niklander and P. Porkka: Implementation plan for the RODAIN prototype. Report C-1996-104, Department of Computer Science, University of Helsinki, 1996.

45. K. Raatikainen and A. Tuoriniemi: Mobility in the intelligent network architecture. In *Proc. 4e Colloque International sur l'Intelligence dans les Reseaux*, France Telecom, 1996, 289–294

46. J. Taina, P. Elovaara, T. Niklander and K. Raatikainen: Platform requirements of a real-time object-oriented database management system for telecommunications. Report C-1996-30, Department of Computer Science, University of Helsinki, 1996.

47. J. Taina and K. Raatikainen: Design issues and experimental database architecture for telecommunications. In *Proc. IFIP TC6 Conference on Intelligent Networks and New Services, IN '95* (eds. J. Nørgaard and W. Iversen), Chapman & Hall, 1996, 121–139.

48. J. Taina and K. Raatikainen: An experimental real-time object-oriented database architecture for intelligent networks. *Engineering Intelligent Systems* **3** (1996), 141–154.

49. J. Taina and K. Raatikainen: RODAIN: A real-time object-oriented database system for telecommunications. In *Proc. Workshop on Databases: Active and Real-Time, DART '96*, ACM, 1996, 12–15.

50. J. Taina and K. Raatikainen: Database usage and requirements in intelligent networks. In *Proc. IFIP TC6 WG6.7 International Conference on Intelligent Networks and Intelligence in Networks* (ed. D. Gaiti), Chapman & Hall, 1997, 261–280.

51. J. Taina and S. Son: A framework for real-time object-oriented database models. In *Proc. IEEE 3rd International Workshop on Object-Oriented Real-Time Dependable Systems, WORDS '97*, IEEE, 1997, 146–152.

52. J. Taina and S. Son: TARTOS — Toolbox for active real-time object-oriented database system models. In *Proc. IEEE Workshop on Parallel and Distributed Systems, WPDRTS '97*, IEEE, 1997, 131–140.

53. J. Taina and S. Son: Requirements for real-time object-oriented database models — How much is too much? In *Proc. 9th Euromicro Workshop on Real-Time Systems*, IEEE, 1997, 258–265.

54. J. Taina and S. Son: A toolbox approach for requirements analysis and modeling of active real-time object-oriented database models. Submitted to *Journal of Systems Architecture, Special Issue on Real-Time systems*, 1997.

Software – Programming Techniques – Object-Oriented Programming (D.1.5)

55. J. Hautamäki: A survey of frameworks. Report A-1997-3, Department of Computer Science, University of Tampere, 1997.

56. K. Hautamäki, J. Tuomi, A. Viljamaa and J. Viljamaa: Four example frameworks in Java. Report B-1997-6, Department of Computer Science, University of Tampere, 1997.

57. A. Viljamaa: Application frameworks in the Java environment. Report C-1997-24, Department of Computer Science, University of Helsinki, 1997.

Logic Programming (D.1.6)

58. T. Eiter, G. Gottlob and H. Mannila: Disjunctive Datalog. *ACM Transactions on Database Systems* **22**, 3 (1997), 364–418.

59. T. Eiter and H. Mannila: Theory distance and similarity: Measures and computation. *Acta informatica* **34**, 2 (1997), 109–133.

60. T. Eiter, H. Mannila and G. Gottlob: On the expressive power of disjunctive logic programming over finite structures. Report, CD-TR 96/90, Christian Doppler Laboratory for Expert Systems, Vienna University of Technology, 1996.

Software Engineering – Tools and Techniques (D.2.2)

61. H. Haapa-aho, H. Hakulinen, A. Hirvonen, T.-K. Kupias, H. Laine, H. Niinistö, S. Räisänen and P. Virkki: *The invasion of objects* (in Finnish). Suomen Atk-kustannus Oy, 1997.

62. M. Hakala, J. Hautamäki, J. Tuomi, A. Viljamaa and J. Viljamaa: Design of a Java framework engineering tool. Report A-1997-12, Department of Computer Science, University of Tampere, 1997.

63. M. Harsu, J. Hautamäki and K. Koskimies: A language implementation framework in Java. In *Proc. Workshop on Language Support for Design Patterns and Object-Oriented Frameworks, LSDF '97* (eds. J. Bosch, G. Hedin and K. Koskimies). Research Report 6/97, University of Karlskrona/Ronneby, 1997.

64. H. Laine: Finnish-English object dictionary. http://www.helsinki.fi/~laine/oliosanasto/, 1997.

65. T. Niklander and K. Raatikainen: Analysis of design tools for potential use in the RODAIN project. Report C-1996-76, Department of Computer Science, University of Helsinki, 1996.

66. J. Paakki (ed.): *Proc. 5th Symposium on Programming Languages and Software Tools*, Jyväskylä, Finland, 1997. Report C-1997-37, Department of Computer Science, University of Helsinki, 1997.

67. J. Paakki, J. Koskinen and A. Salminen: From relational program dependencies to hypertextual access structures. *Nordic Journal of Computing* **4**, 1 (1997), 3–36.

68. J. Viljamaa: Tools supporting the use of design patterns in frameworks. Report C-1997-25, Department of Computer Science, University of Helsinki, 1997.

Testing and Debugging (D.2.5)

69. S.-P. Lahtinen, E. Sutinen, J. Tarhio and A.-P. Tuovinen: Object-oriented visualization of program logic. In *Proc. Technology of Object-Oriented Languages & Systems, TOOLS23*, 1997, 93–107.

Distribution and Maintenance (D.2.7)

70. J. Paakki, A. Salminen and J. Koskinen: Automated hypertext support for software maintenance. *The Computer Journal* **39**, 7 (1996), 577–597.

Programming Languages – Language Classifications (D.3.2)

71. J. Jaakkola, P. Kilpeläinen and G. Lindén: TranSID: An SGML tree transformation language. In *Proc. 5th Symposium on Programming Languages and Software Tools* (ed. J. Paakki), Report C-1997-37, Department of Computer Science, University of Helsinki, 1997, 72–83.

Processors (D.3.4)

72. G. Lindén, H. Tirri and A. I. Verkamo: ALCHEMIST: A general purpose transformation generator. *Software — Practice and Experience*, **26**, 6 (1996), 653–676.

73. A.-P. Tuovinen: Towards a framework for developing processors for visual languages. In *Proc. Workshop on Language Support for Design Patterns and Object-Oriented Frameworks, LSDF '97* (eds. J. Bosch, G. Hedin and K. Koskimies). Research Report 6/97, University of Karlskrona/Ronneby, 1997.

74. A.-P. Tuovinen: A framework for processors of visual languages. In *Proc. Workshop Reader on Object-Oriented Technology, ECOOP '97* (eds. J. Bosch and S. Mitchell), Lecture Notes in Computer Science 1357, Springer, 1998, 119–122.

75. A. P. Tuovinen: Error recovery in parsing relational languages. Submitted 1998.

Data – Data Storage Representations (E.2)

76. O. Nurmi and E. Soisalon-Soininen: Chromatic binary search trees: A structure for concurrent rebalancing. *Acta informatica* **33**, 6 (1996), 547–557.

77. O. Nurmi, E. Soisalon-Soininen and D. Wood: Relaxed AVL trees, main-memory databases and concurrency. *International Journal of Computer Mathematics* **62**, 1 (1996), 23–44.

78. K. Pollari-Malmi, E. Soisalon-Soininen and T. Ylönen: Concurrency control in B-trees with batch updates. *IEEE Transactions on Knowledge and Data Engineering* **8**, 6 (1996), 975–984.

Data Encryption (E.3)

79. P. Orponen: Quantum computers and molecular computation: A threat to public key encryption (in Finnish). *Tietotekniikan maailma*, 1 (1996), 28-33.

Theory of Computation – Computation by Abstract Devices – Models of Computation (F.1.1)

80. P. Orponen: The computational power of discrete Hopfield nets with hidden units. *Neural Computation* **8**, 2 (1996), 403–415.

81. P. Orponen: Computing with truly asynchronous threshold logic networks. *Theoretical Computer Science* **174**, 1-2 (1997), 123–136.

Complexity Classes (F.1.3)

82. H. Buhrman and P. Orponen: Random strings make hard instances. *Journal of Computer and System Sciences* **53**, 2 (1996), 261-266.

Analysis of Algorithms and Problem Complexity – Nonnumerical Algorithms and Problems (F.2.2)

83. G. Das, R. Fleischer, L. Gasieniec, D. Gunopoulos and J. Kärkkäinen: Episode matching. In *Proc. 8th Annual Symposium on Combinatorial Pattern Matching, CPM '97* (eds. A. Apostolico and J. Hein), Lecture Notes in Computer Science 1264, Springer, 1997, 12–27.

84. P. Jokinen, J. Tarhio and E. Ukkonen: A comparison of approximate string matching algorithms. *Software — Practice and Experience* **26**, 12 (1996), 1439–1458.

85. J. Kärkkäinen (ed.): *Proc. Algorithm Day, 10 January 1997.* Report C-1997-11, Department of Computer Science, University of Helsinki, 1997.

86. J. Kärkkäinen and E. Sutinen: Lempel-Ziv index for q-grams. In *Proc. 4th Annual European Symposium on Algorithms, ESA '96* (eds. J. Diaz

and M. Sena), Lecture Notes in Computer Science 1136, Springer, 1996, 378–391.

87. J. Kärkkäinen and E. Ukkonen: Sparse suffix trees. In *Proc. 2nd Annual International Computing and Combinatorics Conference, COCOON '96*, Lecture Notes in Computer Science 1090, Springer, 1996, 219–230.

88. J. Kärkkäinen and E. Ukkonen: Lempel-Ziv parsing and sublinear-size index structures for string matching. In *Proc. 3rd South American Workshop on String Processing, WSP '96*, International informatics series no. 4, Carleton University Press, 1996, 141–155.

89. P. Kilpeläinen: Notes on associative and commutative matching. Submitted for publication 1997.

90. O. Lehtinen, E. Sutinen and J. Tarhio: Experiments on block indexing. In *Proc. 3rd South American Workshop on String Processing, WSP '96*, International informatics series no. 4, Carleton University Press, 1996, 183–193.

91. H. Mannila and P. Ronkainen: Similarity of event sequences. In *Proc. 4th International Workshop on Temporal Representation and Reasoning, TIME '97* (eds. R. Morris and L. Khatib), 1997. IEEE Computer Society Press, 136–139.

92. E. Sutinen: *Using q-Grams in Approximate Pattern Matching*. Ph.Lic. Thesis, Report C-1996-90, Department of Computer Science, University of Helsinki, 1996.

93. E. Sutinen and J. Tarhio: Filtration with q-samples in approximate string matching. In *Proc. 7th Annual Symposium on Combinatorial Pattern Matching, CPM '96* (eds. D. Hirschberg and G. Myers), Lecture Notes in Computer Science 1075, Springer, 1996, 50–63.

94. J. Tarhio: A sublinear algorithm for two-dimensional string matching. *Pattern Recognition Letters* **17**, 8 (1996), 833–838.

95. J. Tarhio and H. Peltola: String matching in the DNA alphabet. *Software — Practice and Experience* **27**, 7 (1997), 851-861.

Logics and Meanings of Programs – Specifying and Verifying and Reasoning about Programs (F.3.1)

96. M. Butler, J. Grundy, T. Långbacka, R. Ruksenas and J. von Wright: The Refinement Calculator: Proof support for program refinement. In *Proc. Formal Methods Pacific: Discrete Mathematics and Theoretical Computer Science, FMP '97* (eds. L. Groves and S. Reeves), Springer, 1997, 40–61.

97. M. Butler and T. Långbacka: Program derivation using the Refinement Calculator. In *Proc. 9th International Conference on Theorem Proving in Higher Order Logics, TPHOLs '96* (eds. J. von Wright, J. Grundy and J. Harrison), Lecture Notes in Computer Science 1125, Springer, 1996, 93–108.

98. J. Eloranta, M. Tienari and A. Valmari: Essential transitions to bisimulation equivalences. *Theoretical Computer Science* **179**, 1/2 (1997), 397–419.

99. K. Granö and J. Paakki: From layer to layer — Object-oriented protocol refinement in Kannel. In *Proc. 2nd International Conference on Coordination Languages and Models, COORDINATION '97* (eds. D. Garlan and D. Le Metayer), Lecture Notes in Computer Science 1282, Springer, 1997, 355–373.

100. R. Kaivola: *Equivalences, Preorders and Compositional Verification for Linear Time Temporal Logic and Concurrent Systems*. Ph.D. Thesis, Report A-1996-1, Department of Computer Science, University of Helsinki, 1996.

101. T. Långbacka: *An Interactive Environment Supporting the Development of Formally Correct Programs*. Ph.D. Thesis, Department of Computer Science, Åbo Akademi University, 1997.

102. T. Långbacka and J. von Wright: Refining reactive systems in HOL using action systems. In *Proc. 10th International Conference on Theorem Proving in Higher Order Logics, TPHOLs '97* (eds. E. L. Gunter and A. Felty), Lecture Notes in Computer Science 1275, Springer, 1997, 183–197

103. A. Valmari and M. Tienari: Compositional failure-based semantic models for basic LOTOS. *Formal Aspects of Computing* **7** (1995), 440–468.

Mathematical Logic and Formal Languages – Mathematical Logic (F.4.1)

104. C. Gefwert, P. Orponen and J. Seppänen (eds.): *Logic, Mathematics and the Computer — Foundations: History, Philosophy and Applications*, vol. 14, Symposium Series, Finnish Artificial Intelligence Society, 1996.

105. J. Grundy and T. Långbacka: Recording HOL proofs in a structured browsable format. In *Proc. 6th International Conference on Algebraic Methodology and Software Technology, AMAST '97*, 1997.

106. R. Kaivola: Axiomatising linear time mu-calculus. In *Proc. 6th International Conference on Concurrency Theory, CONCUR '95* (eds. I. Lee and S. Smolka), Lecture Notes in Computer Science 962, Springer, 1995, 423–437.

107. R. Kaivola: On modal mu-calculus and Büchi tree automata. *Information Processing Letters* **54** (1995), 17–22.

108. R. Kaivola: Axiomatising extended computation tree logic. In *Proc. 21st Colloquium on Trees in Algebra and Programming, CAAP '96* (ed. H. Kirchner), Lecture Notes in Computer Science 1059, Springer, 1996, 87–101.

109. R. Kaivola: Fixpoints for Rabin tree automata make complementation easy. In *Proc. 23rd International Colloquium on Automata, Languages and Programming, ICALP '96* (eds. F. Meyer auf der Heide and B. Monien), Lecture Notes in Computer Science 1099, Springer, 1996, 312–323.

110. R. Kaivola: *Using Automata to Characterise Fixed Point Temporal Logics*. Ph.D. Thesis, Report CST-135-97 (ECS-LFCS-97-356), Department of Computer Science, University of Edinburgh, 1997.

111. R. Kaivola: Axiomatising extended computation tree logic. *Theoretical Computer Science* **190**, 1 (1998), 41–60.

112. T. Långbacka: Developing a graphic theorem proving tool. In *Logic, Mathematics and the Computer — Foundations: History, Philosophy and Applications* (eds. C. Gefwert, P. Orponen and J. Seppänen), vol. 14, Symposium Series, Finnish Artificial Intelligence Society, 1996, 179–188.

113. P. Orponen: On the complexity of computational problems. In *Logic, Mathematics and the Computer — Foundations: History, Philosophy and*

Applications (eds. C. Gefwert, P. Orponen and J. Seppänen), vol. 14, Symposium Series, Finnish Artificial Intelligence Society, 1996, 144–152.

Grammars and Other Rewriting Systems (F.4.2)

114. E. Ohlebusch and E. Ukkonen: On the equivalence problem for E-pattern languages. In *Proc. 21st International Symposium on Mathematical Foundations of Computer Science, MFCS '96* (eds. W. Penczek and A. Szałas), Lecture Notes in Computer Science 1113, Springer, 1996, 457–468.

115. E. Ohlebusch and E. Ukkonen: On the equivalence problem for E-pattern languages. *Theoretical Computer Science* **186**, 1-2 (1997), 231–248.

116. E. Ukkonen: String databases and finite multitape automata. In *Proc. 4th South American Workshop on String Processing, WSP '97*, International informatics series 8, Carleton University Press, 1997, 179.

Formal Languages (F.4.3)

117. P. Kilpeläinen and D. Wood: SGML and exceptions. In *Proc. International Workshop on Principles of Document Processing, PODP '96* (eds. C. Nicholas and D. Wood), Lecture Notes in Computer Science 1293, Springer, 1997, 39–49. (Also technical report HKUST-CS96-30, Department of Computer Science, The Hong Kong University of Science & Technology, 1996).

Mathematics of Computing – Numerical Analysis – Optimization (G.1.6)

118. J. Lahtinen, P. Myllymäki, T. Silander and H. Tirri: Empirical comparison of stochastic algorithms in a graph optimization problem. In *Proc. 2nd Nordic Workshop on Genetic Algorithms and their Applications, 2NWGA* (ed. J. Alander), Report 11, University of Vaasa, 1996, 45–59.

Discrete Mathematics – Combinatorics (G.2.1)

119. O. Yli-Harja, K. Oksanen, K. Egiazarian and J. Astola: The Dedekind problem and perfect shuffle. In *Proc. Algorithm Day, 10 January 1997* (ed. J. Kärkkäinen), Report C-1997-11, Department of Computer Science, University of Helsinki, 1997.

Probability and Statistics (G.3)

120. P. Kontkanen, P. Myllymäki, T. Silander and H. Tirri: A Bayesian approach for retrieving relevant cases. In *Proc. Conference on Artificial Intelligence Applications, EXPERSYS-97* (ed. P. Smith). IITT International, 1997, 67–72.

121. P. Kontkanen, P. Myllymäki, T. Silander and H. Tirri: A Bayesian approach to discretization. In *Proc. European Symposium on Intelligent Techniques*, 1997, 265–268.

122. P. Kontkanen, P. Myllymäki, T. Silander and H. Tirri: On the accuracy of stochastic complexity approximations. To appear in *Causal Models and Intelligent Data Analysis* (ed. A. Gammerman), Springer, 1998. Also in *Proc. Causal Models and Statistical Learning Seminar*, 1997, 103–117.

123. P. Kontkanen, P. Myllymäki, T. Silander and H. Tirri: Comparing stochastic complexity minimization algorithms in estimating missing data. In *Proc. 4th Workshop on Uncertainty Processing, WUPES '97*, 1997, 81–90.

124. P. Kontkanen, P. Myllymäki, T. Silander, H. Tirri and P. Grünwald: On predictive distributions and Bayesian networks. Submitted to *Statistics and Computing*.

125. H. Tirri: *Plausible Prediction by Bayesian Inference*. Ph.D. Thesis, Report A-1997-1, Department of Computer Science, University of Helsinki, 1997.

126. H. Tirri and T. Silander: Stochastic complexity based estimation of missing elements in questionnaire data. To appear in *Proc. Annual Meeting of the American Educational Research Association, AERA '98*, 1998.

127. H. Tirri, T. Silander and K. Tirri: Using neural networks for descriptive statistical analysis of educational data. In *Proc. Annual Meeting of the American Educational Research Association, AERA '97*, SIG Educational Statisticians, ERIC Document Reproduction Service 408 344, 1997. (On microcard.)

128. H. Tirri, T. Silander and K. Tirri: Bayesian finite mixtures for nonlinear modeling of educational data. In *Proc. Annual Meeting of the American Educational Research Association, AERA '97*, Division D,

ERIC Document Reproduction Service 408 345, 1997. (On microcard.)

129. O. Yli-Harja: Statistical characterisation of stack filters. In *Proc. IEEE Nonlinear Signal Processing Workshop*, 1997. (Published as CD-ROM only.)

Information Systems – Database Management (H.2)

130. H. Laine. *A Vocabulary of Relational Databases* (in Finnish). Tietotekniikan kehittämiskeskus, 1996.

131. K. Raatikainen, J. Kiviniemi, T. Niklander and P. Porkka: System design of the RODAIN prototype database. Report C-1997-28, Department of Computer Science, University of Helsinki, 1997.

Physical Design (H.2.2)

132. J. Laine: *Database Reorganization.* Ph.Lic. Thesis, Report C-1997-85, Department of Computer Science, University of Helsinki, 1997.

Languages (H.2.3)

133. L. Akselin and P. Kilpeläinen: Graphical text region queries (in Finnish). In *Proc. SGML Finland 1997* (ed. K. Rytkönen), SGML Users' Group Finland, 1997, 109–122.

134. G. Grahne: Updates and counterfactuals. *Journal of Logic and Computation* **8**, 1 (1998), 87–117.

135. G. Grahne, A. Mendelzon and P. Revesz: Knowledgebase transformations. *Journal of Computer and System Sciences* **54**, 1 (1997), 98–112.

136. G. Grahne and M. Nykänen: Safety, translation and evaluation of alignment calculus. In *Proc. 1st East-European Symposium on Advances in Databases and Information Systems, ADBIS '97*, 1997, 295–304.

137. G. Grahne, M. Nykänen and E. Ukkonen: Reasoning about strings in databases. To appear in *Journal of Computer and System Sciences*.

138. G. Grahne, N. Spyratos and D. Stamate: Semantics and containment of queries with internal and external conjunctions. In *Proc. 6th International Conference on Database Theory, ICDT '97* (eds. F. Afrati and

P. Kolaitis), Lecture Notes in Computer Science 1186, Springer, 1997, 71–82.

139. J. Jaakkola and P. Kilpeläinen: Using sgrep for querying structured text files. In *Proc. SGML Finland 1996* (ed. J. Saarela), SGML User's Group Finland, 1996, 56–67. (Also Report C-1996-83, Department of Computer Science, University of Helsinki, 1996.)

140. M. Nykänen: *Alignment Calculus: A Modal Logic for Querying String Databases*. Ph.Lic. Thesis, Report C-1996-97, Department of Computer Science, University of Helsinki, 1997.

141. M. Nykänen: *Querying String Databases with Modal Logic*. Ph.D. Thesis, Report A-1997-3, Department of Computer Science, University of Helsinki, 1997.

Systems (H.2.4)

142. P. Elovaara and K. Raatikainen: Evaluation of concurrency control algorithms for real-time databases. Report C-1996-52, Department of Computer Science, University of Helsinki, 1996.

143. P. Elovaara and K. Raatikainen: Analysis of WAIT-50 and OCC-TI concurrency control algorithms for real-time databases in telecommunications. Report C-1996-78, Department of Computer Science, University of Helsinki, 1996.

144. T. Imielinski and H. Mannila: A database perspective on knowledge discovery. *Communications of the ACM* **39**, 11 (1996), 58–64.

145. J. Kiviniemi, J. Lindström, T. Niklander, P. Porkka and K. Raatikainen: Experimental design of RODAIN trial. Report C-1997-83, Department of Computer Science, University of Helsinki, 1997.

146. J. Kiviniemi, J. Lindström, T. Niklander, P. Porkka and K. Raatikainen: RODAIN prototype real-time database system, version 2. Report C-1997-84, Department of Computer Science, University of Helsinki, 1997.

147. J. Kiviniemi, T. Niklander, P. Porkka and K. Raatikainen: Transaction processing in the RODAIN real-time database system. In *Real-Time Database and Information Systems* (eds. A. Bestavros and V. Fay-Wolfe), Kluwer Academic Publishers, 1997, 355–375.

148. J. Kiviniemi and K. Raatikainen: Real-time object-oriented data model for telecommunications. Report C-1996-75, Department of Computer Science, University of Helsinki, 1996.

149. A. Lehtola, R. de By, H. Tirri and J. Wäsch: The TransCoop architecture. Chapter 5 in *Transaction Management Support for Cooperative Applications* (eds. R. de By et al.), Kluwer 1998, 93-111.

150. H. Mannila: Data mining: Machine learning, statistics, and databases. In *Proc. 8th International Conference on Scientific and Statistical Database Management, SSDBM '96*, IEEE Computer Society Press, 1996, 2–9.

151. H. Mannila: Schema design and knowledge discovery. In *Proc. 15th International Conference on Conceptual Modeling, ER '96*, Lecture Notes in Computer Science 1157, Springer, 1996, 27.

152. J. Puustjärvi: Negotiation transactions: An approach to increase the automation of workflows. In *Proc. 9th International Conference on Advanced Information Systems Engineering, CAiSE '97* (eds. A. Olivé and J. A. Pastor), Lecture Notes in Computer Science 1250, 1997, 89–102.

153. J. Puustjärvi, H. Tirri and J. Veijalainen: Concurrency control for overlapping and cooperative workflows. *Bulletin of the Technical Committee on Operating Systems and Application Environments, TCOS*, **8**, 1 (1996), 24–30.

154. J. Puustjärvi, H. Tirri and J. Veijalainen: Managing overlapping transactional workflows. In *Proc. 8th International Conference on Advanced Information Systems Engineering, CAiSE '96* (eds. P. Constantopoulos, J. Mypoulos and Y. Vassiliou), Lecture Notes in Computer Science 1080, Springer, 1996, 345–361.

155. J. Puustjärvi, H. Tirri and J. Veijalainen: Reusability and modularity in transactional workflows. *Information Systems* **22**, 2/3 (1997), 101–120.

156. K. Raatikainen: Real-time databases in telecommunications. In *Real-time Database Systems: Issues and Applications*, International Series in Engineering and Computer Science, Real-time Systems, Kluwer, 1997, 93–98.

157. S. Sippu and E. Soisalon-Soininen: An analysis of magic sets and related optimization strategies for logic queries. *Journal of the ACM* **43**, 6 (1996), 1046–1088.

158. J. Veijalainen, J. Wäsch, J. Puustjärvi, H. Tirri and O. Pihlajamaa: Transaction models in cooperative work — An overview. Chapter 3 in *Transaction Management Support for Cooperative Applications* (eds. R. de By et al.), Kluwer 1998, 27-58.

Heterogeneous Databases (H.2.5)

159. A.-P. Tuovinen: Using compiler techniques to convert an SQL application for a new DBMS. In *Proc. 4th Symposium on Programming Languages and Software Tools* (ed. Laszlo Varga), Department of General Computer Science, Eötvös Lorand University of Budapest, 1995, 349–359.

160. A.-P. Tuovinen and J. Paakki: Translating SQL for database reengineering. *ACM SIGPLAN Notices* **31**, 2 (1996), 21–26.

Information Storage and Retrieval – Content Analysis and Indexing (H.3.1)

161. R. Agrawal, H. Mannila, R. Srikant, H. Toivonen and A. I. Verkamo: Fast discovery of association rules. In *Advances in Knowledge Discovery and Data Mining* (eds. U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth and R. Uthurusamy), AAAI Press, 1996, 307–328.

162. Y. Huhtala, J. Kärkkäinen, P. Porkka and H. Toivonen: Efficient discovery of functional and approximate dependencies using partitions. In *Proc. 14th International Conference on Data Engineering, ICDE '98*, IEEE Computer Society Press, 1998, 392-401.

163. H. Mannila and H. Toivonen: Discovering generalized episodes using minimal occurrences. In *Proc. 2nd International Conference on Knowledge Discovery and Data Mining, KDD '96*, AAAI, 1996, 146–151.

164. H. Mannila, H. Toivonen and A. I. Verkamo: Discovery of frequent episodes in event sequences. *Data Mining and Knowledge Discovery* **1**, 3 (1997), 259–289.

165. H. Toivonen: Sampling large databases for finding frequent sets. In *Proc. 22nd International Conference on Very Large Data Bases, VLDB '96* (eds. T. Viljayaraman, A. Buchmann, C. Mohan and N. Sarda), Morgan Kaufmann, 1996, 134–145.

166. H. Toivonen: *Discovery of Frequent Patterns in Large Data Collections*. Ph.D. Thesis, Report A-1996-5, Department of Computer Science, University of Helsinki, 1996.

Information Search and Retrieval (H.3.3)

167. H. Ahonen, O. Heinonen, M. Klemettinen and A. I. Verkamo: Mining in the phrasal frontier. In *Proc. 1st European Symposium on Principles of Data Mining and Knowledge Discovery, PKDD '97* (eds. J. Komorowski and J. Zytkow), Lecture Notes in Computer Science 1263, Springer, 1997, 343–350.

168. H. Ahonen, O. Heinonen, M. Klemettinen and A. I. Verkamo: Mining in the phrasal frontier. Report C-1997-14, Department of Computer Science, University of Helsinki, 1997.

169. H. Ahonen, O. Heinonen, M. Klemettinen and A. I. Verkamo: Applying data mining techniques in text analysis. Report C-1997-23, Department of Computer Science, University of Helsinki, 1997.

170. H. Ahonen, O. Heinonen, M. Klemettinen and A. I. Verkamo: Applying data mining for descriptive phrase extraction in digital document collections. To appear in *Proc. IEEE Conference on Advances in Digital Libraries, ADL '98*, 1998.

171. G. Das, D. Gunopoulos and H. Mannila: Finding similar time series. In *Proc. 1st European Symposium on Principles of Data Mining and Knowledge Discovery, PKDD '97* (eds. J. Komorowski and J. Zytkow), Lecture Notes in Computer Science 1263, Springer, 1997, 88–100.

172. G. Das, H. Mannila and P. Ronkainen: Similarity of attributes by external probes. Report C-1997-66, Department of Computer Science, University of Helsinki, 1997.

173. D. Gunopoulos, H. Mannila and S. Saluja: Discovering all most specific sentences by randomized algorithms (extended abstract). In *Proc. 6th International Conference on Database Theory, ICDT '97* (eds. F. Afrati and P. Kolaitis), Lecture Notes in Computer Science 1186, Springer, 1997, 215–229.

174. D. Gunopulos, S. Saluja and H. Mannila: Discovering all most specific sentences by randomized algorithms. Research Report, Max-Planck-Institut für Informatik, Saarbrücken, 1996.

175. K. Hätönen, M. Klemettinen, H. Mannila, P. Ronkainen and H. Toivonen: Rule discovery in alarm databases. Report C-1996-7, Department of Computer Science, University of Helsinki, 1996.

176. K. Hätönen, M. Klemettinen, H. Mannila, P. Ronkainen and H. Toivonen: Knowledge discovery from telecommunication network alarm databases. In *Proc. 12th International Conference on Data Engineering, ICDE 96* (ed. S. Su), IEEE Computer Society Press, 1996, 115–122.

177. D. Heckerman, H. Mannila, D. Pregibon and R. Uthurusamy (eds.): *Proc. 3rd International Conference on Knowledge Discovery and Data Mining, KDD '97*, AAAI Press, 1997.

178. K. Korpimies: *World-Wide Web Navigation* (in Finnish). Suomen Atkkustannus Oy, 1995.

179. K. Korpimies and E. Ukkonen: Searching for general documents. To appear in *Proc. Conference on Flexible Query Answering Systems, FQAS '98*, Lecture Notes in Artificial Intelligence, Springer, 1998.

180. K. Lemström, J. Korte, P. Kuusi, P. Kyheröinen and P. Päiväkumpu: An open platform for an image retrieval System. In *Proc. 6th Scandinavian Conference On Artificial Intelligence, SCAI '97* (ed. G. Grahne), Frontiers in Artificial Intelligence and Applications 40, IOS Press, 1997, 277–278.

181. K. Lemström, P. Kuusi, P. Kanerva, S. Koponen, S. Männynsalo, J. Riihimäki, J. Teinilä and T. Virtanen: A client-server tool for image retrieval applications. In *The Challenge of Image Retrieval* (eds. J. P. Eakins, D. J. Harper and J. Jose), 1998.

182. K. Lemström, J. Korte, P. Kuusi, P. Kyheröinen and P. Päiväkumpu: PICSearch — A platform for image content-based searching algorithms. In *Proc. 6th International Conference in Central Europe on Computer Graphics and Visualisation, WSCG '98* (ed. V. Skala), 1998, 222-229.

183. H. Mannila: From data to knowledge: Data mining and enrichment (in Finnish). *Tietoyhteys* **1**, 4 (1997), 4–5.

184. H. Mannila: Methods and problems in data mining. In *Proc. 6th International Conference on Database Theory, ICDT '97* (eds. F. Afrati and P. Kolaitis), Lecture Notes in Computer Science 1186, Springer, 1997, 41–55.

185. H. Mannila: Inductive databases and condensed representations: concepts for data mining. To appear in *Proc. International Logic Program Symposium*, 1997.

186. H. Toivonen: Data mining: Digging into mountains of data in the search for knowledge (in Finnish). *MikroPC*, 12 (1997), 62–66.

Information Interfaces and Presentation – User Interfaces (H.5.2)

187. H. Erkiö: The purpose of evaluations, their benefits, and methods (in Finnish). In *Design of graphical user interfaces* (in Finnish, ed. A. Kalimo), Suomen Atk-kustannus Oy, 1996, 83–88.

188. H. Erkiö: Model-based evaluation methods (in Finnish). In *Design of graphical user interfaces* (in Finnish, ed. A. Kalimo), Suomen Atk-kustannus Oy, 1996, 113–123.

Computing Methodologies – Artificial Intelligence (I.2)

189. T. Elomaa: Artificial intelligence does not match human intelligence (in Finnish). Guest author column, *Helsingin Sanomat*, 20 March 1996, A2.

190. G. Grahne (ed.): *Proc. 6th Scandinavian Conference on Artificial Intelligence, SCAI '97*, Frontiers in Artificial Intelligence and Applications 40, IOS Press, 1997.

191. G. Grahne (ed.): *Research Announcements of the 6th Scandinavian Conference on Artificial Intelligence, SCAI '97*, Report C-1997-49, Department of Computer Science, University of Helsinki, 1997.

192. P. Myllymäki and H. Tirri: *Prospects of Bayesian Networks* (in Finnish). Technology Review 58/98, The Technology Development Centre (TEKES), 1998.

Applications and Expert Systems (I.2.1)

193. E. Koskimäki, J. Göös, P. Kontkanen, P. Myllymäki and H. Tirri: Comparing soft computing methods in prediction of manufacturing data. To appear in *Proc. 11th International Conference on Industrial and Engineering Applications of Artificial Intelligence & Expert Systems, IEA-98-AIE*, 1998.

Knowledge Representation Formalisms and Methods (I.2.4)

194. S. Eloranta: On iterated belief change. In *Proc. 6th Scandinavian Conference on Artificial Intelligence* (ed. G. Grahne), Frontiers in Artificial Intelligence and Applications 40, IOS Press, 1997, 83-94.

Learning (I.2.6)

195. R. Aarts and J. Rousu: Towards CBR for bioprocess planning. In *Proc. 3rd European Workshop on Advances in Case-Based Reasoning, EWCBR '96* (eds. I. Smith and B. Faltings), Lecture Notes in Artificial Intelligence 1168, Springer, 1996, 16-27.

196. R. Aarts and J. Rousu: Adaptation cost as a criterion for solution evaluation. In *Proc. 3rd European Workshop on Advances in Case-Based Reasoning, EWCBR '96* (eds. I. Smith and B. Faltings) Lecture Notes in Artificial Intelligence 1168, Springer, 1996, 354–361.

197. R. Aarts and J. Rousu: Qualitative knowledge to support reasoning about cases. In *Proc. 2nd International Conference in Case-Based Reasoning, ICCBR '97*, Lecture Notes in Artificial Intelligence 1266, Springer, 1997, 489–498.

198. H. Ahonen: Automatic generation of SGML content models. In *Proc. 6th International Conference on Electronic Publishing, Document Manipulation and Typography, EP '96* (eds. A. Brown, A. Brüggemann-Klein and A. Feng), Wiley Publishers, 1996, 195–206.

199. H. Ahonen: Disambiguation of SGML content models. In *Proc. International Workshop on Principles of Document Processing, PODP '96* (eds. C. Nicholas and D. Wood), Lecture Notes in Computer Science 1293, Springer, 1997, 27–37.

200. H. Ahonen: Generation of SGML DTDs for tagged documents. In *Proc. SGML Finland 1996* (ed. J. Saarela), SGML User's Group Finland, 1996, 40–46.

201. H. Ahonen: *Generating Grammars for Structured Documents Using Grammatical Inference Methods*. Ph.D. Thesis, Report A-1996-4, Department of Computer Science, University of Helsinki, 1996.

202. H. Ahonen, H. Mannila and E. Nikunen: Generating grammars for SGML tagged texts lacking DTD. *Mathematical and Computer Modelling* **26**, 1 (1997), 1–13.

203. A. Brazma, I. Jonassen, E. Ukkonen and J. Vilo: Discovering patterns and subfamilies in biosequences. In *Proc. 4th International Conference on Intelligent Systems for Molecular Biology, ISMB '96*, AAAI Press, 1996, 34–43.

204. A. Brazma, E. Ukkonen and J. Vilo: Discovering unbounded unions of regular pattern languages from positive examples (extended abstract). In *Proc. 7th International Symposium on Algorithms and Computation, ISAAC '96*, Lecture Notes in Computer Science 1178, Springer, 1996, 95–104.

205. A. Brazma, J. Vilo and E. Ukkonen: Finding transcription factor binding site combinations in the yeast genome (extended abstract). In *Computer Science and Biology: Proc. German Conference on Bioinformatics, GCB '97*, Genome informatics: Function, Structure, Phylogeny, MIPS Munich Information Center for Protein Sequences, 1997, 57–59.

206. A. Brazma, J. Vilo, E. Ukkonen and K. Valtonen: Data mining for regulatory elements in yeast genome. In *Proc. 5th International Conference on Intelligent Systems for Molecular Biology, ISMB '97*, AAAI Press, 1997, 65–74.

207. L. Dehaspe and H. Toivonen: Frequent query discovery: A unifying ILP approach to association rule mining. Report CW-258, Department of Computer Science, Katholieke Universiteit Leuven, Belgium, 1998.

208. T. Elomaa: *Tools and Techniques for Decision Tree Learning*. Ph.D. Thesis, Report A-1996-2, Department of Computer Science, University of Helsinki, 1996.

209. T. Elomaa: Machine intelligence and learning (invited public lecture). In *Proc. 7th Finnish Artificial Intelligence Conference, STeP '96: Genes, Nets and Symbols* (eds. J. Alander, T. Honkela and M. Jakobsson), Finnish Artificial Intelligence Society, 1996, 90–95.

210. T. Elomaa: The vacuum-cleaner robot is not yet at sight. Commentary, Computing Section, *Helsingin Sanomat*, 13 December 1996, D2.

211. T. Elomaa and J. Rousu: Finding optimal multi-splits for numerical attributes in decision tree learning. In *Proc. 7th Finnish Artificial Intelligence Conference, STeP '96: Genes, Nets and Symbols* (eds. J. Alander,

T. Honkela and M. Jakobsson), Finnish Artificial Intelligence Society, 1996, 104–111.

212. T. Elomaa and J. Rousu: Efficient multisplitting on numerical data. In *Proc. 1st European Symposium on Principles of Data Mining and Knowledge Discovery, PKDD '97* (eds. J. Komorowski and J. Zytkow), Lecture Notes in Artificial Intelligence 1263, Springer, 1997, 178–188.

213. T. Elomaa and J. Rousu: On the well-behavedness of important attribute evaluation functions. In *Proc. 6th Scandinavian Conference on Artificial Intelligence* (ed. G. Grahne), Frontiers in Artificial Intelligence and Applications 40, IOS Press, 1997, 95–106.

214. T. Elomaa and J. Rousu: Well-behaved attribute evaluation functions for numerical attributes. In *Proc. 10th International Symposium on Foundations of Intelligent Systems* (eds. Z. Ras and A. Skowron), Lecture Notes in Artificial Intelligence 1325, Springer, 1997, 147–156.

215. T. Elomaa and J. Rousu: General and efficient multisplitting of numerical attributes. To appear in *Machine Learning*. (Report C-1996-82, Department of Computer Science, University of Helsinki, 1996, and Technical Note No. I.98.06, Institute for Systems, Informatics and Safety, Joint Research Centre, European Commission, 1998.)

216. K. Fredriksson: Genetic algorithms and generative encoding of neural networks for some benchmark classification problems. In *Proc. 3rd Nordic Workshop on Genetic Algorithms and their Applications, 3NWGA* (ed. J. Alander), Finnish Artificial Intelligence Society, 1997, 123-134.

217. P. Grünwald, P. Kontkanen, P. Myllymäki, T. Silander and H. Tirri: Minimum encoding approaches for predictive modeling. To appear in *Proc. 14th Conference on Uncertainty in Artificial Intelligence, UAI ´98*, 1998.

218. D. P. Helmbold, J. Kivinen and M. K. Warmuth: Worst-case loss bounds for sigmoided neurons. In *Advances in Neural Information Processing Systems 8* (eds. D. S. Touretzky, M. C. Mozer and M. E. Hasselmo), MIT Press, 1996, 309–315.

219. M. Jaeger, H. Mannila and E. Weydert: Data mining as selective theory extraction in probabilistic logic. In *Proc. ACM SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery, SIGMOD '96 DMKD*, 1996.

220. J. Kivinen and M. K. Warmuth: Exponentiated gradient versus gradient descent for linear predictors. *Information and Computation*, **132**, 1 (1997), 1–63.

221. J. Kivinen and M. K. Warmuth: Relative loss bounds for multidimensional regression problems. To appear in *Advances in Neural Information Processing Systems 1997, NIPS '97*, 1997.

222. J. Kivinen, M. K. Warmuth and P. Auer: The perceptron algorithm versus Winnow: Linear versus logarithmic mistake bounds when few input variables are relevant. *Artificial Intelligence*, **97**, 1-2 (1997), 325–343.

223. P. Kontkanen, P. Myllymäki, T. Silander and H. Tirri: Bayes optimal instance-based learning. In *Proc. 10th European Conference on Machine Learning, ECML '98* (eds. C. Nédellec and C. Rouveirol), Lecture Notes in Artificial Intelligence 1398, Springer, 1998, 77–88.

224. P. Kontkanen, P. Myllymäki, T. Silander and H. Tirri: Batch classifications with discrete finite mixtures. In *Proc. 10th European Conference on Machine Learning, ECML '98* (eds. C. Nédellec and C. Rouveirol), Lecture Notes in Artificial Intelligence 1398, Springer, 1998, 89–94.

225. P. Kontkanen, P. Myllymäki, T. Silander and H. Tirri: Bayesian Classification and Feature Selection with BAYDA. In *10th European Conference on Machine Learning, ECML '98: Demonstration and Poster Papers* (eds. C. Nédellec and C. Rouveirol), Report CSR-98-07, Technische Universität Chemnitz, 1998.

226. P. Kontkanen, P. Myllymäki, T. Silander, H. Tirri and P. Grünwald: Comparing predictive inference methods for discrete domains. In *Proc. 6th International Workshop on Artificial Intelligence and Statistics*, 1997, 311–318.

227. P. Kontkanen, P. Myllymäki, T. Silander, H. Tirri and P. Grünwald: Bayesian and information-theoretic predictive distributions for Bayesian networks. In *Proc. 7th Belgian-Dutch Conference on Machine Learning, BeNeLearn '97* (eds. W. Daelemans, P. Flach and A. van den Bosch), 1997, 59–68.

228. P. Kontkanen, P. Myllymäki, T. Silander, H. Tirri and P. Grünwald: Bayesian and information-theoretic priors for Bayesian network parameters. In *Proc. 10th European Conference on Machine Learning,*

ECML '98 (eds. C. Nédellec and C. Rouveirol), Lecture Notes in Artificial Intelligence 1398, Springer, 1998, 208–213.

229. P. Kontkanen, P. Myllymäki and H. Tirri: Some experimental results with finite mixture models. In *Proc. 1st European Conference on Highly Structured Stochastic Systems*, 1996, 112–115.

230. P. Kontkanen, P. Myllymäki and H. Tirri: Predictive data mining with finite mixtures. In *Proc. 2nd International Conference on Knowledge Discovery and Data Mining, KDD '96*, 1996, 176–182.

231. P. Kontkanen, P. Myllymäki and H. Tirri: Comparing Bayesian model class selection criteria by discrete finite mixtures. In *Proc. Conference on Information, Statistics and Induction in Science, ISIS '96*, (eds. D. L. Dowe, K. B. Korb and J. J. Oliver), World Scientific, 1996, 364–374.

232. P. Kontkanen, P. Myllymäki and H. Tirri: Experimenting with the Cheeseman-Stutz evidence approximation for predictive modeling and data mining. In *Proc. 10th Annual Florida Artificial Intelligence International Research Symposium, FLAIRS '97*, 1997, 204–211.

233. P. Kontkanen, P. Myllymäki and H. Tirri: Constructing Bayesian finite mixture models by the EM algorithm. Technical Report NC-TR-97-003, ESPRIT Working Group on Neural and Computational Learning (NeuroCOLT), 1997.

234. H. Mannila: Data mining and machine learning. In *Proc. 13th International Conference on Machine Learning, ICML '96*, Morgan Kaufmann, 1996, 555.

235. H. Mannila and H. Toivonen: Multiple uses of frequent sets and condensed representations. In *Proc. 2nd International Conference on Knowledge Discovery and Data Mining, KDD '96*, AAAI, 1996, 189–194.

236. H. Tirri, P. Kontkanen and P. Myllymäki: Probabilistic instance-based learning. In *Proc. 13th International Conference on Machine Learning* (ed. L. Saitta). Morgan Kaufmann, 1996, 507–515.

237. H. Tirri, P. Kontkanen and P. Myllymäki: A Bayesian framework for case-based reasoning. In *Advances in Case-Based Reasoning* (eds. I. Smith and B. Faltings). Lecture Notes in Artificial Intelligence 1168, Springer, 1996, 413–427.

Problem Solving, Control Methods and Search (I.2.8)

238. N. Bollobás, G. Das, D. Gunopoulos and H. Mannila: Time-series similarity problems and well-separated geometric sets. In *Proc. 13th Annual Symposium on Computational Geometry*, ACM, 1997, 454–456.

239. R. Greiner and P. Orponen: Probably approximately optimal satisficing strategies. *Artificial Intelligence* **82**, 1-2 (1996), 21–44.

240. D. Gunopoulos, R. Khardon, H. Mannila and H. Toivonen: Data mining, hypergraph transversals, and machine learning (extended abstract). In *Proc. 16th ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, PODS '97*, 1997, 209–216.

241. H. Mannila and H. Toivonen: On an algorithm for finding all interesting sentences. In *13th European Meeting on Cybernetics and Systems Research* (ed. R. Trappl), vol. 2, Austrian Society for Cybernetic Studies, 1996, 973-978.

242. H. Mannila and H. Toivonen: Levelwise search and borders of theories in knowledge discovery. *Data Mining and Knowledge Discovery* **1**, 3 (1997), 241–258.

Image Processing – Enchancement (I.4.3)

243. K. Lemström, J. Tarhio and T. Takala: Color dithering with n-best algorithm. In *Proc. 4th International Conference in Central Europe on Computer Graphics and Visualization, WSCG '96* (ed. N. Thalmann and V. Skala), Department of Computer Science, University of West Bohemia, 1996, 162–170.

Pattern Recognition – Clustering (I.5.3)

244. O. Heinonen and H. Mannila: Attribute-oriented induction and conceptual clustering. Report C-1996-2, Department of Computer Science, University of Helsinki, 1996.

Simulation and Modeling – Simulation Support Systems (I.6.7)

245. E. Arjas, H. Mannila, M. Salmenkivi, R. Suramo and H. Toivonen: BASS: Bayesian analyzer of event sequences. In *Proc. XII Symposium on Computational Statistics, COMPSTAT '96*, 1996, 199–204.

246. H. Toivonen, H. Mannila, M. Salmenkivi and K.-P. Laakso: Bassist — a tool for MCMC simulation of statistical models. To appear in *3rd International Congress of the Federation of European Simulation Societies, EUROSIM '98*, 1998.

Text Processing – Document Preparation (I.7.2)

247. H. Ahonen, B. Heikkinen, O. Heinonen, J. Jaakkola, P. Kilpeläinen and G. Lindén: Design and implementation of a document assembly workbench. In *Proc. International Conference on Electronic Publishing, Artistic Imaging, and Digital Typography, EP '98/RIDT '98* (eds. R. D. Hersch, J. André and H. Brown), Lecture Notes in Computer Science 1375, 1998, 476–486.

248. H. Ahonen, B. Heikkinen, O. Heinonen, J. Jaakkola, P. Kilpeläinen, G. Lindén and H. Mannila: Intelligent assembly of structured documents. Report C-1996-40, Department of Computer Science, University of Helsinki, 1996.

249. H. Ahonen, B. Heikkinen, O. Heinonen, J. Jaakkola, P. Kilpeläinen, G. Lindén and H. Mannila: Constructing tailored SGML documents. In *Proc. SGML Finland 1996* (ed. J. Saarela), SGML User's Group Finland, 1996, 106–116.

250. H. Ahonen, B. Heikkinen, O. Heinonen, J. Jaakkola and M. Klemettinen: Analysis of document structures for element type classification. In *Preliminary Proc. 4th International Workshop on Principles of Digital Document Processing, PODDP '98* (eds. C. Nicholas and D. Wood), 1998.

251. H. Ahonen, B. Heikkinen, O. Heinonen and P. Kilpeläinen: A system for assembling specialized textbooks from a pool of documents. Report C-1997-22, Department of Computer Science, University of Helsinki, 1997.

252. H. Ahonen, B. Heikkinen, O. Heinonen and P. Kilpeläinen: Assembling documents from digital libraries. In *Proc. 8th International Conference on Database and Expert Systems Applications, DEXA '97* (eds. A. Hameurlain and A. M. Tjoa), Lecture Notes in Computer Science 1308, Springer, 1997, 419–429.

253. H. Ahonen, B. Heikkinen, O. Heinonen and M. Klemettinen: Improving the accessibility of SGML documents — A content-

analytical approach. In *Proc. SGML Europe '97*, Graphics Communications Association, 1997, 321–327.

254. H. Ahonen, B. Heikkinen, O. Heinonen and M. Klemettinen: Discovery of reasonably-sized fragments using inter-paragraph similarities. Report C-1997-67, Department of Computer Science, University of Helsinki, 1997.

255. P. Kilpeläinen: The utilization of the content information of books (in Finnish). Manuscript of a chapter in the final report of the SÄÄTÖ II project, The Finnish National Board of Education, 1998.

256. G. Lindén: *Structured Document Transformations*. Ph.D. Thesis, Report A-1997-2, Department of Computer Science, University of Helsinki, 1997.

Computer Applications – Physical Sciences and Engineering (J.2)

257. A. Korhola, J. Weckström, H. Olander, H. Seppä, T. Blom, S. Sorvari, J. Virkanen, H. J .B. Birks, H. Toivonen and H. Mannila: Climate changes as recorded by ecologically sensitive arctic lakes in Fennoscandia. To appear in *Proc. International SCANTRAN Meeting: An IGBP Terrestrial Transect for Scandinavia/Northern Europe*, 1998.

Life and Medical Sciences (J.3)

258. M. Eerola, H. Mannila and M. Salmenkivi: Frailty factors and time-dependent hazards in modelling ear infections in children using BASSIST. To appear in *Proc. XIII Symposium on Computational Statistics, COMPSTAT '98*, 1998.

259. M. Huttunen, E. Ukkonen and B. Vehviläinen: Neural networks as a part of watershed-model in ice-reduction of discharge observations. In *Proc. Nordic Hydrological Conference*, NHP rapport 0900-0267 nr. 40, Icelandic Hydrological Committee, 1996, 286–293.

260. M. Huttunen, E. Ukkonen and B. Vehviläinen: Coding a conceptual model into a neural network in modeling ice-correction. In *Proc. Conference Artificial Neural Networks in Engineering: Smart Engineering Systems: Neural Networks, Fuzzy Logic and Evolutionary Programming, ANNIE '96*, Intelligent Engineering Systems through Artificial Neural Networks, ASME Press, 1996, 1001–1006

261. M. Huttunen, E. Ukkonen and B. Vehviläinen: Using trainable computing networks in the optimization of lake regulation. In *Proc. International Conference on Neural Information Processing and Intelligent Information Systems: Progress in Connectionist-Based Information Systems*, vol. 2, Springer, 1998, 975–978.

262. M. Huttunen, B. Vehviläinen and E. Ukkonen: Neural networks in the ice-correction of discharge observations. *Nordic Hydrology* **28**, 4/5 (1997), 283–296.

263. R. Krahe, H. Julkunen, H. Mannila, H. Lokki, V. Ollikainen, J. Ignatius, A. de la Chapelle and J. Kere: Exclusion of three candidate susceptibility regions on human chromosomes 1q31-q42, 5q23-q32 and 9p22-p21 for systemic lupus erythematosus (SLE) in a Finnish founder population. In *Proc. 47th Annual Meeting of the American Society of Human Genetics, American Journal of Human Genetics* **61** (suppl.), A1638, 1997.

Arts and Humanities (J.5)

264. T. Kirkinen and H. Toivonen: GIS-assisted data analysis — Finding meanings in complex spatial data sets. To appear in *Proc. 26th Conference on Computer Applications in Archaeology, CAA '98*, 1998.

Computers in Other Systems (J.7)

265. J. Laakso, H. Aaltonen, A. Leppävuori, H. Toivonen, V. Kilpi, H. Mononen and J. Viitasalo: Using laser radar in ski testing (in Finnish). Report, Research Institute for Olympic Sports, 1997.

Computing Milieux – Computers and Education – Computer Uses in Education (K.3.1)

266. T. Eriksson, J. Kurhila and E. Sutinen: An agent-based framework for special needs education. In *Proc. 6th Scandinavian Conference on Artificial Intelligence, SCAI '97* (ed. G. Grahne), Frontiers in Artificial Intelligence and Applications 40, IOS Press, 1997, 270–271.

267. D. Joyce, J. Gerhardt-Powals, E. Koffman, W. Kreuzer, C. Laxer, K. Loose, E. Sutinen and R. A. Whitehurst: Developing laboratories for the SIGCSE Computing Laboratory Repository: Guidelines, recommendations, and sample labs. In *Working Group Reports and*

Supplemental Proc. on Integrating Technology into Computer Science Education, ITiCSE '97, ACM, 1997, 1–12.

268. S.-P. Lahtinen, E. Sutinen, A.-P. Tuovinen and J. Tarhio: Learning and problem solving assisted with animated objects. In *Abstracts of the European Conference on Educational Research in 1995, ECER '95*, University of Bath, 1995, 153.

269. M. Mäkelä: Experiences in computer uses in education (in Finnish). *Peda-forum: korkeakoulupedagoginen tiedotuslehti* **3**, 1 (1996), 3–5.

270. M. Mäkelä: Information technology — a tool and an obstacle in the education of the future. In *Information Technology in Educational Management for the Schools of the Future* (eds. A. Fung, A. Visscher, B. Barta and D. Teather), Chapman & Hall, 1997, 23–30.

271. M. Mäkelä et al.: Impact of information and communication technology on the management of future schools. In *Information Technology in Educational Management for the Schools of the Future* (eds. A. Fung, A. Visscher, B. Barta and D. Teather), Chapman & Hall, 1997, 185–188.

272. M. Parkkinen, P. Parkkinen and E. Sutinen: On the feasibility of multimedia in a Tanzanian context. In *Matematiikan ja luonnontieteiden opetuksen tutkimuspäivien esitykset* (ed. V. Meisalo), Research Report 162, Department of Teacher Education, University of Helsinki, 1996, 173–179.

Computer and Information Science Education (K.3.2)

273. J. Haajanen, M. Pesonius, E. Sutinen, J. Tarhio, T. Teräsvirta and P. Vanninen: Animation of user algorithms on the web. In *Proc. IEEE Symposium on Visual Languages, VL '97*, IEEE Computer Society Press, 1997, 360–367.

274. J. Kurhila: WWW supporting computer graphics studies (in Finnish). In *Proc. Research Seminar on Digital Media: Computer-Aided Instruction* (in Finnish), Helsinki University of Technology, 1997, 121–132.

275. J. Kurhila and M. Mäkelä: WWW as a source to support computer graphics studies. Published 1997 at http://www.cs.helsinki.fi/group/goa/artikkeli/article.html.

276. J. Kurhila and M. Mäkelä: Ready to use learning material from the Web — a utopia or a reality? (in Finnish). *Peda-forum: korkeakoulupedagoginen tiedotuslehti* **4**, 2 (1997), 35–37.

277. S.-P. Lahtinen, T. Lamminjoki, V. Ollikainen, E. Sutinen, J. Tarhio, and A.-P. Tuovinen: Animation technology as a method of learning to write algorithms. In *Matematiikan ja luonnontieteiden opetuksen tutkimuspäivien esitykset* (ed. V. Meisalo), Research Report 162, Department of Teacher Education, University of Helsinki, 1996, 147–151.

278. S.-P. Lahtinen, T. Lamminjoki, E. Sutinen, J. Tarhio and A.-P. Tuovinen: Towards automated animation of algorithms. In *Proc. 4th International Conference in Central Europe on Computer Graphics and Visualization, WSCG '96* (ed. N. Thalmann and V. Skala), Department of Computer Science, University of West Bohemia, 1996, 150–161.

279. S.-P. Lahtinen, V. Meisalo, E. Sutinen and J. Tarhio: Unified challenges of algorithm design and problem solving. In *Matematiikan ja luonnontieteiden opetuksen tutkimuspäivien esitykset* (ed. V. Meisalo), Research Report 162, Department of Teacher Education, University of Helsinki, 1996, 153–161.

280. S.-P. Lahtinen, A. Porttikivi, E. Sutinen and A.-P. Tuovinen: Algorithm animation as a method of learning. In *Uudet menetelmät ja mahdollisuudet matemaattis-luonnontieteellisten aineiden oppimisessa* (ed. J. Enkenberg and K. Sormunen), University of Joensuu, 1994, 65–70.

281. S.-P. Lahtinen, E. Sutinen and J. Tarhio: Automated animation of algorithms with Eliot. To appear in *Journal of Visual Languages and Computing*.

282. J. Markkanen, E. Rautama, E. Sutinen, J. Tarhio and T. Teräsvirta: Animation of algorithms with Eliot and Jeliot (demonstration). In *Working Group Reports and Supplemental Proc. on Integrating Technology into Computer Science Education, ITiCSE '97*, ACM, 1997, 152.

283. V. Meisalo, E. Rautama, E. Sutinen and J. Tarhio: Teaching algorithms with animation: A case study using Eliot. In *Proc. International Symposium on Learning Technology and Telematics in Education and Training, LeTTET '96*, University of Joensuu, 1997, 79–84.

284. V. Meisalo, E. Sutinen and J. Tarhio: CLAP: Teaching data structures in a creative way. In *Proc. Integrating Technology into Computer Science Education, ITiCSE 97* (ed. G. Davies), *ACM SIGCSE bulletin* **29**, 3 (1997), 117–119.

285. V. Meisalo, E. Sutinen and J. Tarhio: Impacts of interdisciplinary dialogue to computer science education. To appear in *Proc. IFIP WG 3.2. Working Conference 1997: Computer Science as a Discipline and in other Disciplines*, IFIP, 1997.

286. V. Meisalo, E. Sutinen, J. Tarhio and T. Teräsvirta: Combining algorithmic and creative problem solving on the Web. To appear in *Proc. Teleteaching '98*.

287. E. Rautama, E. Sutinen and J. Tarhio: Learning algorithms by animation (in Finnish). *Peda-Forum: korkeakoulupedagoginen tiedotuslehti* **3**, 2 (1996), 21

288. E. Rautama, E. Sutinen and J. Tarhio: From algorithm to animation (in Finnish). *Ote — opetus & teknologia* **4**, 6 (1996), 24–26.

289. E. Rautama, E. Sutinen and J. Tarhio: Excel as an algorithm animation environment. *Proc. Integrating Technology into Computer Science Education, ITiCSE 97* (ed. G. Davies), *ACM SIGCSE bulletin* **29**, 3 (1997), 24–26.

290. E. Rautama, E. Sutinen and J. Tarhio: Supporting learning process with concept map scripts. *Journal of Interactive Learning Research* **8**, 3/4 (1997), 407-420.

291. E. Sutinen and J. Tarhio: Eliot - what happens inside a program?. In *Proc. Interaktiivinen teknologia koulutuksessa, ITK '97* (eds. V. Kuuliala et al.), Summer University of Hämeenlinna, 1997, 94–98.

292. E. Sutinen, J. Tarhio, S.-P. Lahtinen, A.-P. Tuovinen, E. Rautama and V. Meisalo: Eliot — An algorithm animation environment. Report A-1997-4, Department of Computer Science, University of Helsinki, 1997.

Computers and Society (K.4)

293. P. Kuuppelomäki: Critical computer science groups in the world (in Finnish). *Tietotekniikan maailma* **1** (1996), 34–35.

294. J. Vilo: Estonia's way into a computer society (in Finnish). *Tietotekniikan maailma* **1** (1996), 23–25.

Legal Aspects of Computing — Governmental Issues (K.5.2)

295. H. Ahonen: Communications Decency Act: Network censorship or not? (in Finnish). *Tietotekniikan maailma* **1** (1996), 2–7.

Chapter 3

Faculty

The members of the faculty are introduced by giving their position, e-mail address, WWW home page, research interests, a sample of recent publications, and recent academic activities.

All e-mail addresses are given in the Internet format. In the domain `cs.helsinki.fi` there is a general address format containing the first and the last name of a person separated with a period, for example `hannu.erkio@cs.helsinki.fi` for Hannu Erkiö (note the necessary transliteration é→e, ä→a and ö→o).

The publication numbers refer to the classified listing of publications in Section 2.3.

AHONEN, Helena Ph.D., Research Assistant

• *E-mail:* `Helena.Ahonen@cs.helsinki.fi`

• *Url:* `http://www.cs.helsinki.fi/hahonen/`

• *Interests:* document management, data mining in text, machine learning

• *Publications:* [167–170, 198–202, 247–254, 295]

• *Memberships of program committees:* Knowledge Discovery and Data Mining, KDD '97, Newport Beach, California, USA

ALANKO, Timo Ph.D., Assistant Professor

• *E-mail:* `Timo.Alanko@cs.helsinki.fi`

• *Url:* `http://www.cs.helsinki.fi/~alanko/`

• *Interests:* distributed operating systems, mobile systems, performance analysis

• *Publications:* [6–9, 12, 13, 19]

BACK, Ralph-Johan Ph.D., Docent, Professor (Åbo Akademi)

• *E-mail:* `backrj@abo.fi`

• *Url:* `http://www.abo.fi/~backrj/`

• *Interests:* program construction methods and tools, formal methods, programming language semantics, program verification, multiprocessor technology

ELOMAA, Tapio Ph.D., Assistant Professor

• *E-mail:* `Tapio.Elomaa@cs.helsinki.fi`

• *Url:* `http://www.cs.helsinki.fi/~elomaa/`

• *Interests:* machine learning, robotics, artificial intelligence, algorithms

• *Publications:* [189, 208–215]

• *Memberships of program committees:* World Multiconference on Systemics, Cybernetics and Informatics, SCI '97, Caracas, Venezuela ◦ International Conference on Information Systems Analysis and Synthesis, Orlando, Florida, USA, 1998 ◦ 6th Scandinavian Conference on Artificial Intelligence, SCAI '97, Helsinki, Finland (co-chairman)

• *Other activities:* Referee for Data Mining and Knowledge Discovery, 1997–

ERKIÖ, Hannu Ph.D., Docent, Assistant Professor

• *E-mail:* `Hannu.Erkio@cs.helsinki.fi`

• *Url:* `http://www.cs.helsinki.fi/~erkio/`

• *Interests:* user interfaces, computer-supported cooperative work

• *Publications:* [187, 188]

GRAHNE, Gösta Ph.D., Docent, Assistant Professor (Currently Associate Professor, Concordia University)

• *E-mail:* `grahne@cs.concordia.ca`

• *Url:* `http://www.cs.concordia.ca/~faculty/grahne/`

• *Interests:* artificial intelligence, belief revision, data and knowledge bases, knowledge representation, sequence databases

• *Publications:* [134–138, 190, 191]

• *Memberships of program committees:* 6th Scandinavian Conference of Artificial Intelligence, SCAI '97, Helsinki, Finland, (chair)

GRANÖ, Kari M.Sc., Research Assistant

• *E-mail:* `Kari.Grano@cs.helsinki.fi`

• *Url:* `http://www.cs.helsinki.fi/~grano/`

• *Interests:* distributed computing, programming languages

• *Publications:* [10, 17, 24, 99]

HÄKKINEN, Auvo M.Sc., System Administrator

• *E-mail:* `Auvo.Hakkinen@cs.helsinki.fi`

• *Url:* `http://www.cs.helsinki.fi/~ahakkine/`

• *Interests:* computer systems organization, operating systems, software engineering,

HEIKKINEN, Barbara M.Sc., Research Assistant

• *E-mail:* `Barbara.Hekkinen@cs.helsinki.fi`

• *Url:* `http://www.cs.helsinki.fi/~bheikkin/`

• *Interests:* document management, digital libraries, information retrieval, data mining

• *Publications:* [247–254]

HEINONEN, Oskari M.Sc., Research Assistant

• *E-mail:* `Oskari.Heinonen@cs.helsinki.fi`

• *Url:* `http://www.cs.helsinki.fi/~oheinone/`

• *Interests:* document management, data mining, information retrieval

• *Publications:* [167–170, 244, 247–254]

JÄRVINEN, Pertti Ph.D., Docent, Professor (University of Tampere)

• *E-mail:* `pj@cs.uta.fi`

• *Url:* `http://www.uta.fi/~pj/`

• *Interests:* information systems, research methodologies, social effects of computing systems, management of computing activities

KAIVOLA, Roope Ph.D., Docent, Assistant Professor

• *E-mail:* `Roope.Kaivola@cs.helsinki.fi`

• *Url:* `http://www.cs.helsinki.fi/~rkaivola/`

• *Interests:* concurrency, temporal logic, process algebra, infinitary automata

• *Publications:* [18, 100, 106–111]

• *Other activities:* Referee for *Theoretical Computer Science* ◦ *SIAM Journal on Computing* ◦ 9th International Conference on Computer-Aided Verification, CAV '97, Haifa, Israel ◦ Conference on Computer Science Logic, CSL '97, Aarhus, Denmark ◦ 24th International Colloquium on Automata, Languages and Programming, ICALP '97, Bologna, Italy ◦ 6th Scandinavian Conference on Artificial Intelligence, SCAI '97, Helsinki, Finland ◦ Logic in Computer Science, LICS '96, New Jersey, USA ◦ International Conference on Concurrency Theory, CONCUR '96, Pisa, Italy

KÄRKKÄINEN, Juha M.Sc., Research Assistant

• *E-mail:* `Juha.Karkkainen@cs.helsinki.fi`

• *Url:* `http://www.cs.helsinki.fi/~tpkarkka/`

• *Interests:* algorithms and data structures, string algorithms, data mining, computational biology

• *Publications:* [83, 85–88, 162]

KARVI, Timo Ph.Lic., Assistant Professor (acting)

• *E-mail:* `Timo.Karvi@cs.helsinki.fi`

• *Url:* `http://www.cs.helsinki.fi/~karvi/`

• *Interests:* formal specification, process algebras, distributed computing

KEROLA, Teemu Ph.D., Assistant Professor

• *E-mail:* `Teemu.Kerola@cs.helsinki.fi`

• *Url:* `http://www.cs.helsinki.fi/~kerola/`

• *Interests:* performance evaluation, software metrics, data visualization, operating systems

KILPELÄINEN, Pekka Ph.D., Docent, Assistant Professor

• *E-mail:* `Pekka.Kilpelainen@cs.helsinki.fi`

• *Url:* `http://www.cs.helsinki.fi/~kilpelai/`

• *Interests:* document management, pattern matching, data mining and knowledge discovery

• *Publications:* [71, 89, 117, 133, 139, 247–249, 251, 252, 255]

• *Memberships of program committees:* International Conference on Electronic Publishing, EP '98, Saint Malo, France

• *Other activities:* Preliminary inspector for the Ph.D. thesis of Eila Kuikka (University of Kuopio) in 1996 ◦ Referee for *Acta Informatica* ◦ *Algorithmica* ◦ *Journal of Algorithms* ◦ *Nordic Journal on Computing* ◦ *SIAM Journal on Computing* ◦ Scandinavian Workshop on Algorithm Theory ◦ Combinatorial Pattern Matching ◦ Programming Language Implementation and Logic Programming ◦ ACM SIGIR ◦ Member of the Dictionary Committee of the Finnish Information Processing Association since 1994 ◦ Treasurer of the Finnish Society for Computer Science in 1995–1997

KIVINEN, Jyrki Ph.D., Docent, Assistant Professor

• *E-mail:* `Jyrki.Kivinen@cs.helsinki.fi`

• *Url:* `http://www.cs.helsinki.fi/~jkivinen/`

• *Interests:* machine learning, neural networks

• *Publications:* [218, 220–222]

• *Other activities:* Coordinator of the Graduate School in Computer Science and Engineering (HeCSE) 1998–

KLEMETTINEN, Mika M.Sc., Research Assistant

• *E-mail:* `Mika.Klemettinen@cs.helsinki.fi`

• *Url:* `http://www.cs.helsinki.fi/~mklemett/`

• *Interests:* data mining, text mining, information retrieval, user interfaces

• *Publications:* [20–23, 167–170, 175, 176, 250, 253, 254]

• *Other activities:* Referee for *IEEE Transactions on Knowledge and Data*

Engineering, Special Issue on Data Mining ◦ Technical editor of *Tietojenkäsittelytiede* (Finnish Journal of Computer Science)

KOJO, Markku M.Sc., Assistant Professor

• *E-mail:* `Markku.Kojo@cs.helsinki.fi`

• *Url:* `http://www.cs.helsinki.fi/~kojo/`

• *Interests:* mobile computing, distributed systems, data communication

• *Publications:* [6–9, 11–14, 19, 38]

• *Other activities:* Referee for the 2nd International Mobile Computing Conference, 1996, Hsinchu, Taiwan, R.O.C. ◦ *IEEE Transactions on Computers*, Special Issue on Mobile Computing, 1996

KORPIMIES, Kai M.Sc., Research Assistant, Assistant Professor (acting)

• *E-mail:* `Kai.Korpimies@cs.helsinki.fi`

• *Url:* `http://www.cs.helsinki.fi/~korpimie/`

• *Interests:* information retrieval

• *Publications:* [1, 2, 4, 178, 179]

KUJALA, Teija M.Sc., Amanuensis

• *E-mail:* `Teija.Kujala@cs.helsinki.fi`

• *Url:* `http://www.cs.helsinki.fi/~tkujala/`

• *Interests:* computer-aided software engineering, database design

KUTVONEN, Lea Ph.Lic., Laboratory Engineer, Researcher

• *E-mail:* `Lea.Kutvonen@cs.helsinki.fi`

• *Url:* `http://www.cs.helsinki.fi/~summanen/`

• *Interests:* open distributed processing, heterogeneous system architectures

• *Publications:* [28–37]

• *Memberships of program committees:* 2nd International Working Conference on Distributed Applications and Interoperable Systems, DAIS '99, Helsinki, Finland, 1999

• *Other activities:* Finnish delegate in ISO/IEC JTC1/SC21 WG7 on Open Distributed Processing (ODP) standardisation

KUTVONEN, Petri Chief Systems Analyst

• *E-mail:* `Petri.Kutvonen@cs.helsinki.fi`

• *Url:* `http://www.cs.helsinki.fi/~kutvonen/`

• *Interests:* operating systems, distributed systems, open distributed processing (ODP), trading function

• *Publications:* [37]

KUUPPELOMÄKI, Päivi M.Sc., Research Assistant

• *E-mail:* `Paivi.Kuuppelomaki@cs.helsinki.fi`

• *Url:* `http://www.cs.helsinki.fi/~kuuppelo/`

• *Interests:* modelling of concurrency

• *Publications:* [293]

LAINE, Harri Ph.Lic., Assistant Professor

• *E-mail:* `Harri.Laine@cs.helsinki.fi`

• *Url:* `http://www.cs.helsinki.fi/~laine/`

• *Interests:* computer-aided software engineering (CASE), database design

• *Publications:* [61, 64, 130]

LEMSTRÖM, Kjell M.Sc., Research Assistant

• *E-mail:* `Kjell.Lemstrom@cs.helsinki.fi`

• *Url:* `http://www.cs.helsinki.fi/~klemstro/`

• *Interests:* computer graphics, digital image processing, information retrieval, pattern recognition (pattern matching)

• *Publications:* [180–182, 243]

LINDÉN, Greger Ph.D., Assistant Professor (acting)

• *E-mail:* `Greger.Linden@cs.helsinki.fi`

• *Url:* `http://www.cs.helsinki.fi/~linden/`

• *Interests:* document management, structured documents, structured transformations, SGML

• *Publications:* [71, 72, 247–249, 256]

• *Other activities:* Member of the board of the SGML Users' Group Finland 1996– ○ Publicity manager of SCAI '97 ○ Secretary of the Publishing Association Nordic Journal of Computing 1998–

LINNAINMAA, Seppo Ph.D., Docent, Research Professor (VTT Information Technology)

• *E-mail:* `Seppo.Linnainmaa@vtt.fi`

• *Url:* `http://www.vtt.fi/tte/staff/lis/`

• *Interests:* knowledge-based systems, object-oriented software, intelligent operations management

LOKKI, Heikki Ph.Lic., Assistant Professor

• *E-mail:* `Heikki.Lokki@cs.helsinki.fi`

• *Url:* `http://www.cs.helsinki.fi/~lokki/`

• *Interests:* computer science in biological sciences, analytical methods for ringing recoveries

• *Publications:* [263]

• *Other activities:* Referee of EURING '97 ○ Coordinator of the Graduate School in Computational Biology, Bioinformatics and Biometry (ComBi) 1998–

MÄKELÄ, Matti Dr.Techn., Associate Professor

• *E-mail:* `Matti.Makela@cs.helsinki.fi`

• *Url:* `http://www.cs.helsinki.fi/~mamakela/`

• *Interests:* numerical methods, mathematical software, computers in education, computer graphics, scientific visualization

• *Publications:* [269–271, 275, 276]

• *Other activities:* International coordinator and referee of ACM SIGNUM Newsletter

MANNILA, Heikki Ph.D., Professor

• *E-mail:* Heikki.Mannila@cs.helsinki.fi

• *Url:* http://www.cs.helsinki.fi/~mannila/

• *Interests:* data mining, databases, machine learning, algorithms

• *Publications:* [20–23, 58–60, 91, 144, 150, 151, 161, 163, 164, 171–177, 183–185, 202, 219, 234, 235, 238, 240–242, 244–246, 248, 249, 257, 258, 263]

• *Memberships of program committees:* 13th European Meeting on Cybernetics and Systems Research, EMCSR '96, Vienna, Austria ◦ 15th ACM SIGACT-SIGART-SIGMOD Symposium on Principles of Database Systems, PODS '96, Montréal, Canada ◦ 3rd International Workshop on Principles of Document Processing, PODP '96, Palo Alto, California, USA ◦ 22nd International Conference on Very Large Databases, VLDB '96, Mumbai, India ◦ ACM SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery, SIGMOD '96 DMKD, Montréal, Canada, ◦ 1st European Symposium on Principles of Data Mining and Knowledge Discovery, PKDD '97, Trondheim, Norway ◦ 3rd International Conference on Knowledge Discovery and Data Mining, KDD '97, Newport Beach, California, USA (co-chair) ◦ 2nd ACM SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery, SIGMOD '97 DMKD, Tucson, Arizona, USA ◦ European Conference on Machine Learning, ECML '97, Prague, Czech Republic ◦ 9th International Conference on Scientific and Statistical Database Management Systems, SSDBM '97, Olympia, Washington, USA ◦ 2nd Pacific-Asia Conference on Knowledge Discovery and Data Mining, PAKDD '98, Melbourne, Australia ◦ 7th ACM SIGACT-SIGART-SIGMOD Symposium on Principles of Database Systems, PODS '98, Seattle, Washington, USA ◦ 14th International Conference on Data Engineering, ICDE '98, Orlando, Florida, USA ◦ 4th International Workshop on Principles of Digital Document Processing, PODDP '98, Saint Malo, France ◦ 2nd European Symposium on Principles of Data Mining and Knowledge Discovery, PKDD '98, Nantes, France ◦ 15th National Conference on Artificial Intelligence, AAAI '98, Madison, Wisconsin USA ◦ 4th International Conference on Knowledge Discovery and Data Mining, KDD '98, New York, USA

• *Other activities:* Chairman of the board of the Rolf Nevanlinna Institute, 1995– ◦ Reviewer for ESPRIT project ILP2, 1997– ◦ Editor-in-chief of *Data Mining and Knowledge Discovery*, 1997– ◦ Member of the KDD Steering Committee, 1996– ◦ External reviewer for professorships at the University of Jyväskylä, Rochester University, Uppsala University, Bar-Ilan

University and University of Freiburg.

MARTTINEN, Liisa M.Sc., Assistant Professor (acting)

• *E-mail:* `Liisa.Marttinen@cs.helsinki.fi`

• *Url:* `http://www.cs.helsinki.fi/~marttine/`

• *Interests:* computer communication, computer networks, network management, open distributed processing

• *Publications:* [28–30]

MYLLYMÄKI, Petri Ph.D., Assistant Professor

• *E-mail:* `Petri.Myllymaki@cs.helsinki.fi`

• *Url:* `http://www.cs.helsinki.fi/~myllymak/`

• *Interests:* probabilistic modeling, uncertain reasoning, machine learning, case-based reasoning, neural computing

• *Publications:* [118, 120–124, 192, 193, 217, 223–233, 236, 237]

• *Memberships of program committees:* 11th Annual Florida Artificial Intelligence International Research Symposium, FLAIRS '98, special track on uncertain reasoning, Florida, USA ◦ 2nd Nordic Workshop on Genetic Algorithms and Their Applications, NWGA '96, Vaasa, Finland

• *Other activities:* Referee for *IEEE Transactions on Pattern Analysis and Machine Intelligence* ◦ *Knowledge Engineering Review* ◦ *Nordic Journal of Computing* ◦ 13th European Conference on Artificial Intelligence, ECAI '98

NIKLANDER, Tiina M.Sc., Amanuensis

• *E-mail:* `Tiina.Niklander@cs.helsinki.fi`

• *Url:* `http://www.cs.helsinki.fi/~niklande/`

• *Interests:* real-time databases, fault-tolerance

• *Publications:* [40–44, 46, 65, 131, 145–147]

NURMI, Otto Dr.rer.pol., Assistant Professor

• *E-mail:* `Otto.Nurmi@cs.helsinki.fi`

• *Interests:* databases, algorithms, data structures, computational geometry

• *Publications:* [76, 77]

NYKÄNEN, Matti Ph.D., Assistant Professor (acting)

• *E-mail:* `Matti.Nykanen@cs.helsinki.fi`

• *Url:* `http://www.cs.helsinki.fi/~mnykanen/`

• *Interests:* query languages for sequence databases, bioinformatics, theory of computation

• *Publications:* [136, 137, 140, 141]

• *Other activities:* Treasurer of the 6th Scandinavian Conference on Artificial Intelligence, SCAI '97 ○ Treasurer of the Finnish Society of Computer Science 1998

ORPONEN, Pekka Ph.D., Docent, Professor (University of Jyväskylä)

• *E-mail:* `orponen@math.jyu.fi`

• *Url:* `http://www.math.jyu.fi/~orponen/`

• *Interests:* computational complexity, neural networks, genetic algorithms, cellular automata

PAAKKI, Jukka Ph.D., Associate Professor

• *E-mail:* `Jukka.Paakki@cs.helsinki.fi`

• *Url:* `http://www.cs.helsinki.fi/~paakki/`

• *Interests:* software engineering, programming languages: design and implementation

• *Publications:* [10, 17, 24, 66, 67, 70, 99, 160]

• *Memberships of program committees:* 7th Nordic Workshop on Programming Environment Research, NWPER '96, Aalborg, Denmark ○ 3rd International Workshop on Automated and Algorithmic Debugging, AADEBUG '97, Linköping, Sweden ○ 9th International Symposium on Programming Languages: Implementations, Logics, and Programs, PLILP '97, Southampton, UK ○ 5th Symposium on Programming Languages and Software Tools, Jyväskylä, Finland, 1997 (chair) ○ 10th International Symposium on Programming Languages: Implementations, Logics, and Programs, PLILP '98, Pisa, Italy

• *Other activities:* Editor of *Tietojenkäsittelytiede* (Finnish Journal of Computer Science) ○ Columnist of *Tietoyhteys* (Magazine by the Centre for

Scientific Computing) ◦ *Tietoviikko* (Weekly Journal of the Finnish Information Processing Association) ◦ *Taltio* (Magazine of Novo Group Ltd.) ◦ Referee for 8th International Symposium on Programming Languages: Implementations, Logics, and Programs, PLILP '96, Aachen, Germany, 1996 ◦ 11th International Symposium on Computer and Information Systems, ISCIS-XI, Antalya, Turkey, 1996 ◦ External reviewer for the Academy of Finland from 1995

POLLARI-MALMI, Kerttu Techn.Lic., Assistant Professor (acting)

• *E-mail:* Kerttu.Pollari-Malmi@cs.helsinki.fi

• *Url:* http://www.cs.helsinki.fi/~pollari/

• *Interests:* databases, concurrency control, B-trees

• *Publications:* [78]

PUUSTJÄRVI, Juha Ph.Lic., Research Assistant

• *E-mail:* Juha.Puustjarvi@cs.helsinki.fi

• *Url:* http://www.cs.helsinki.fi/~puustjar/

• *Interests:* concurrency control, multidatabases, workflows

• *Publications:* [152–155, 158]

RAATIKAINEN, Kimmo Ph.D., Associate Professor

• *E-mail:* Kimmo.Raatikainen@cs.helsinki.fi

• *Url:* http://www.cs.helsinki.fi/~kraatika/

• *Interests:* distributed systems, mobile computing, telecommunications, intelligent networks, performance evaluation, real-time databases

• *Publications:* [6–9, 11–16, 19, 38–50, 65, 131, 142, 143, 145–148, 156]

• *Memberships of program committees:* IFIP 1996 Conference on Intelligent Networks, IN '96, Melbourne, Australia

• *Other activities:* Member of IFIP TC6 Special Interest Group of Intelligent Networks ◦ Editor of Tietojenkäsittelytiede (Finnish Journal of Computer Science)

RÄIHÄ, Kari-Jouko Ph.D., Docent, Professor (University of Tampere)

• *E-mail:* kjr@cs.uta.fi

• *Url:* `http://www.cs.uta.fi/~kjr/`

• *Interests:* human-computer interaction

RONKAINEN, Pirjo M.Sc., Research Assistant

• *E-mail:* `Pirjo.Ronkainen@cs.helsinki.fi`

• *Url:* `http://www.cs.helsinki.fi/~ronkaine/`

• *Interests:* data mining, databases, knowledge bases, hypertext

• *Publications:* [20, 91, 172, 175, 176]

SALMENKIVI, Marko M.Sc., M.Th., Research Assistant

• *E-mail:* `Marko.Salmenkivi@cs.helsinki.fi`

• *Url:* `http://www.cs.helsinki.fi/~salmenki/`

• *Interests:* data mining, computational statistics

• *Publications:* [245, 246, 258]

SILANDER, Tomi M.Sc., Research Assistant

• *E-mail:* `Tomi.Silander@cs.helsinki.fi`

• *Url:* `http://www.cs.helsinki.fi/~tsilande/`

• *Interests:* probabilistic modeling

• *Publications:* [118, 120–124, 126–128, 217, 223–228]

SIPPU, Seppo Ph.D., Associate Professor

• *E-mail:* `Seppo.Sippu@cs.helsinki.fi`

• *Url:* `http://www.cs.helsinki.fi/~sippu/`

• *Interests:* database structures, query processing, logic databases, spatial databases, parsing

• *Publications:* [157]

SUTINEN, Erkki Ph.Lic., Research Assistant

• *E-mail:* `Erkki.Sutinen@cs.helsinki.fi`

• *Url:* `http://www.cs.helsinki.fi/~sutinen/`

• *Interests:* algorithms and data structures, string algorithms, algorithm animation, computer-aided learning

• *Publications:* [69, 86, 90, 92, 93, 266–268, 272, 273, 277–292]

TAINA, Juha Ph.Lic, Research Assistant

• *E-mail:* `Juha.Taina@cs.helsinki.fi`

• *Url:* `http://www.cs.helsinki.fi/~taina/`

• *Interests:* real-time databases, object orientation, intelligent networks, main memory databases, database models

• *Publications:* [5, 46–54]

TARHIO, Jorma Ph.D., Docent, Associate Professor (University of Joensuu)

• *E-mail:* `Jorma.Tarhio@cs.joensuu.fi`

• *Url:* `http://www.cs.joensuu.fi/~tarhio/`

• *Interests:* computer science education, string algorithms, animation of algorithms,

TIENARI, Martti Ph.D., Professor, Chairman of the Department

• *E-mail:* `Martti.Tienari@cs.helsinki.fi`

• *Url:* `http://www.cs.helsinki.fi/~tienari/`

• *Interests:* modeling concurrency, computer networks and distributed systems, programming languages and compilers, management of computing

• *Publications:* [3, 7, 8, 98, 103]

• *Memberships of program committees:* IFIP WG 6.1 International Conference on Distributed Applications and Interoperable Systems, DAIS '97, Cottbus, Germany

• *Other activities:* Activities in IFIP (International Federation of Information Processing): General Assembly 1987–96, IFIP Trustee 1989–95, Cognizant Officer of TC2 (Software: Theory and Practice) and TC12 (Artificial Intelligence), Member of Publication Committee, Chairman of the Activity Management Board 1992–95, Chairman of the Technical Assembly 1996 ◦ Member of the Finnish Academy of Science and Letters ◦ Member of the Finnish Academy of Technical Sciences

TIRRI, Henry Ph.D., Docent, Senior Research Scientist

• *E-mail:* `Henry.Tirri@cs.helsinki.fi`

• *Url:* `http://www.cs.helsinki.fi/~tirri/`

• *Interests:* probabilistic modeling, uncertain reasoning, foundations of machine learning, case-based reasoning, neural computing, transaction processing, nonlinear modeling in education

• *Publications:* [72, 118, 120–128, 149, 153–155, 158, 192, 193, 217, 223–233, 236, 237]

• *Memberships of program committees:* 3rd European Conference on Case-Based Reasoning, EWCBR '96, Lausanne, Switzerland, 1996 ◦ Workshop on Case-Based Learning: Beyond Classification and Feature Vectors, Prague, Czech Republic, 1997 ◦ 4th European Conference on Case-Based Reasoning, EWCBR '98, Dublin, Ireland, 1998

• *Other activities:* Editor of *The Computer Journal*, 1995– ◦ Referee for *IEEE Transactions on Information Theory* ◦ 15th International Joint Conference on Artificial Intelligence, IJCAI '97 ◦ 6th Scandinavian Conference on Artificial Intelligence, SCAI '97, Helsinki, Finland ◦ 3rd European Conference on Case-Based Reasoning, EWCBR '96, Lausanne, Switzerland, 1996 ◦ Workshop on Case-Based Learning: Beyond Classification and Feature Vectors, Prague, Czech Republic, 1997 ◦ 4th European Conference on Case-Based Reasoning, EWCBR '98, Dublin, Ireland, 1998 ◦ Preliminary inspector of the Ph.D. Theses of Seppo Kuusisto, Tampere University of Technology, 1998 and Petteri Pajunen, Helsinki University of Technology, 1998 ◦ Associate node coordinator of the European Community ESPRIT III Network of Excellence in Neural Networks (NEURONET) ◦ Official opponent for PhD defense of Seppo Kuusisto, Tampere University of Technology, 1998

TOIVONEN, Hannu Ph.D., Assistant Professor

• *E-mail:* `Hannu.Toivonen@cs.helsinki.fi`

• *Url:* `http://www.cs.helsinki.fi/~htoivone/`

• *Interests:* data mining, data analysis, machine learning, artificial intelligence, stochastic simulation

• *Publications:* [20–23, 161–166, 175, 176, 186, 207, 235, 240–242, 245, 246, 257, 264, 265]

• *Memberships of program committees:* 3rd International Conference on

Knowledge Discovery and Data Mining, KDD '97, Newport Beach, California, 1997 ∘ 4th International Conference on Knowledge Discovery and Data Mining, KDD '98, New York, 1998

• *Other activities:* Referee for *Data Mining and Knowledge Discovery* ∘ *Decision Support Systems*, Special Issue on Knowledge Discovery and Its Applications to Business Decision Making, 1996

TUOVINEN, Antti-Pekka M.Sc., Research Assistant

• *E-mail:* `Antti-Pekka.Tuovinen@cs.helsinki.fi`

• *Url:* `http://www.cs.helsinki.fi/~aptuovin/`

• *Interests:* visual languages, object-oriented programming, language engineering, software engineering

• *Publications:* [69, 73–75, 159, 160, 268, 277, 278, 280, 292]

UKKONEN, Esko Ph.D., Professor

• *E-mail:* `Esko.Ukkonen@cs.helsinki.fi`

• *Url:* `http://www.cs.helsinki.fi/~ukkonen/`

• *Interests:* algorithms and data structures, string algorithms, machine learning, computational biology

• *Publications:* [84, 87, 88, 114–116, 137, 179, 203–206, 259–262]

• *Memberships of program committees:* 24th International Colloquium on Automata, Languages, and Programming, ICALP '97, Bologna, Italy ∘ 8th Annual Symposium on Combinatorial Pattern Matching, CPM '97, Aarhus, Denmark ∘ 4th South American Workshop on String Processing, WSP '97, Valparaíso, Chile ∘ 1st International Conference on Discovery Science, DS '98, Fukuoka, Japan ∘ 9th International Workshop on Algorithmic Learning Theory, ALT '98, Otzenhausen, Germany

• *Other activities:* Editor-in-Chief of the *Nordic Journal of Computing*, from 1994 ∘ Member of the Steering Committee of the Scandinavian Workshop on Algorithm Theory (SWAT), from 1991 ∘ External reviewer for professorships at University of California at San Diego, University of California at Davis, University of Colorado at Boulder, the University of Utah, 1996, University of Tampere 1997, and University of Kuopio 1998 ∘ Director of the Graduate School in Computational Biology, Bioinformatics, and Biometry (ComBi), 1998– ∘ Biotechnology Prize of 1996 awarded by the Alko-Group Ltd

VALMARI, Antti Dr.Techn., Docent, Associate Professor (Tampere University of Technology)

• *E-mail:* `Antti.Valmari@tut.fi`

• *Url:* `http://www.cs.tut.fi/~ava/`

• *Interests:* computer-aided verification and validation of concurrent and distributed systems

VEIJALAINEN, Jari Dr.-Ing., Docent, Professor, (University of Jyväskylä)

• *E-mail:* `veijalai@jytko.jyu.fi`

• *Url:* `http://www.cs.jyu.fi/~veijalai/`

• *Interests:* heterogeneous autonomous distributed systems, heterogeneous transaction processing, transactional workflow systems, mobile computing, verification and validation of concurrent and distributed systems, multimedia databases

VERKAMO, A. Inkeri Ph.D., Docent, Assistant Professor

• *E-mail:* `Inkeri.Verkamo@cs.helsinki.fi`

• *Url:* `http://www.cs.helsinki.fi/~verkamo/`

• *Interests:* software performance evaluation, software engineering, data mining

• *Publications:* [72, 161, 164, 167–170]

• *Memberships of program committees:* 8th International Conference on Software Engineering and Knowledge Engineering, SEKE '96, Lake Tahoe, Nevada

• *Other activities:* Coordinator of the Graduate School in Computer Science and Engineering (HeCSE) 1995–7

VIHAVAINEN, Juha Ph.Lic., Assistant Professor

• *E-mail:* `Juha.Vihavainen@cs.helsinki.fi`

• *Url:* `http://www.cs.helsinki.fi/~vihavain/`

• *Interests:* principles and implementation of programming languages, language processors, programming paradigms, multi-paradigm programming, object-oriented programming

• *Other activities:* Referee for ECOOP 1991-95 and *Theory and Practice of Object Systems* 1997.

VILO, Jaak M.Sc. (University of Tartu), Research Assistant

• *E-mail:* `Jaak.Vilo@cs.helsinki.fi`

• *Url:* `http://www.cs.helsinki.fi/~vilo/`

• *Interests:* computational biology, pattern discovery, machine learning

• *Publications:* [203–206, 294]

WIKLA, Arto M.Sc., Assistant Professor (acting)

• *E-mail:* `arto.wikla@cs.helsinki.fi`

• *Url:* `http://www.cs.helsinki.fi/~wikla/`

• *Interests:* computer science education, WWW, programming languages, Java

YLI-HARJA, Olli Dr.Techn., Assistant Professor (acting)

• *E-mail:* `Olli-Yli-Harja@cs.helsinki.fi`

• *Url:* `http://www.cs.helsinki.fi/~yliharja/`

• *Interests:* image processing, nonlinear signal processing, median filters, algorithm analysis, computer-aided sound and music analysis, data structures

• *Publications:* [119, 129]

Chapter 4

Education

4.1 Educational Program

The students of the department normally start their university studies at the age of 19. Their goal is to receive a B.Sc. (Bachelor of Science) or M.Sc. (Master of Science) degree in computer science requiring three to four or five years of full time study. Beyond the first degree there are two alternative graduate degrees: Ph.Lic. (Licentiate of Philosophy) and Ph.D. (Doctor of Philosophy). The academic year has two semesters: the fall semester lasts from 1 September to 20 December (classes from 11 September to 10 December), while the spring semester lasts from 16 January to 31 May (classes from 16 January to 10 May, excluding one week of Easter vacation). It is also possible to study in summer. Intensive courses of 4–5 weeks covering introductory topics are offered in June and August. Graduate courses are also organized in cooperation with other Finnish universities during the summer. These courses typically last for one week and are intended for Ph.Lic. and Ph.D. students. These are often given in English by foreign visitors.

In order to obtain a B.Sc. degree a student must earn 120 units of academic credit. For a M.Sc. degree, 160 units of credit as well as a thesis is required. One credit should normally correspond to roughly one week (40 hours) of study. Our students typically register for 12 credits ("study weeks") in the fall semester and 15 credits in the spring semester. During the summer sessions a student can earn an additional 8–10 credits. Most students, however, work in industry during the summer to gain practical experience in data processing and to improve their financial situation. This is actually what the department recommends. Thus, a normal student should earn 27 credits a year, an exceptionally diligent full-year student 40 credits. Nevertheless, there is a considerable variation in study efficiency among students.

Our typical course consists of about 50 to 60 lectures (a lecture lasts 45 minutes) and of about 20 to 30 hours of problem solving, discussion and repetition sessions in small groups of about 10 to 20 students. Each course is examined individually with grades: 3/3 = excellent, 2/3 = good, 1/3 = satisfactory. A typical course is worth 4 or 5 credits. The computer laboratory is supervised in small groups of 6 to 12 students. Students also attend seminar courses, the enrollment of which ranges from 5 to 15 students. In these seminars the students read current literature, write essays and give oral presentations. A seminar group normally meets 2 hours per week yielding 2 credits per semester.

In order to receive a M.Sc. degree in computer science, students are required to earn their credits as follows:

Computer science	$\geq$ 95 cr
Mathematics	$\geq$ 26 cr
Physics	$\geq$ 15 cr
General studies	9 cr
Total	$\geq$160 cr

In mathematics, the obligatory courses are calculus (11 cr), algebra (5 cr), logic (5 cr), and probability (5 cr). Physics can be replaced with almost any other subject, such as economics, administration, statistics, or psychology. For a B.Sc. degree, 55 credit units of computer science is sufficient.

The computer science studies for a M.Sc. degree can be subdivided as follows:

Obligatory courses and laboratories	35 cr
Elective courses	$\geq$ 28 cr
Seminars	$\geq$ 4 cr
Project work	8 cr
M.Sc. Thesis, Scientific writing	20 cr
Total	$\geq$ 95 cr

The obligatory computer science courses and laboratory work currently cover (academic year 1997–98) the following areas:

Introductory programming (Java)	6 cr
Data structures	6 cr
Operating systems and hardware architecture	7 cr
Information systems and databases	8 cr
Theory of computation	4 cr
Total	31 cr

In principle, students are fairly free to choose any elective courses. They normally follow the recommendation of the department by building up a specialized background knowledge for a successful thesis in one of our research groups. Thus, a student might specialize according to his/her study goals, interests and talents, towards, e.g., theoretical computer science, information systems, telecommunications software, distributed systems, operating systems, artificial intelligence, or software engineering.

To start studies for the postgraduate degrees Ph.Lic. (Licentiate of Philosophy) and Ph.D. (Doctor of Philosophy) in Computer Science, a student having shown good academic standing in his/her M.Sc. studies contacts a professor of the department. At first, a personal study program is designed for the student. It outlines the field of specialization of the studies, the topic for the thesis, and the content and the schedule of the coursework. Each student is assigned a personal advisor. See Section 4.2 (Graduate Schools) for more details.

The requirements for the Ph.Lic. degree can be summarized as follows:

Elective courses and seminars	
in computer science	20 cr
in mathematics	20 cr
Ph.Lic. thesis	50 cr
Total	90 cr

The elective courses in mathematics can be replaced with coursework in other subjects such as physics, economy, psychology, etc., or additional courses in computer science. The Ph.Lic. thesis is normally written in English. The allocation of credits for thesis research indicates that after the required coursework it should take 1–2 years to prepare a Ph.Lic. thesis.

It is important that the student takes the courses and the seminars early enough to obtain sufficient background for writing the thesis. Active participation in seminar courses is particularly useful as is attending international schools and specialized research courses. Such courses are also regularly given at the department.

The requirements for the Ph.D. degree are otherwise the same as for the Ph.Lic. degree, but a Ph.D. thesis demands more work, from 2 to 3 years, roughly one year more than a Ph.Lic. thesis. The Ph.D. degree can be achieved directly, although we often recommend that our students take the Ph.Lic. degree first, and then by improving and extending their Ph.Lic. research, achieve the Ph.D. level.

The Ph.D. theses are written in English. A thesis should include a scientific contribution which is significant enough to be publishable internationally. A Ph.D. thesis (as well as a Ph.Lic. thesis) can also be assembled from a number of published articles or congress papers, possibly written jointly with other authors. A dissertation of this type, which is actually fairly common, consists of an introductory survey written by the candidate alone, with the individual articles as appendices.

Preparing the thesis is clearly the most demanding part of the Ph.D. and Ph.Lic. studies. To succeed with the thesis it is recommendable that a student works within a research group at the department. The support and the criticism given by the group is often essential for making progress in the work.

4.2 Graduate Schools

The department is at the moment involved in two separate graduate schools, the Helsinki Graduate School in Computer Science and Engineering (HeCSE) and the Graduate School in Computational Biology, Bioinformatics, and Biometry (ComBi).

4.2.1 HeCSE

The Helsinki Graduate School in Computer Science and Engineering (HeCSE) is a postgraduate program in computer science and computer engineering jointly offered by the Helsinki University of Technology and the University of Helsinki. In addition to the Department of Computer Science at the University of Helsinki, the following laboratories of the University of Technology participate in HeCSE:

- Laboratory of Information Processing Science (Prof. Martti Mäntylä),
- Laboratory of Telecommunications Software and Multimedia (Prof. Arto Karila),
- Digital Systems Laboratory (Prof. Leo Ojala),

- Laboratory of Signal Processing and Computer Technology (Prof. Iiro Hartimo),
- Laboratory of Computer and Information Science (Prof. Erkki Oja), and
- Neural Networks Research Centre (Prof. Teuvo Kohonen).

The main areas of HeCSE are:

- *Software Systems* (information technology for production, multimedia, database systems, data structures),
- *Software Engineering* (embedded systems, specification methods),
- *Telecommunication Software and Distributed Systems* (formal methods, data communication software, signal processing), and
- *Learning and Intelligent Systems* (pattern recognition, neural networks, machine learning, knowledge based systems).

HeCSE provides a program aiming at a doctoral degree in four years on the basis of a master's degree in computer science or a related field. The student is appointed a supervising professor, with whom a study plan is made. The study plan contains an individual timetable. The student participates in the work of a research group. The progress of the student is followed by his/her supervisor and at seminars. Additionally, the student must report on his/her work twice a year. The student is formally enrolled at the university of his/her supervising professor.

The studies consist of courses and seminars (20–25%) and research work ending in a thesis (75–80%). There is a compulsory course on Writing Scientific English and five courses on the main areas of HeCSE, three of which are recommended to be taken by the student. The other courses are chosen individually. The student takes the courses and most of the seminars during the first two years. At the end of the second year, the student presents a thesis proposal to the board of HeCSE. The doctoral thesis is expected to be finished by the end of the fourth year. The studies end with a public doctoral dissertation.

HeCSE has in all 25 graduate assistantships, providing funding for about half of the students of the graduate school. Other sources of funding are employment in research projects as research assistants or in the participating laboratories as part time teachers. During the first three years of its existence (1995–1997), HeCSE has produced 14 doctors (7 of these at the University of Helsinki) and 13 licentiates (3 of these at the University of Helsinki), with the number steadily growing by the year.

The Finnish graduate schools in computer science (Helsinki, Jyväskylä Turku, Tampere and Eastern Finland) co-operate. Joint courses are arranged and students of one graduate school can participate in the courses arranged by another graduate school. Also participation in some courses arranged by industry is possible.

4.2.2 ComBi

The Graduate School in Computational Biology, Bioinformatics, and Biometry (ComBi) is a new postgraduate program jointly offered by the University of Helsinki, the University of Turku and the Center for Scientific Computing of Finland.

The research goal of ComBi is to develop computational, mathematical, and statistical methods and models for biological sciences. To that end, ComBi will educate Ph.D.'s with high-quality methodological expertise. In their thesis work the students are expected to apply this expertise to computational, data analysis, or modeling problems in biology or in some related field. The thesis projects are carried out in close cooperation with one or more research groups in the application area.

Computational biology is a new field of research which develops models and software implementations for computational problems in molecular biology, biotechnology, and genetics. Such means are needed both in the basic research and in the industrial applications of biotechnology. Computer-aided DNA sequencing, sequence comparison, prediction of protein structures, docking of molecules, and the interpretation of electron microscopy or NMR data are examples of typical computationally intensive tasks. Bioinformatics refers to the development and use of (molecular) biological databases.

Biometry investigates statistical modeling problems related for example to genetic mapping (linkage analysis), to the genetic and environmental risk factors of complex diseases, or to the spreading of infectious diseases. Further topics include population dynamics and numerical taxonomy with applications to ecology, microbiology, and genetics. Management of spatial data (GIS) is also often needed.

ComBi incorporates the following departments and institutions that provide the methodological education of the school:

- Department of Computer Science, University of Helsinki

  - Prof. Heikki Mannila (data mining, databases, computational data-analysis)
  - Prof. Esko Ukkonen (computational biology, bioinformatics, pattern matching, machine learning),

- Rolf Nevanlinna Institute, University of Helsinki
  - Prof. Elja Arjas (statistics, especially biometry),

- Department of Mathematical Sciences, University of Turku
  - Prof. Mats Gyllenberg (population dynamics, numerical taxonomy), and

- Center for Scientific Computing
  - Prof. Olle Teleman (molecular modeling, structural biology, scientific computing).

The application oriented part of the thesis work of the students of ComBi will be done in cooperation with suitable research groups working at the universities mentioned above or at other institutions.

The school is coordinated by the Department of Computer Science of the University of Helsinki. Professor Esko Ukkonen is the Director of ComBi.

ComBi provides a program aiming at a doctoral degree in four years on the basis of a master's degree. The master's degree should preferably be in computer science, (applied) mathematics, or statistics. However, a degree in natural sciences (e.g., biological sciences, biochemistry, chemistry, physics, biotechnology) or in engineering is possible, provided that the student has sufficient knowledge of computational and mathematical methods.

The student is appointed a supervising professor, with whom a personal study plan is made. A central principle of the school is that a major part of the thesis work is carried out by participating in the research work in some biological application area. The student is formally enrolled at the university of his/her supervising professor.

The studies consist of courses and seminars (about 20 %) and research work leading to a Ph.D. thesis. The student should take at least three of the five general courses organized by the school. The rest of the courses should be selected individually according to the special field of the thesis project. Most of the courses and seminars are arranged during the normal academic semesters (January to May, September to December), but some intensive courses are also planned during summer and winter breaks. Moreover, the school aims at organizing and financing working periods abroad for its students.

The progress of the student is followed by his/her supervisor and at the joint seminars arranged by the school. The student must report on his/her progress twice a year. The student should present a thesis proposal by the end of the second year, and the thesis is expected to be finished by the end of the fourth year. The studies end with a public defense of the thesis.

4.3 Course Descriptions

Undergraduate Courses 1996–98

Introduction to Computing (2 cu) Introduction to computers and data processing. Algorithms. Computer hardware. Operating systems. Applications software. Programming languages. Database systems. Communication networks. System analysis and design.

Introduction to Programming (4 cu) The course provides the student with the basic principles of programming: algorithms, programming techniques, and object-orientation. Java is the language used, but all of its features are not covered.

Computer Systems Organization (3 cu) Introduction. Data presentation, error detection and correction. Computer organization. Conventional machine level. Assembly language. Compilation, linking, loading. Input/Output. Secondary storage. Operating system. Data communication equipment and software.

Information Systems (4 cu) Principles of relational databases, SQL, database programming, application development, databases and the WWW, user interfaces, information system development, object-oriented analysis, use case model, introduction to database design.

Programming Project (2 cu) The student designs, documents and programs a complete, realistic program. In the course of the development she/he also gives small lectures and demonstrations about the project.

Data Structures (4 cu) Basic data structures. Applications to algorithms. Analysis of algorithms. Implementations of data structures and algorithms. Memory management.

Concurrent Systems (4 cu) Structure and implementation of concurrent and distributed systems. The main emphasis is on solving problems in operating systems.

Database Systems I (4 cu) Databases and database management systems. Relational databases, relational algebra and calculus. File and index structures. Query processing. Transaction processing. Relational database design, functional dependencies and normalization. Object-oriented databases.

Theory of Computation (4 cu) Finite automata and regular languages. Context-free grammars and languages. Rudiments of parsing theory and attribute grammars. Context-sensitive and type-0 grammars. Turing machines. Recursive and recursively enumerable sets. Computability and computational complexity.

Data Structures Project (2 cu) A simulator or some other fairly large program is designed, programmed, tested and documented.

Information Systems Project (2 cu) A small ADP-system is designed and programmed.

Computing Methodologies (3 cu) A characterization of computer science, its problems, methods and applications by selected examples and discussions. Models as tools. A learning model. Human and scientific information processing: Perspectives on data, information, coding and computing, perception and thinking. Algorithms, why and how. L-systems as models. Iterated function systems. Discrete numerical computing. Pictures and graphics. Individual work.

Languages for Artificial Intelligence (3 cu) Fundamentals of declarative and symbolic programming. LISP and PROLOG basics. Hands-on experimentations and implementations.

Computer Graphics (4 cu) Overview of graphics systems. Output primitives and their attributes. Two-dimensional transformations. Windowing and clipping. Segments. Interactive input methods. Three-dimensional concepts, representations, transformations, viewing. Hidden-surface and hidden-line removal. Shading and color models. Modeling methods. Design of the user interface. Individual practical work.

Computer Uses in Education (4 cu) Fundamentals of computer applications in education. The computer as a tutor, tool, and tutee. Computer assisted instruction (CAI) systems. Courseware design, development, and evaluation. Authoring systems and languages. Multimedia CAI. Intelligent CAI. Applications and research. Practical courseware designing in small groups.

Semantics of Programs (3 cu) Axiomatic semantics of programs. Weakest precondition calculus for the guarded command language of Dijkstra. Development of small programs based on the programming logic.

Data Communications (4 cu) The electrical interface. Data transmission. Data link control protocols. Local area networks. High-speed and bridged local area networks. Wide area networks. Internetworking. Transport protocols.

Programming in C (2 cu) Language definition. Programming tools. General programming principles.

Programming Techniques (C++) (3 cu) Abstract data types, class libraries, generic programming, object-oriented programming.

Programming Techniques (Ada) (3 cu) Basic Ada data and control structures, packages, concurrency and object-oriented programming. Applicability of Ada in the implementation of a real-time system.

Software Engineering (4 cu) Introduction to software engineering as a professional discipline. Models of software engineering. Team work. Project planning and organization. Requirements analysis and engineering. Software design. Implementation techniques. Testing. Debugging and maintenance. Software configuration management. Software quality assurance.

Software Engineering Laboratory (6 cu) Each student takes part in a project where a group of students analyzes the requirements of a software product, designs, implements, and tests the product, using systematic software engineering methods and tools. The group assignment may also be focused on some sub-phase of the software life-cycle, such as evolution or maintenance of an existing software system.

UNIX Principles (1 cu) Principles of the UNIX environment for end users, includes principles of file system, shell, wildcards, protection, I/O, text editing, regular expressions, sorting and searching, awk, program development, project maintenance, and networking commands.

UNIX Platform (3 cu) The programming interface to the UNIX system: system calls and library functions for process control, memory management, file systems and peripherals, tools for interprocess communication.

Network Programming (4 cu) Internet networking services provided to application programs and TCP/IP protocols, including design and implementation principles of these protocols. The focus is on practical aspects of designing and implementing distributed applications using these protocols.

Wireless Data Communication Networks (1 cu) Basics of wireless data communication. GSM, wireless LANs, cellular telephone networks, packet radio networks, wireless LANs, satellites and UMTS.

Digital Signal Processing (3 cu) Basics of digital signal processing. Applications, especially in the area of biological sciences.

Management of Research Data (3 cu) Metadata. Data presentation and storage: files, trees, databases. Statistical packages. Explorative data analysis. Sampling. Bootstrapping and randomization. Data visualization.

Introduction to Document Preparation (2 cu) HTML and WWW techniques, document preparation and electronic publishing.

Scientific Writing (4 cu) Sources of scientific information. Use of libraries and scientific data bases. The structure and details of a scientific publication. Examples of scientific Finnish or Swedish. Three individual survey writing exercises.

Tutoring (1 cu) In tutoring the second year students are divided into groups of ca. 15 students. Each group is led by a teacher or researcher (tutor) and the group meets monthly during two academic years. In the meetings students introduce matters related to their studies for discussion. In addition each student meets regularly with his/her tutor for private discussions to solve problems in his/her studies. The objective of tutoring is to reduce the number of student dropping out, uncover bottlenecks in the studies, and improve the relationships between students and teachers.

Graduate Courses 1996–98

a) General Computer Science

Design and Analysis of Algorithms (5 cu) Analysis techniques. Design techniques. Models of computation and lower bounds. Algorithms on sets. Graph algorithms. Approximation algorithms for NP-complete problems. Probabilistic algorithms. Parallel algorithms.

String Processing Algorithms (5 cu) Exact string matching. Approximate string matching. Pattern matching in static strings. Text databases and hypertext. Algorithm implementation and a comparison project.

Machine Learning (5 cu) History. Inductive learning: Learning in the blocks world, identification in the limit, version spaces. Learning classifiers: Finite automata, case-based, rules, decision trees, neural networks, genetic algorithms. PAC-learning: basics, Occam's razor, Vapnik-Chervonenkis dimension, learning by queries, PAC and noise, relation of different models. PAC and classifier learning. Inductive logic programming. Real-world applications.

Elements of Uncertain Reasoning (4 cu) Intuitions behind Bayesian modeling. Elements of Bayesian inference. Bayesian networks and their construction from data. Minimum encoding modeling. Applications of Bayesian modeling.

Problem Solving (2 cu) Creative problem solving in learning, teaching, and research.

Data Compression (4 cu) Text compression. Image compression. Compression in telecommunications. An implementation project.

Advanced Computer Graphics (4 cu) A selection of advanced topics such as tray tracing, radiosity, solid modeling, illumination and color, scientific visualization, etc. are taken as a theme of the course. Individual and group work, report writing and oral presentations by the participants.

Fundamentals of Image Processing (2 cu) Basic principles and methods of digital image processing.

Applications of Image Processing (2 cu) How to use digital image processing in various applications, possibilities and restrictions.

Robotics (4 cu) Types and applications of robots. Components of a robot. Architectures. Autonomous mobile robots: Navigation and motion planning. Robot learning: Reinforcement learning, Q learning.

Computational Biology (3 cu) Molecular biology. Sequence comparison and database search. Fragment assembly of DNA. Physical mapping of DNA. Phylogenetic trees. Genome rearrangements. Molecular structure prediction.

b) Computer Software

Distributed Operating Systems (4 cu) Kernel functionality. File service. Name service. Time and coordination. Replication. Distributed transactions. Recovery and fault tolerance.

Performance Evaluation (2 cu) General performance modeling concepts. Queueing network models and their solutions. Workload modeling. Emphasis on applications.

Computer Networks (4 cu) Formal specification and analysis of communication protocols, specification case studies. Security in computer networks: encryption, authentication, digital signatures etc.

Temporal Logic and Verification (3 cu) An introduction to linear and branching-time temporal and modal logics and their applications in specifying and verifying requirements of concurrent and reactive systems.

Principles of Concurrent Programming I (3 cu) Basic abstractions in concurrent programming, analysis and verification of algorithms and synchronization primitives for shared-memory parallel programs.

Principles of Concurrent Programming II (3 cu) An overview of concurrent programming in a distributed memory computer architecture where communication is realized with message passing. Besides formal analysis of this computing model a collection of distributed algorithms are studied to illustrate and concretize this approach.

Computer Architectures (4 cu) Structure of computer architectures, from instruction sets to I/O systems. The main emphasis is on uniprocessor systems.

Object-Oriented Programming (4 cu) Introduction to object-oriented thinking. Pure object-oriented languages. Object-oriented analysis and design. Hybrid object-oriented languages. Design patterns.

Compilers (6 cu) Lexical analysis, syntax analysis, semantic analysis, and code generation; use of metatools, laboratory assignments.

Mobile (Wireless) Data Communications (3 cu) Basic technologies, major challenges, and future of wireless data networking including: wireless data networks (GSM, Mobitex, CDPD, etc.), challenges of mobile (wireless) data networking, mobile IP, current solutions to the challenges, and possible applications for nomads. The course includes also minor projects in which participants design systems that employ wireless data networking.

The LOTOS Specification Language (3 cu) Lotos and its extensions. Specification and analysis of distributed systems. Lotos programming environments.

c) Information Systems

Database Systems II (5 cu) Physical data organization in databases. Index structures for files. B-trees. Dynamic hashing. Query processing and optimization for relational database systems. Join algorithms. Query optimization for distributed databases. Crash recovery. Concurrency control. Transaction management in client-server architectures. Distributed transactions.

Object-Oriented Databases (4 cu) Basic concepts of object data management. Object-relational database systems. Object data modelling in ODMG-93. Object query languages. Object calculus. Views. Object storage. Storage of large objects. Single-level store. Performance of a memory-mapped ODB. Indexing. Clustering. Transaction management. Object data caching.

Spatial Information Systems (3 cu) Spatial data and geographic database systems. Topology of planar graphs. Spatial data types based on finite-resolution geometry. Spatial index structures. Spatial joins. Space-filling curves and the Peano model. Topological relationships. Constraint databases and the polynomial model. Object-relational spatial databases.

Transaction Processing (4 cu) Serializability theory. Locking and non-locking schedulers. Multiversion concurrency control. Centralized and distributed recovery. Management of replicated data. Multidatabase transaction management. Cooperative transaction management. Prototype systems.

User Interfaces (4 cu) Concrete user interface solutions and their foundations. Aspects of user interface design: cognitive psychology, graphical user interfaces, user interface management systems, usability and testing. Current fields of research, e.g., WWW, multimedia, computer-supported co-operative work, virtual reality.

Computer-Supported Cooperative Work (3 cu) Concepts and history of CSCW. Advanced electronic mail and news systems. Meeting support systems. Collaborative writing. Hypertext and WWW in collaboration. Shared data. Workflow principles. Social and organizational aspects.

Information Retrieval Methods (3 cu) Traditional and new information retrieval methods, including networked retrieval and information exploration. Information filtering. Digital libraries.

Information Systems Development Methodologies (4 cu) Information system life-cycle models. Comparison of methods. Research on development methodologies. SA, SSADM, JSD, IE, OMT, OOSE, OOA and new object-oriented methods. Principles of Computer-Aided Software Engineering (CASE). Meta CASE, CASE repositories and processing of CASE data.

Knowledge Discovery in Databases (3 cu) Data mining. Finding frequent patterns in sequences. Integrity constraints in databases. Advanced techniques.

Seminar Courses in 1996–98

a) General Computer Science

Analogic Computation
Animation of Algorithms
Artificial Life
Biodatabases
Computers in Special Needs Education
Computer Science Education
Computer Uses in Education
Evolutionary Computation
Information Technology as a Means for Communication
Quantum Computing
Reasoning About Knowledge
Research Seminar on Computational Biology
Research Seminar on Machine Learning and Data Mining

b) Computer Software

Computer Network Architectures ATM Networking
CASE Systems
Colloquium on Distributed Systems
The CORBA Platform for Distributed Software Systems
Data Communications Technology
Security in Distributed Systems

Design and Implementation of the Linux Operating System
Implementation of Object-Oriented Languages
Internet Protocols
Object Architectures
Object-Oriented Analysis and Design
Object-Oriented Programming
Open Distributed Computing
Real Time Systems
Research Seminar on Formal Specification and Verification
Research Seminar on Mobile Workstations
Software Engineering
Software Metrics
Software Testing
Supercomputers with Distributed Memory
Telecommunications Technology
TINA
The Z Specification Language

c) Information Systems

Geographical Information Systems
Hypertext Systems
Information Retrieval Methods
Knowledge Discovery
Knowledge Discovery and Pattern Recognition
Object-Oriented Databases
Research on User Interfaces
Research Seminar on Databases
Research Seminar on Information Systems
Text Databases
User Interface Research

4.4 Accepted Theses

Theses for Doctor of Philosophy

Seppo Linnainmaa: Analysis of Some Known Methods of Improving the Accuracy of Floating-Point Sums. *BIT* **14** (1974), 167–202.

Eljas Soisalon-Soininen: Characterization of LL(k) Languages by Restricted LR(k) Grammars. A-1977-3.

Esko Ukkonen: On the Effect of Rounding Errors on the Flow of Control in Numerical Processes. A-1977-7.

Ralph-Johan Back: On the Correctness of Refinement Steps in Program Development. A-1978-4.

V.-E. Juhani Virkkunen: A Unified Approach to Floating-Point Rounding with Applications to Multiple-Precision Summation. A-1980-1.

Hannu Erkiö: Studies on the Efficiency of Certain Internal Sort Algorithms. A-1980-4.

Seppo Sippu: Syntax Error Handling in Compilers. A-1981-1.

Kari-Jouko Räihä: A Space Management Technique for Multi-Pass Attribute Evaluators. A-1981-4.

Timo Alanko: Empirical Studies of Program Behaviour in Virtual Memory. A-1983-3.

Kai Koskimies: Extensions of One-Pass Attribute Grammars. A-1983-4.

Heikki Mannila: Instance Complexity for Sorting and NP-complete Problems. A-1985-1.

Ilkka Haikala: Program Behaviour in Memory Hierarchies. A-1986-2.

Pekka Orponen: The Structure of Polynomial Complexity Cores. A-1986-3.

A. Inkeri Verkamo: Sorting in Hierarchical Memories. A-1988-1.

Jorma Tarhio: Attribute Grammars for One-Pass Compilation. A-1988-11.

Gösta Grahne: The Problem of Incomplete Information in Relational Databases. A-1989-1.

Niklas Holsti: Script Editing for Recovery and Reversal in Textual User Interfaces. A-1989-5.

Kimmo Raatikainen: Modelling and Analysis Techniques for Capacity Planning. A-1989-6.

Jukka Paakki: Paradigms for Attribute-Grammar-Based Language Implementation. A-1991-1.

Jyrki Kivinen: Problems in Computational Learning Theory. A-1992-1.

Patrik Floréen: Computational Complexity Problems in Neural Associative Memories. A-1992-5.

Pekka Kilpeläinen: Tree Matching Problems with Applications to Structured Text Databases. A-1992-6.

Jaana Eloranta: Minimal Transition Systems with Respect to Divergence Preserving Behavioural Equivalences. A-1994-1.

Petri Myllymäki: Mapping Bayesian Networks to Stochastic Neural Networks: A Foundation for Hybrid Bayesian-Neural Systems. A-1995-1.

Roope Kaivola: Equivalences, Preorders and Compositional Verification for Linear Time Temporal Logic and Concurrent Systems. A-1996-1.

Tapio Elomaa: Tools and Techniques for Decision Tree Learning. A-1996-2.

Helena Ahonen: Generating Grammars for Structured Documents Using Grammatical Inference Methods. A-1996-4.

Hannu Toivonen: Discovery of Frequent Patterns in Large Data Collections. A-1996-5.

Henry Tirri: Plausible Prediction by Bayesian Inference. A-1997-1.

Greger Lindén: Structured Document Transformations. A-1997-2.

Matti Nykänen: Querying String Databases with Modal Logic. A-1997-3.

Selection of Theses for Licentiate of Philosophy

Kari-Jouko Räihä: On Attribute Grammars and Their Use in Compiler Writing Systems. A-1977-4.

Kai Koskimies: A Study on the Programming Language Euclid. A-1980-2.

Harri Laine: Semantic Integrity and Data Base Update in the Grammatical Data Base Model. C-1981-48.

Jorma Tarhio: Attribute Evaluation During LR Parsing. A-1982-4.

Juha Vihavainen: Design of the Simulation Language Mode (in Finnish). C-1987-42.

Jukka Paakki: Generating One-Pass Semantic Analysis for a Compiler. A-1988-8.

Eeva Hartikainen: Specification and Design of Distributed Programs That Use Broadcasting. A-1988-13.

Jukka Mutikainen: An Experimental Study of Attribute Selection Criteria in Decision Tree Induction. C-1989-67.

Erja Nikunen: Views in Structured Text Databases. C-1990-60.

Liisa Räihä: Sequence Comparison: Computation of the Edit Distance. C-1991-59.

Juha Puustjärvi: Management of Transactions in Heterogeneous Distributed Database System. C-1992-31.

Pasi Tapanainen: Finite-State Parsing for Natural Languages (in Finnish). C-1993-7.

Greger Lindén: Incremental Updates in Structured Documents. C-1993-19.

Helena Ahonen: Generating Grammars for Structured Documents Using Grammatical Inference Models. C-1994-65.

Päivikki Parpola: Object-Oriented Knowledge Acquisition (in Finnish). C-1995-11.

Pekka Siltanen: Free Form Modeling Using Cubic Bezier Curves and Surfaces. C-1995-19.

Lea Kutvonen: The Role of Trading Function in Open Distributed Processing Infrastructure. C-1996-84.

Erkki Sutinen: Using q-Grams in Approximate Pattern Matching. C-1996-90.

Matti Nykänen: Alignment Calculus: A Modal Logic for Querying String Databases. C-1996-97.

Juha Taina: Database Architecture for Intelligent Networks. C-1997-50.

Jukka Laine: Database Reorganization. C-1997-85.

Selection of Master's Theses in 1996–1998[1]

a) General Computer Science

1996

Erkki Heilakka: Corrosion Models Using Procedural Textures. C-1996-32.

Kari Huttunen: Scanning in Approximate String Matching. C-1996-85.

Markus Huttunen: Using Machine Learning Methods for Modelling Hydrological Time Series. C-1996-38.

[1] Written mostly in Finnish, some in Swedish or English.

Risto Kankkunen: Automatic Construction of Hyphenation Algorithms Using the Synthesis Method of Finite Automata. C-1996-62

Hanna Keskinen: The Computer Driving License Examination and Its Relationship to the Basic Computer Science Studies at Universities. C-1996-100.

Simo-Pekka Lahtinen: Visual Debuggers. C-1996-60.

Tei Laine: Using a Genetic Algorithm for Optimization in a Changing Environment. C-1996-56.

Stefano De Pascale: An Implementation of Interval Computations on the Spreadsheet. C-1996-89.

Antti Pesonen: An Interval Classifier. C-1996-18.

Juho Rousu: Constructing Decision Trees and Lists Using the MDL Principle. C-1996-15.

Mika Seppänen: Constructing Bayesian Networks from Data. C-1996-33.

Mika Sorsa: Qualitative Bayesian Networks. C-1996-98.

1997

Teemu Antti-Poika: Using Bayesian Networks in Classification. C-1997-68.

Kimmo Fredriksson: Learning the Structure of Neural Networks Using Genetic Algorithms. C-1997-34.

Raul Hakli: Extending a Database Management System with a String Query Language. C-1997-89.

Jaakko Kurhila: Computer Graphics Studies and Learning Materials on the Web. C-1997-26.

Mikko Oksalahti: Simulating Light Reflection and Shadows in Computer Graphics. C-1997-39.

Jan Täppinen: Neural Networks in Financial Applications. C-1997-76.

b) Computer Software

1996

Heikki Helin: A Description Language for Generic Events in the Mowgli System. C-1996-88.

Nina Hutteger: Short Message Service in the Mowgli Architecture. C-1996-102.

Jani Kiiskinen: A Protocol Adapted for Wireless Data Communication and Its Performance when Compared to the TCP/IP Protocol. C-1996-74.

Oskari Koskimies: An Agent-Oriented Application Framework in the Mowgli Architecture. C-1996-87.

Petteri Lappalainen: Process Migration in a Fault-Tolerant Real-Time System. C-1996-86.

Mika Liljeberg: Performance of World-Wide-Web over a Cellular Telephone Network. C-1996-51.

Jussi Ollikainen: Implementation of a TINA Communication Session in an ATM Network. C-1996-34.

Timo Sivonen: Security Management on Distributed Systems. C-1996-37.

1997

Leif Björklund: Mobile Internet Packet Routing with Cellular Networks. C-1997-73.

Max Hamberg: ODP Viewpoint Specifications as a Methodology. C-1997-30.

Mika Hatanpää: Software Reuse. C-1997-64.

Vesa Karhu: Manual Verification of Message Passing Algorithms with the Use of Temporal Logic. C-1997-62.

Petteri Kaskenpalo: Data Security in the Mowgli Architecture. C-1997-52.

Esko Koskimaa: Meta Facilities in Object-Oriented Languages. C-1997-86.

Ville Lavonius: Implementation of POSIX/DMX Coworking in the DX200 System. C-1997-35.

Sami Lilja: Adapting the DX-200 Software for the Chorus Microkernel. C-1997-58.

Matti Luukkainen: Applying Process Algebras in Modelling and Verifying Data Security. C-1997-3.

Samu Paajanen: Use Cases in System Testing. C-1997-20.

Mikko Puuskari: Standardization Framework for Slow Wireless Link Protocols in an Indirect Architecture. C-1997-74.

Marita Raja: The Mobile-Connection Host Management in the Mowgli System. C-1997-90.

Quanyi Sun: Testing Real-Time Embedded Software. C-1997-80.

Timo Suominen: Analysis of Embedded SQL in the HyperSoft System. C-1997-44.

Janne-Pekka Surakka: A Global Mailbox-Service for Intelligent Telephones. C-1997-54.

Linus Torvalds: Linux: A Portable Operating System. C-1997-12.

Katri Turunen: A Distributed Memory-File System Based on the Virtual Memory Concept. C-1997-77.

Antti Viljamaa: Graphical Simulation of Data Communication Protocols. C-1997-4.

Jukka Viljamaa: Interpretation of the Protocol Language Kannel. C-1997-21.

c) Information Systems

1996

Petri Elovaara: Evaluation of Concurrency Control Algorithms for Real-Time Databases. C-1996-64.

Juha Gustafsson: Concept Learning and Theory Construction in Logical Databases. C-1996-39.

Barbara Heikkinen: Qyerying Structured Hypertext Documents in an Object Database. C-1996-27.

Jani Kokkonen: Implementation Techniques of Collaborative Drawing Applications. C-1996-59.

Katariina Lapatto: Aggregation Operators in Logic Databases. C-1996-65.

Kalevi Nummelin: Transfer of Data between Health Care Organizations. C-1996-91.

Ristomatti Partanen: Using Sounds in User Interfaces with Focus on Patient Control. C-1996-41.

Marko Salmenkivi: Modeling Event Series and Generating MCMC-Simulation Programs to Approximate the Posterior Distribution of Model Parameters. C-1996-70.

Marko Vartiainen: User Interface Design for a Customer Service Application. C-1996-71.

1997

Tiina Aalto-Ylönen: Intelligent Agent-Based Collaborative Learning Environments. C-1997-95.

Pasi Ahonen: Implementation Principles for Digital Libraries. C-1997-1.

Lasse Akselin: Graphical Representation of Text Region Queries. C-1997-27.

Tero Jankko: Testing of Graphical User Interfaces. C-1997-87.

Jukka Kiviniemi: Real-Time Object-Oriented Data Model for Telecommunications. C-1997-71.

Tommi Korhonen: Windowing and Coordination of Multiple Windows. C-1997-70.

Kaisa Kostiainen: Digital Aids for Constructive Learning. C-1997-56.

Susanna Lehtinen: A Multidimensional Index Structure for a Main-Memory Database. C-1997-94.

Antti Leino: The Influence of Conference Systems on Conference Discussion. C-1997-41.

Heli Liukkonen: Data Caching in Object Databases. C-1997-19.

Mia Lähteenmäki: Evaluation of the ODMG-93 Object-Oriented Database Standard. C-1997-55.

Vesa Ollikainen: Evaluation Algorithms for Efficient Distance. C-1997-92.

Harri Rautio: An Implementation of Particle Systems in a PC. C-1997-42.

Riitta Sillanpää: Hypertext Visualization. C-1997-31.

Minna Walden: Benefiting from a User History in a User Interface. C-1997-61.

4.5 Abstracts of Recent Ph.D. Theses

Helena Ahonen: *Generating Grammars for Structured Documents Using Grammatical Inference Methods.* A-1996-4.

Dictionaries, user manuals, encyclopedias, and annual reports are typical examples of structured documents. Structured documents have an internal, usually hierarchical, organization that can be used, for instance, to help in retrieving information from the documents and in transforming documents into another form. The document structure is typically represented by a context-free or regular grammar. Many structured documents, however, lack the grammar: the structure of individual documents is known but the general structure of the document class is not available. Examples of this kind of documents include documents that have Standard Generalized Markup Language (SGML) tags but not a Document Type Definition (DTD).

In this thesis we present a technique for generating a grammar describing the structure of a given structured document instances. The technique is based on ideas from machine learning. It forms first finite-state automata describing the given instances completely. These automata are modified by considering certain context conditions; the modifications correspond to generalizing the underlying language. Finally, the automata are converted into regular expressions, which are then used to construct the grammar. Some refining operations are also presented that are necessary for generating a grammar for a large and complicated document. The technique has been implemented and it has been experimented using several document types.

Hannu Toivonen: *Discovery of Frequent Patterns in Large Data Collections.* A-1996-5.

Data mining, or knowledge discovery in databases, aims at finding useful regularities in large data sets. Interest in the field is motivated by the growth of computerized data collections and by the high potential value of patterns discovered in those collections. For instance, bar code readers at supermarkets produce extensive amounts of data about purchases. An analysis of this data can reveal useful information about the shopping behavior of the customers. Association rules, for instance, are a class of patterns that tell which products tend to be purchased together.

The general data mining task we consider is the following: given a class of patterns that possibly have occurrences in a given data collection, determine which patterns occur frequently and are thus probably the most useful ones. It is characteristic for data mining applications to deal with high volumes of both data and patterns.

We address the algorithmic problems of determining efficiently which patterns are frequent in the given data. Our contributions are new algorithms, analyses of problems, and pattern classes for data mining. We also present extensive experimental results. We start by giving an efficient method for the discovery of all frequent association rules, a well known data mining problem. We then introduce the problem of discovering frequent patterns in general, and show how the association rule algorithm can be extended to cover this problem. We analyze the problem complexity and derive a lower bound for the number of queries in a simple but realistic model. We then show how sampling can be used in the discovery of exact association rules, and we give algorithms that are efficient especially in terms of the amount of database processing. We also show that association rules with negation and disjunction can be approximated efficiently. Finally, we define episodes, a class of patterns in event sequences such as alarm logs. An episode is a combination of event types that occur often close to each other. We give methods for the discovery of all frequent episodes in a given event sequence.

The algorithm for the discovery of association rules has been used in commercial data mining products, the episode algorithms are used by telecommunication operators, and discovered episodes are used in alarm handling systems.

Henry Tirri: *Plausible Prediction by Bayesian Inference.* A-1997-1.

The capability to perform inference with uncertain and incomplete information is characteristic to intelligent systems. Many of the research issues in artificial intelligence and computational intelligence can actually be viewed as topics in the "science of uncertainty," which addresses the problem of plausible inference, i.e., optimal processing of incomplete information. The various different approaches to model and implement intelligent behavior such as neural networks, fuzzy logic, non-monotonic (default) logics and Bayesian networks all address the same problem of finding an appropriate language and inference mechanism to perform plausible inference, needed to implement such activities as prediction, decision making, and planning.

In this work we study the problem of plausible prediction, i.e., the problem of building predictive models from data in the presence of uncertainty. Our approach to this problem is based on the language of Bayesian probability theory both in its traditional and information theoretic form. We study Bayesian prediction theoretically and empirically with finite mixture models. Such models are interesting due to their ability to accurately model complex distributions with few parameters. In addition, finite mixture models can be viewed as a probabilistic formulation of many model families commonly used in machine learning and computational intelligence. We first address the question of how an intelligent system should predict given the available information. We present three alternatives for probabilistic prediction: single model based prediction, evidence based prediction, and minimum encoding based prediction. We then compare the empirical performance of these alternatives by using a class of finite mixture models. The empirical results demonstrate that, especially for small data sets, both the evidence and the minimum encoding approaches outperform the traditionally used single model approach.

We then focus on the problem of constructing finite mixture models from the given data and a priori information. We give the Bayesian solution for inferring both the most probable finite mixture model structure, i.e., the proper number of mixture components, and the most probable model within the class. For general mixture models the exact solution in both problems is computationally infeasible. Thus we also evaluate the quality of approximate approaches.

The Bayesian predictive approach presented can be applied to a wide class of prediction problems appearing in various application domains, e.g., medical and fault diagnostic problems, design problems and sales support systems. Using publicly available data sets, we demonstrate empirically that Bayesian prediction with finite mixtures is highly competitive when compared to the results achieved with other popular non-Bayesian approaches using, for example, neural network and decision tree models. The Bayesian prediction method presented constitutes the kernel of the `D-SIDE/C-SIDE` software currently used in industrial applications.

Greger Lindén: *Structured Document Transformations.* A-1997-2.

We present two techniques for transforming structured documents. The first technique, called TT-grammars, is based on earlier work by Keller et al., and has been extended to fit structured documents. TT-grammars assure that the constructed transformation will produce only syntactically correct output even if the source and target representations may be specified with two unrelated context-free grammars. We present a transformation generator called ALCHEMIST which is based on TT-grammars. ALCHEMIST has been extended with semantic actions in order to make it possible to build full scale transformations. ALCHEMIST has been extensively used in a large software project for building a bridge between two development environments. The second technique is a tree transformation method especially targeted at SGML documents. The technique employs a transformation language called TranSID, which is a declarative, high-level tree transformation language. TranSID does not require the user to specify a grammar for the target representation but instead gives full programming power for arbitrary tree modifications. Both ALCHEMIST and TranSID are fully operational on UNIX platforms.

Matti Nykänen: *Querying String Databases with Modal Logic.* A-1997-3.

New application areas for databases demand the possibility of managing not only the traditional atomic but also structured data types. One type of this kind is a finite sequence of characters drawn from a finite alphabet. These string databases are important for example in molecular biology data management, because they can for instance represent DNA sequences directly as strings from alphabet $\{\mathtt{A}, \mathtt{C}, \mathtt{G}, \mathtt{T}\}$. Then the query language must be able to manipulate strings both as indivisible entities and as ordered sequences of distinct characters, perform pattern matching in strings, compare several strings with each other, and generate new strings not yet in the database. This work presents Alignment Calculus, a new modal logic extension of the relational calculus, which satisfies all these requirements with a new unified string manipulation formalism. This formalism is based on the concept of multiple alignment of several strings; computational molecular biology employs an analogous concept. In the language, alignments are described logically with string formulae, a new extension of linear-time temporal logic into multiple strings. The abstract expressive power of Alignment Calculus is shown to equal the Arithmetical Hierarchy. The work develops also a syntactically safe sublanguage, whose queries require only finite computational resources. This sublanguage is constructed by first identifying a decidable subcase of string formula safety analysis, even though the general problem is shown to be undecidable. This safe sublanguage is shown to equal the Polynomial-time Hierarchy in its expressive power, and therefore to capture all the string queries occurring in practical situations. The operational counterpart of Alignment Calculus, Alignment Algebra, is developed by replacing the selection operator of relational algebra with one employing multitape nondeterministic finite state automata corresponding to the aforementioned string formulae, and adding an explicit domain symbol. The aforementioned safe sublanguage has also a counterpart in this algebra: expressions, in which all the domain symbol occurrences are restricted by immediately enclosing automata. A finite evaluation strategy is developed for these expressions.

Chapter 5

Library

The department maintains a library with large collections of literature in computer science. The library is jointly financed with the University Computing Centre and is mainly used by the staff and advanced students of the department.

Established in 1967, the library now holds about 52,000 volumes of literature, making it the largest computer science library in Finland. The annual cumulation is about 1,200 monographic titles and 300 journal subscriptions. The floor area is 408 sq. meters including a reading room of 60 seats. Admission to the premises is free and collections are freely available to all visitors. Home loans, however, are normally granted only to university personnel and advanced students of the department (for exceptions, consult the library staff).

To help users search and locate the required literature, the library maintains a www-browsable database of its holdings (http://www.cs.helsinki.fi/kirjasto/). The database includes all journal titles and about 37,000 monographic titles, classified according to the CR Classification System of the ACM. The library is also responsible for the distribution of departmental reports, including Ph.D. theses. Paper copies may be requested from the library, and electronic versions are accessible through the department's FTP server.

The library has two full-time employees, one librarian and one secretary, assisted in their work by a library committee comprising of several members of the faculty. Any inquires to the library are welcome by post, telex, telephone or e-mail (library@cs.helsinki.fi).

Chapter 6

Computing Facilities

The department offers a wide range of services to support computing activities of the academic staff and students. The policy is to provide access to advanced hardware and software systems.

The computing facilities include a farm of servers (general purpose computers, file servers, and dedicated servers for mail, WWW, FTP, etc.) and a network of workstations and PC microcomputers. The departmental general purpose computers are an Alpha based Citum Power System (a repackaged Aspen server), a SPARCserver 670MP and a SPARCserver 10. (The SPARCservers will soon be replaced with a more powerful server.) The main file servers are Pentium based systems running Linux and utilizing RAID technology. The total disk space is currently well over 100 Gbytes. All the Alpha and Pentium based machines use Linux, but the SPARC computers run SunOS/Solaris. Together these systems support a wide variety of services, languages and software tools including electronic mail and news, graphics and visualization tools, several typesetting systems, and relational database systems. Special attention has been placed on security and reliability.

The workstation network consists of about 10 SPARCstations and about 280 PCs (mostly Pentiums with high resolution monitors) running Linux. Windows 3.1, Windows 95 or Windows NT can be used as an alternative for Linux. About 30 of the Linux workstations are mobile laptops which can join and leave the network dynamically. Networking is based on an ATM backbone with Ethernet edge devices. The department has six ATM-switches installed. The mobile laptops can also utilize a departmental 1 Mbit/s radio network which currently has 7 base stations. On the UNIX side (Linux, SunOS/Solaris), NFS is used to share common resources. On the Windows side, Samba (a UNIX hosted Lan Manager Server) is used. The workstations are used as tools for software development, in research and all levels of teaching.

The network of the department is connected to the university backbone network, giving access to computers at the University Computing Centre as well as to the FUNET wide area network that links Finnish universities and research establishments. The computers operated by the Computing Centre include SPARC (Sun, Solbourne, Axil), Digital Alpha and HP machines running under UNIX. Services provided by the Computing Centre include the Oracle and Ingres database management systems, the SAS statistical analysis package, the NAG numerical library, and Pascal, Ada, and Prolog programming environments.

In addition, the department has access to a Cray C94, a Cray T3E, two SGI Power Challenge/Onyx, a Digital AlphaServer 8400, and other supercomputers at the Center for Scientific Computing.

The national FUNET network is further connected to the Nordic University Network, Nordunet, with a dedicated G.703(E3) 34 Mbit/s line. The Nordunet has also a 79 Mbit/s terrestrial connection to NAP Pennsauken in the United States as well as many connections to the European network infrastructure. This means that the department is very well connected to the Internet.

Chapter 7

International Relations

Student Exchange

The department is involved in the Sokrates/Erasmus programme (European Community Action Scheme for the Mobility of University Students). Through this programme students of the department have an opportunity to study at several European universities, including the Technical Universities of Darmstadt, Graz and Munich, the University of Bath, the University of Catalonia in Barcelona and the National Institute of Applied Sciences in Lyon.

The department also accepts visiting students from these and other partner institutions. Exchange of students within the Nordic countries is also easy with the financial support of NORDPLUS, a program coordinated by the Nordic Council of Ministers.

Postgraduate studies at the department can usually be accomplished in English. The department may admit foreign graduate and postgraduate students, who want to study for either the M.Sc. or Ph.D. degree, or visiting students. The Academy of Finland and the Ministry of Education and Science offer grants for foreign postgraduate students within the international exchange programs. Postgraduate studies are usually performed within one of the graduate schools, HeCSE or HeCSE-http://www.cs.helsinki.fi/combi (see Section 4.2).

Research Cooperation

The department has active contacts with many European and American computer science departments. Faculty members have joint research efforts with individual researchers from abroad. The results of this international cooperation are partly presented in the publication list. Many faculty members have spent long periods of time abroad as visiting researchers or guest professors. In turn, apart from more than 20 short term visitors every year, the department has hosted several visiting foreign scholars. A list of long term visitors in 1996–98 is given below.

The department has actively taken part in the European research cooperation. The department has participated in the COST 247 project *Formal methods in Communication Protocol Design*.

The department has also taken part in the ESPRIT Programme in several projects. The largest of these are *KESO – Knowledge Extraction for Statistical Offices* (ESPRIT III 20596, Heikki Mannila, 1996–98) and *TransCoop – Transaction Management Support for Cooperative Applications* (ESPRIT III P8012, Henry Tirri, 1994–97). The department is a member institution in the working group *NEUROCOLT II – Neural and Computational Learning* (ESPRIT Working group EP 27150, Esko Ukkonen, 1994–) and in networks of excellence like *ML-net – Network of excellence in Machine Learning* (ESPRIT III P7115, Esko Ukkonen, 1994–) and *NEURONET - Network of excellence in Neural Networks* (ESPRIT III P8961, Henry Tirri, 1994–).

In addition the department is a member institution of EC ACTS-project DOLMEN *Service Machine Development for an Open Long-Term Mobile and Fixed Network Environment* (AC036, Kimmo Raatikainen, 1995–98). The project *Object-oriented programming and compiler construction* (Jukka Paakki, 1995–97) was funded jointly by the Academy of Finland and Deutscher Akademischer Austauschdienst.

Faculty Members Abroad in 1996–98

Helena Ahonen, Eberhard-Karls-Universität Tübingen, Germany, 8/97–

Tapio Elomaa, Institute for Systems, Informatics and Safety, Joint Research Centre, European Commission, Ispra (Va), Italy 2/97-

Gösta Grahne, Université de Paris-Sud, Centre d'Orsay, France, 5/96

Gösta Grahne, Concordia University, Montreal, Canada, 9/97–

Kari Granö, Technische Universität München, Germany, 2–4/96

Teemu Kerola, University of Texas, Austin, USA, 8/97-7/98

Jyrki Kivinen, University of California at Santa Cruz, USA, 12/97

Heikki Mannila, Max Planck Institut für Informatik, Saarbrücken, Germany, 9/95–2/96

Jukka Paakki, Technische Universität München, Germany 2–3/96

Juha Taina, University of Virginia, Charlottesville, USA 9–12/96

Martti Tienari, University of British Columbia, Vancouver, Canada, 5-7/96

Martti Tienari, La Trobe University, Bundoora, Australia, 8/96

Olli Yli-Harja, University East Anglia, Norwich, UK, 1–5/98

Foreign Visitors in 1996–98

Prof. Amihood Amir, Bar-Ilan University, Israel, 10/97

Dr. Hiroki Arimura, Kyushu University, Fukuoka, Japan, 5–10/96

Dr. Jean-François Boulicaut, INSA, Lyon, France, 11/97-8/98

Dr. Alvis Brazma, University of Latvia, Riga, 8–12/96, 1/98

Dr. Peter Cheeseman, NASA Ames Research Center, Moffett Field, CA, USA, 5/97

Dr. Gautam Das, University of Memphis, USA, 8/96, 5–6/97, 5–6/98

Dr. Dimitri Dracopoulos, Brunel University, London, UK, 8/96

Prof. Jürgen Eickel, Technische Universität München, Germany, 10/97

Dr. Ronen Feldman, Bar-Ilan University, Israel, 8/96

Dr. Peter Grünwald, CWI, Amsterdam, The Netherlands, 5/97-7/97

Dr. Dimitrios Gunopulos, IBM, Almaden, USA, 9–10/96

Mr. Alberto de las Heras, Universitat Politechnica de Catalunya, Spain 3/97-7/97

M.Sc. Riitta Höllerer, Technische Universität München, Germany, 9/96, 9–10/97

Dr. Inge Jonassen, University of Bergen, Norway, 11/96, 1/98

Dr. Roni Khardon, Harvard University, USA, 10/96

Prof. Hartmut König, Brandenburgische Technische Universität Cottbus, Germany, 8/97

Prof. Gad M. Landau, University of Haifa, Israel, 10/97

B.Sc. David Nespoli, USA, 1/96–12/96, 6/97

Mr. Torsten Polle, Universität Hildesheim, Germany, 8/96

Prof. Thomas Richardson, University of Washington, Seattle, USA, 4–7/97

Mr. Jacques Scubich, INSA, Lyon, France, 6/97

Dr. Kuldar Taveter, Tallinn Technical University, Estonia, 10–11/96

Prof. Leo Vohandu, Tallinn Technical University, Estonia, 12/96

Dr. Son T. Vuong, University of British Columbia, Vancouver, Canada, 6/97

Prof. Emmanuel Waller, Université de Paris-Sud, France, 8–9/96

Mr. Hannes Wettig, Universität Köln, Germany, 8/96-7/97

Other Activities

Faculty members regularly serve as referees for international computer science journals and conferences (see Section 3). Some faculty members are also reviewers for various review publications, such as Computing Reviews and Mathematical Reviews. Many faculty members have served in the program committees of various conferences (see Section 3). Prof. Esko Ukkonen is the editor-in-chief of the Nordic Journal of Computing. Prof. Heikki Mannila is the editor-in-chief of the Journal of Data Mining and Knowledge Discovery. Assoc. Prof. Matti Mäkelä is an international coordinator of ACM SIGNUM Newsletter.

The department hosted the joint standardisation meeting of ISO/IEC JTC1/SC21 WG7 and ITU-T SG4 Q.12/Q.14 in Helsinki, 14–25 July, 1997. The joint meeting covered 12 separate standardisation issues, and about 60 delegates were present from 13 countries.

The department organized the 6th Scandinavian Conference on Artificial Intelligence in Helsinki in August 1997. About 80 researchers took part. The conference proceedings were edited by Gösta Grahne and published by IOS Press in the Frontiers in Artificial Intelligence and Applications Series.

ISSN 1238-8645
ISBN 951-45-8138-5
Helsinki 1998
Helsinki University Printing House