C O N T E N T S

University of Helsinki
Department of Computer Science
Annual Report 2001

1. Department of Computer Science in 2001

After an expansive period of several years in the information technology sector, the year 2001 witnessed an economic regression in the form of falling stock options, companies going bankrupt, and employees of start-up companies losing their jobs. Especially the dreams in new e-based and mobile (3G) services were not realised to the extent some optimistic visionaries had envisaged.

4

On the other hand, the slight disappointment and disbelief in the ever-growing power of information technology also had some positive effects. The importance of substance, education and research was once again appreciated when analysing the reasons for failures in the industrial sector: an IT company without a well-founded focus and educated staff will soon be out of business.

The return of the academic ideals showed up at the Department, not only in the substantial and even increased funding for research and education, but also in the large number of students attending the basic courses of the curriculum. As the labour market was a bit colder than in the previous years, at least some of the part-time students wanted or had to turn into full-time ones. This trend did not yet result in a larger number of degrees but even this formally acknowledged improvement will hopefully be seen within the next few years (at least if we are saved from a new radical expansion of the IT sector).

PERSONNEL

Even under the hard pressure from the outside world, the staff of the Department remained as committed as ever. Most of them carried out both teaching and research of high quality. New positions could be filled with qualified persons, with one notable exception: the new professorship in distributed systems and data communications had to be left open due to having no scientifically qualified applicants. This example demonstrates the lack of postgraduates, doctors and thereby professors in areas where the universities have to compete with the industries in their personnel.

When establishing new positions, the Department followed its strategic plans for the years 2001-2003. Special emphasis was laid on administrative positions that would free the teachers and researchers from tasks that are beyond their main functions and competence. The most notable administrative post was that of an office manager whose task it is to coordinate the general administration of the Department. The position was announced in Autumn 2001 and filled in March 2002 when Päivi Karimäki-Suvanto was appointed to the post. Other major administrative arrangements included the assignment of Professor Hannu Erkiö as the "educational manager" of the Department (coming into effect in Autumn 2002) and the nomination of Janne Saarinen as a full-time "information officer" (from June 2002).

The total number of man-years in 2001 was

126.6, a slight increase when compared to year 2000 with 125.1 man-years.

EDUCATION

The year 2001 can well be nominated the year of University education. A global trend was the emphasis on high-level education at universities and departments, as well as on the teaching merits of the staff. At the individual level, University portfolios with educational merits in a central role were taken into use at the University, both in tenure nominations and in the evaluation of academic competence. At the University level, a comprehensive evaluation project of the education was started, with ultimate results in the form of improvement activities expected in the years 2003 to 2005. At the national level, the "virtual university" project attained wide interest and generated a great deal of new ideas and pilots in computer-assisted education.

The Department took part in the education evaluation project of the University by producing (together with the students) a self-evaluation report. The Department report was then integrated with the reports from the Departments of Mathematics, Chemistry and Physical Sciences into a joint report that was given as input to an international educational panel. The panel visited the Departments in March 2002 and gave very positive feedback on the quality of their education. The findings in the self-evaluation report and the panel report will be used in the next few years to improve the quality of teaching at the Department.

The amount of classroom teaching sank to 12962 classroom hours from 15050 hours in 2000, but the number of credit units earned in computer science passed the level of 20,000 cu again. The number of MSc degrees was 61 and the number of PhD degrees was four.

The Department took an active role in virtual education by establishing a virtual strategy working group whose main task was to produce the Department strategy in virtual education.

The strategy was completed in Spring 2002, and its implementation will start in Autumn 2002 under the management of the working group. Another major investment in virtual education was made when the Department joined the European TUeLIP association, whose goal it is to produce a series of joint virtual courses in computer science to all its members.

The "upgrading programme" in computer science, started in 1999, was continued in 2001. Based on the experiences from the previous years, the department organised a selection of courses in the programme in the evenings and weekends so as to make them easier for the upgrading students to attend. The upgrading programme has turned out to be a significant source of funding for the Department. In 2001, 4.9 million FIM was allocated to the Department for organising the upgrading programme (the amount of funding being FIM 2.8 million in 2000). Together with the national information society programme (with the funding of FIM 3.9 million in 2001), the language technology programme (FIM 0.5 million) and the graduate school programme (FIM 2.8 million), the special educational programmes provided a funding of FIM 12.1 million in 2001 (almost 30% of the total Department budget). Such a dependence on short-term educational projects may prove a difficult problem for the Department in the next years if this kind of temporary funding will not be moved into the more stable form of basic funding.

RESEARCH

Most of the research activities at the Department are funded by external sources. In addition, the Department tries to allocate time for research for its regular teaching staff, but this goal has not been met to a satisfactory level due to the pressure on university education in the IT sector.

The external research funding decreased slightly when compared to year 2000 (FIM 12.4 million vs. FIM 13.2 million). On the other

5

hand, the funding from the Academy of Finland increased from FIM 2.6 million to FIM 3.4 million, nearly reaching the funding from the National Technology Agency Tekes (FIM 3.6 million) and passing the funding from the industries (FIM 3.3 million). This distribution can be regarded as a trend from industry-driven "applied" research to more scientific "basic research", probably a move in the right direction. The research output in the form of publications (96) slightly decreased from 2000 (107).

In 2001, the most visible research highlights at the Department were the nomination of the *From Data to Knowledge* (FDK) research unit as a Centre of Excellence by the Academy of Finland and the establishment of the *Basic Research Unit* (BRU) within Helsinki Institute for Information Technology, a joint venture between the University of Helsinki and Helsinki University of Technology. It is expected that FDK and BRU will raise the scientific level of research at the Department even higher within the next few years.

As a whole, the year 2001 can be considered successful both in research and education, thanks to the intensive and committed work carried out at the Department. For this, I want to express my thanks to the staff and also to students without whom at least the education would be rather useless.

Jukka Paakki
Head of the Department

2. Overview

2.1 THE STRATEGY OF THE DEPARTMENT

The Department carries out high quality research in its core fields with a long research tradition as well as opens new research fields on topics that are of emerging interest in the international research community. The selection of the new fields depends on the substance development of computer science, the needs of the information society, and the interest profiles of senior researchers. In its research, the Department emphasises the importance of co-operation with IT companies and research units whose scientific progress is largely depending on the availability of intensive computer applications.

In its education, the Department offers an up-to-date and versatile programme emphasising the conceptual basis of computer science. This programme educates experts on research and development for the IT industry, as well as researchers in computer science. The educational strategy of the Department is to combine theory and practice with a substantial amount of constructive exercises and projects based on the conceptual lecture courses.

The education is based on the traditional core of computer science, adapting to new demands and needs according to the developments in the field. The base of the educational programme, organised around the selected specialisation areas, is kept stable while the courses in the new areas are in most cases provided as optional units of the curriculum. The Department carries out constant development of its educational methods, the central approaches being problem-based and student-centred learning and the use of "virtual" computer-assisted methods and techniques at different levels of depth.

2.2 ORGANISATION

The research and education of the Department are organised into five specialisation areas: algorithms and data structures, distributed systems and data communication, information systems, intelligent systems, and software engineering. Each area is responsible for advancing scientific work in the field and for planning and implementing the corresponding educational programme.

A student can organise her studies according to one of the four education subprogrammes: computer science, applied computer science, teacher in computer science, and computer mathematician. The subprogrammes differ mainly in the volume of the major subject, computer science, included and in the selection and quantity of the minor subjects. The specialisation areas of research and education compose the computer science subprogramme, each of them sharing a common base curriculum (the cum laude approbatur level) and providing a specialised curriculum at the laudatur level.

The Department hosts two extensive research units with special funding. The *From Data to Knowledge* (*FDK*) research unit has been nominated as a Center of Excellence by the Academy of Finland for a six years term 2002-2007, and the *Basic Research Unit* (*BRU*) of the Helsinki Institute for Computer Science (HIIT, a joint research institute with Helsinki University of Technology) has been founded in 2001 with basic funding from the Rector of the University. In addition, the Department infrastructure offers the services of high-level computing facilities and of an extensive library (being formally part of the Kumpula Science Library of the Faculty of Science).

7

The organisation of the Department is sketched in the following figure.

Department Steering Committee		
Administration, services	Education and research areas	Research units
Office	Subprogrammes	FDK
IT maintenance	Specialisation areas	HIIT-BRU
Library		
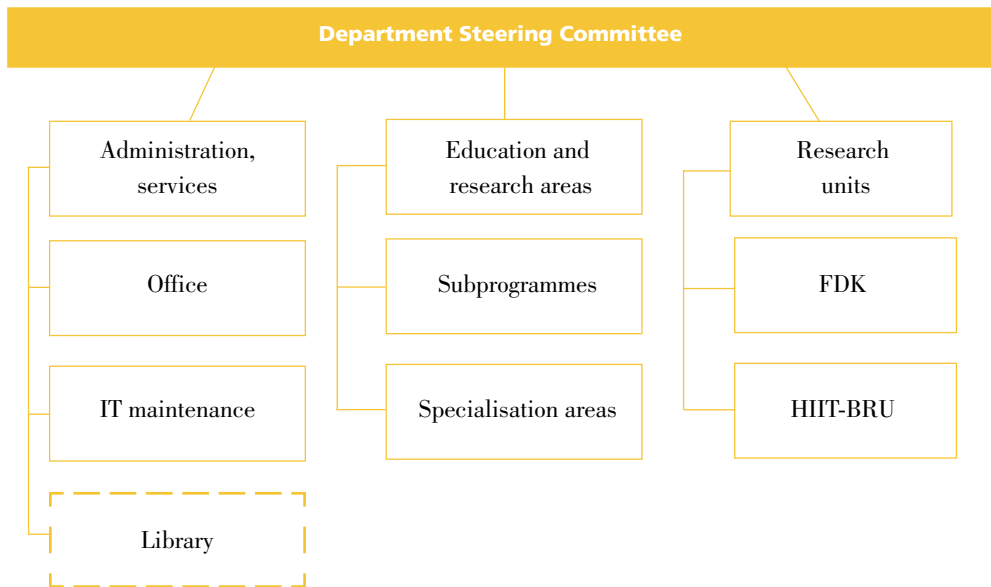
Each organisational unit of the Department is led by a designated person in charge. The Steering Committee was chaired by the head of the Department (Professor Timo Alanko 1st Jan – 31st July 2001, Professor Jukka Paakki 1st Aug – 31st Dec 2001). The office, providing general administrative services in personnel, budgeting and education management, is led by the newly appointed office manager Päivi Karimäki-Suvanto, and the IT maintenance staff was led by IT manager Petri Kutvonen. The educational subprogrammes and specialisation areas were led by the professors responsible for them (algorithms and data structures, and teacher in computer science: Professor Tapio Elomaa; distributed systems and data communication: Professor Timo Alanko; information systems: Professor Hannu Erkiö; intelligent systems: Professor Henry Tirri; software engineering: Professor Inkeri Verkamo; applied computer science and computer mathematician: Professor Matti Mäkelä 1st Jan – 31st July 2001, Professor Eero Hyvönen 1st Aug – 31st Dec 2001). The FDK research unit was led by Academy Professor Esko Ukkonen, and the HIIT-BRU research unit by research director Heikki Mannila.

2.3 THE DEPARTMENT IN NUMBERS

Personnel

Personnel (man-years)	1997	1998	1999	2000	2001
Teachers	46	43	49	59,6	43,5
Professors	8	9	9	9,6	12,1
Senior Assistants	6	4	3	3,4	1,4
Lecturers	9	10	11	12,8	14,2
Assistants	13	8	9	12,9	10,4
Other full-time teachers	7	7	7	4,4	4,2
Part-time teachers	3	5	10	16,9	11,2
Postgraduate students	11	14	11	14,9	17,7
Researchers and research assistants	35	35	37	37,2	39
Administrative personnel	10	11	13	13,0	16,5
Total	108	113	118	125,1	126,6

Man-years according to funding 2001

Other 33 %
Foundations 1 %
Academy 11 %
Budget 55 %

Legend: Budget, Academy, Foundations, Other
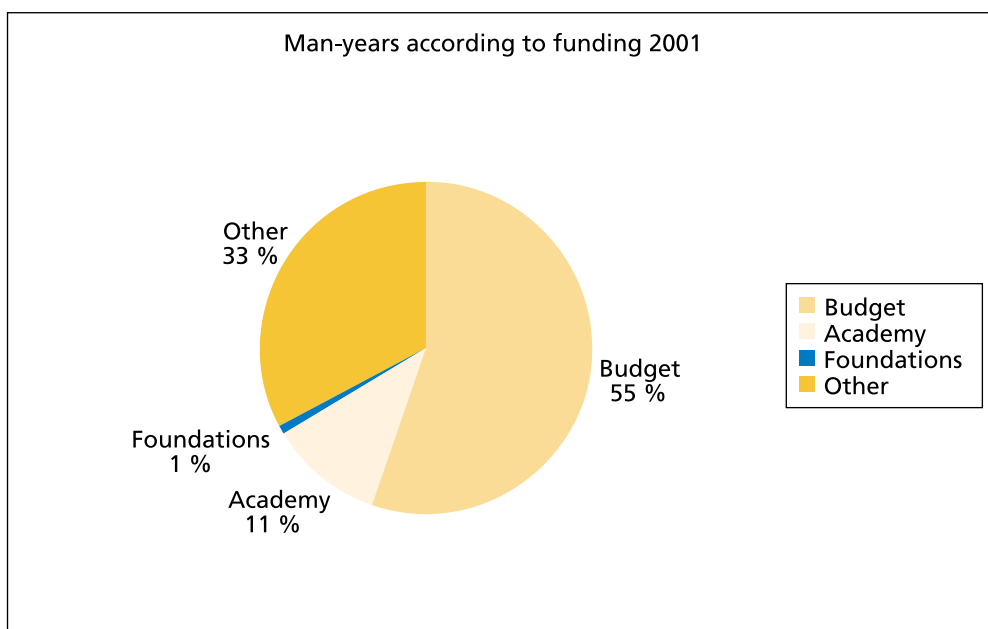
Funding

Funding (million FIM)	1999	2000	2001
Basic funding	11,4	11,4	11,5
Funds carried over from 2000	-	-	5,6
Additional funding	7,2	10,2	12,7
External research funding	11,1	13,2	12,4
Total	29,7	34,8	42,2
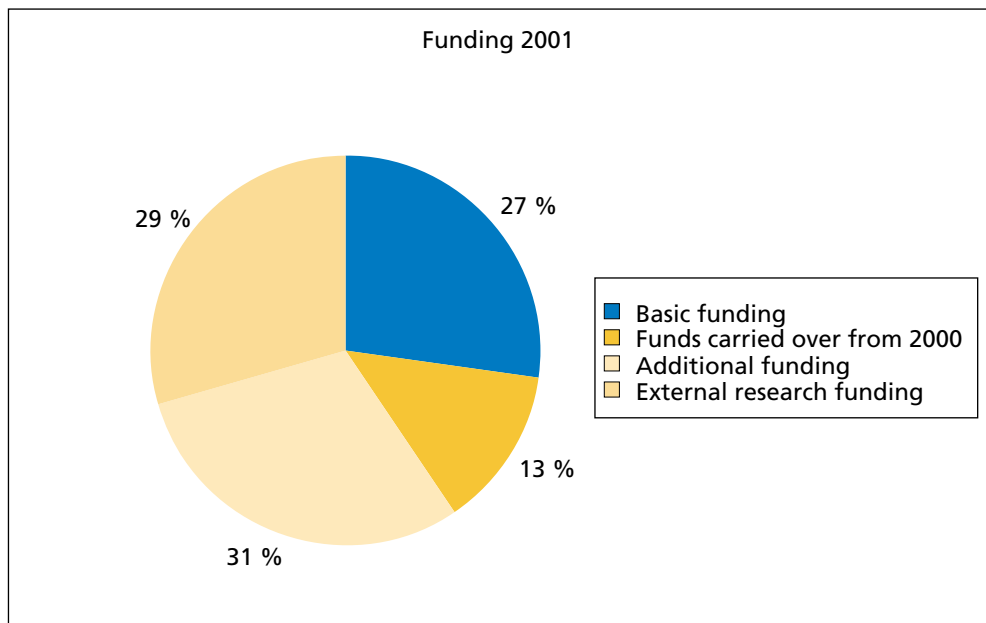
Additional funding for education/Funds for special purposes (million FIM)

	2000	2001
National Information Society Programme	3,9	3,9
Language Technology Programme	-	0,5
Graduate Schools	2,1	2,8
Upgrading studies	2,8	4,9
Other	1,4	0,6
Total	10,2	12,7

External research funding (realized)	1999	2000	2001
Academy of Finland	2,3	2,6	3,4
TEKES & Industry	7,1	9,2	6,9
European Commission	1,4	1	1,3
Other	0,3	0,4	0,8
Total	11,1	13,2	12,4

External research funding 1999 - 2001
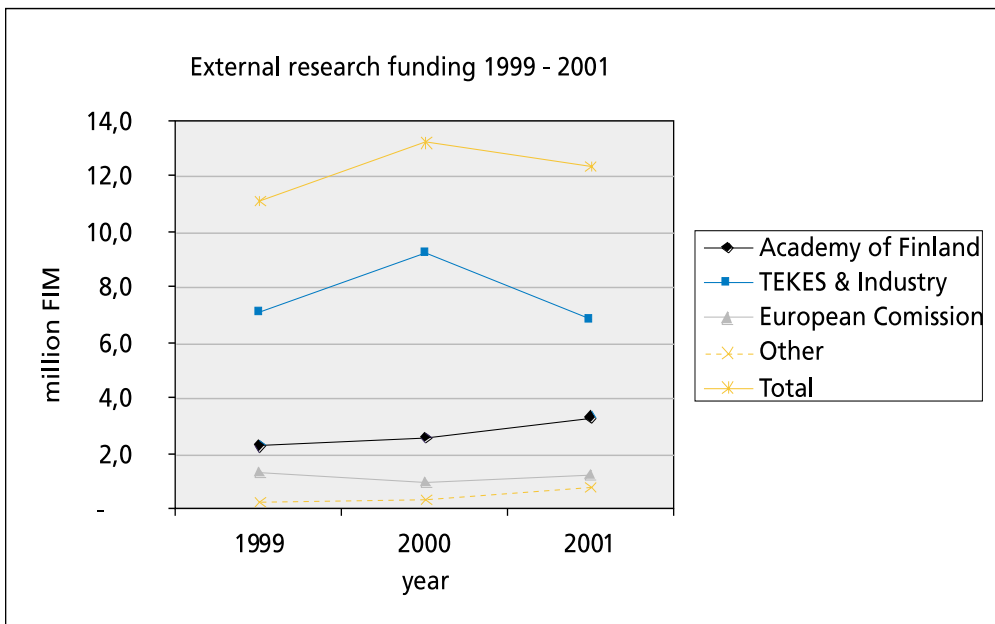
External research funding (realized) 2001 (million FIM)

		%
Academy of Finland	3,4	27
TEKES	3,6	30
Industry	3,3	27
EC	1,3	10
Other national resources	0,8	6
Total	12,4	100

External research funding (realized) in 2001
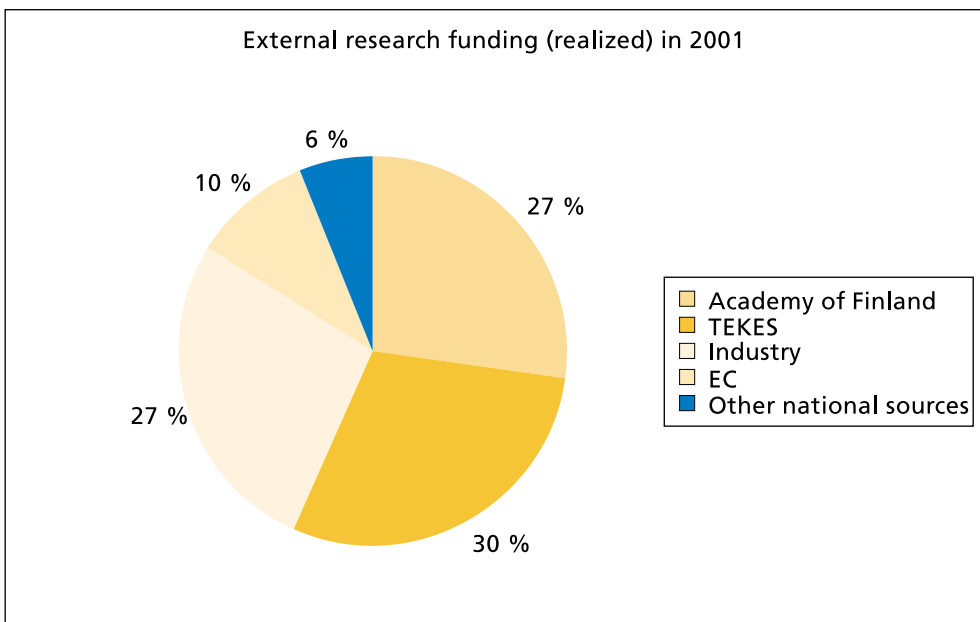
The value of research funds granted in 2001 (income) and the total of granted funds in 2001 (million FIM)

	value	%	total granted	%
Academy of Finland	4.5	33	5.6	35
TEKES & industry	7.8	57	10.2	65
European Commission	1.3	10	0.0	0
Total	13.6	100	15.8	100

Students
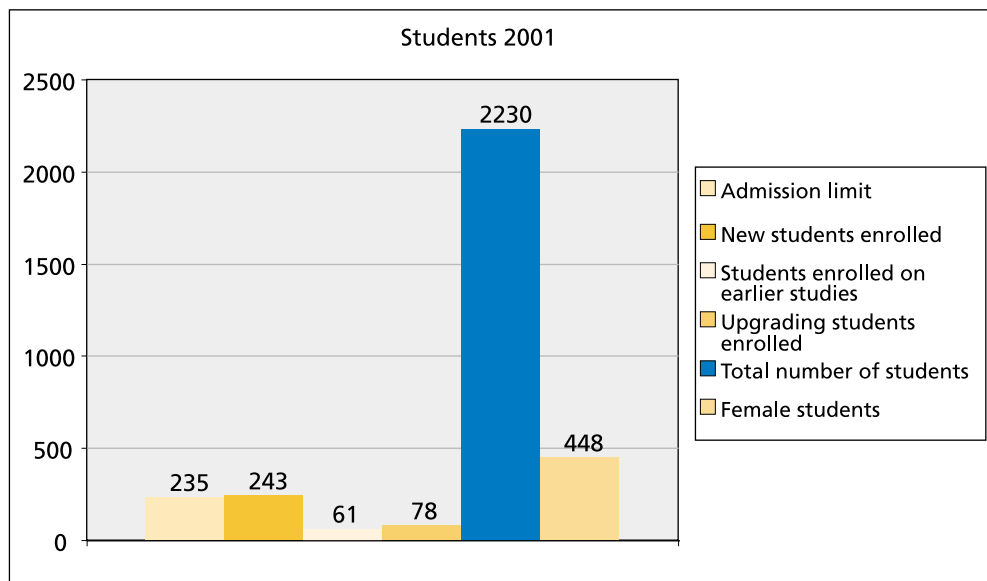
Enrolment (number of students)	1997	1998	1999	2000	2001
Admission limit	260	270	270	270	235
New students enrolled	265	279	258	267	243
Students enrolled on earlier studies	22	55	42	61	61
Upgrading students enrolled	-	-	40	81	78
Total number of students enrolled	1688	1793	1982	2110	2230
Percentage of female students	22	21	22	20	20

Volume of teaching

Teaching	1997	1998	1999	2000	2001
Credits (credit units)	17178	17494	17863	21345	20554
Given instruction (hours)	14480	13739	16529	15050	12962
Teachers (man-years)	46	43	47	60	54
Credit units / teacher years	380	415	401	343	408
Students / teacher years	37	42	43	35	41

Number of degrees

Degrees and theses (number of)	1997	1998	1999	2000	2001
Master's Theses	67	53	63	63	75
MSc degrees	62	48	55	64	61

Postgraduate Education

Postgraduate degrees (number of)	1997	1998	1999	2000	2001
PhLic degrees	4	4	2	5	6
PhD degrees	3	2	3	4	4

Number of Publications

Publications (number of)	1997	1998	1999	2000	2001
Books	3	0	3	0	0
Journal articles	16	10	11	8	6
Peer-reviewed conference and collection articles	70	64	46	49	49
Other publications	60	54	49	50	42
Total	149	128	109	107	97

13
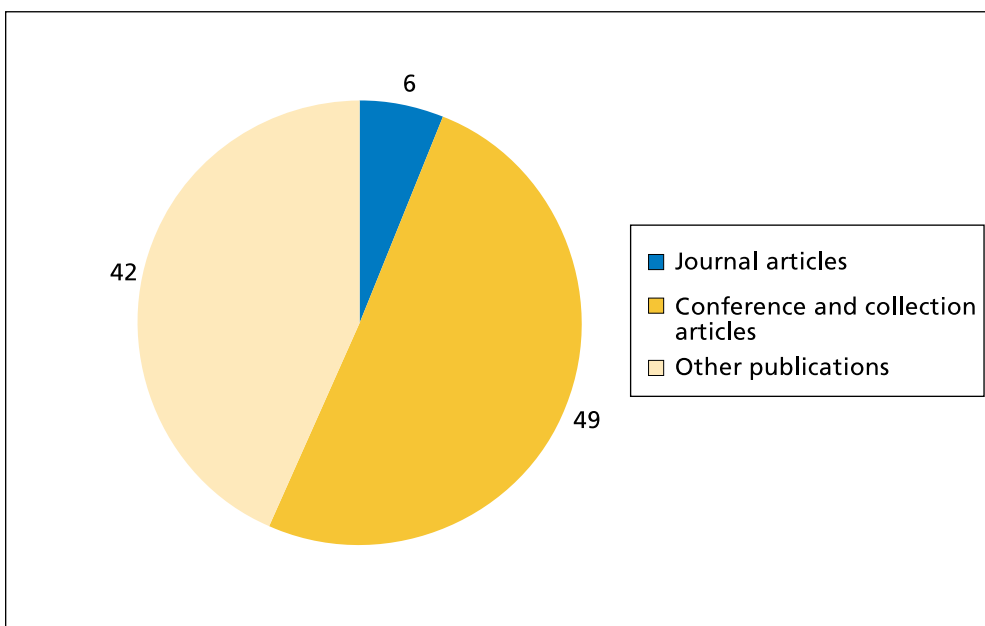
2.4 Nota Bene 2001

Events

In co-operation with the Helsinki Institute for Information Technology, the Finnish Artificial Intelligence Society, XML Finland and Elisa Communications Ltd, the Department arranged the national Semantic Web Kick-Off in Finland seminar on November 2, 2001. In the seminar with more than 200 participants, over 20 experts from the W3C consortium and the Finnish universities and research institutes presented their views on the topical concept of "semantic Web". Professor Eero Hyvönen from the Department acted as the chair of the seminar.

Together with the Department of Computer Science of the Petrozavodsk State University, the Department organised the Annual Finnish Data Processing Week in May 2001. This was the sixth time the seminar was organised.

The seventh Open Day information event for students and guests was arranged on 6th April 2001.

Education

In 2001 a workgroup, led by Lecturer Heikki Lokki, was set up with the objective of strategic development of the Department's teaching and learning methods. The general direction in education will be moving from passive teacher-centred courses to more interactive and student-centred forms of learning.

The wide national enthusiasm for distance and virtual learning was being met at the Department by continuing the development of courses on the web, by establishing a virtual strategy working group (led by Lecturer Teemu Kerola), and by starting its work on the Department's strategy in virtual education. The most serious undertaking in the field was joining the European TUeLIP association for designing and realising e-learning courses commonly targeted to all parties of the association.

The other members of TUeLIP are IBM and the universities of Compeigne, Darmstadt, Delft, Lyon, Milan, and Vienna. The design of the first e-learning courses was started in 2001, and they will be running in 2002.

The problems reported by the students with starting and completing their MSc theses were attacked by radically updating the thesis instructions on the web. In addition, four thesis start-up events were organised, with presentations both from Master students and thesis advisors.

The upgrading programme funded by the Ministry of Education was continued in 2001 and 78 new students were accepted into the programme by the Department. To ease the studies of the upgrading students, most of whom work, some basic courses were given in the evenings and during weekends. The upgrading programme was also reinforced by appointing two University Lecturers, Pirjo Moen and Matti Nykänen, into the programme.

The undertaking in the national language technology education network was strengthened by establishing special courses in the area and by joining the nation-wide graduate school in language technology (KIT). With the appointment of Professor Helena Ahonen-Myka and University Lecturer Greger Lindén, language technology obtained two dedicated research and education positions.

In co-operation with Helsinki University of Technology and Helsinki School of Economics and Business Administration the Department established a joint minor subject in software business. The first students were accepted to study the subject in 2001 and they start their studies in 2002.

The first teacher awards of the Department were granted to Lecturer Arto Wikla and Assistant Anna Pienimäki. The selection was based on proposals by the students and staff and on student feedback.

The scholarships for advancing the MSc thesis were granted to Lauri Eronen, Niklas Laine, Teemu Lampinen, Ilja Ponka, Saara

Huhmarniemi, Marko Ikonen, Lincoyan Kekki, Samuli Siivonen, Jarno Virtanen, Wei Li, and Jaakko Pesonen.

Lecturer in user interfaces, Kjell Lemström Doctor Assistant in computer science, and Jan Lindström Assistant in computer science.

Research

The *From Data to Knowledge* (*FDK*) research unit was appointed a Centre of Excellence in research by the Academy of Finland for the term 2002-2007. The FDK research unit, with one section running at Helsinki University of Technology, is led by Academy Professor Esko Ukkonen.

On the basic funding from the University, the *Basic Research Unit* (*BRU*) of Helsinki Institute for Information Technology (HIIT) was established. Heikki Mannila (Professor at Helsinki University of Technology and the leader of its section in the FDK research unit) was appointed as the director of BRU.

To promote research and development of technologies for the World Wide Web, the Department joined the international W3C consortium. Professor Kimmo Raatikainen was designated as the W3C advisory committee representative of the Department.

Awards and Honours

Eero Hyvönen, Professor, received a "Nokia Educational Award" from Nokia Foundation.

Hannu Erkiö, Professor, was awarded the SVR Knight of the Order of the White Rose of Finland.

15

Personnel

Professor Matti Mäkelä retired on 1st November 2001, becoming the second Professor Emeritus of the Department (the first one being Martti Tienari).

In addition to the nominations listed above, Hannu Toivonen was appointed Professor in information systems, Sari A. Laakso University

3. Research and Education Areas

3.1 COMPUTER SCIENCE

3.1.1 Algorithms

The specialisation area studies the algorithmic and theoretical aspects of formulating and solving computational problems in different application areas. The focus is on discovering efficient solutions for specific problems. The interaction between theory and practice is also of central interest. The application of well-motivated algorithms to real-world problems is within the scope of the area.

The curriculum of the specialisation area covers the core of computer science. The central obligatory courses are: Data Structures and Algorithms, Theory of Computation, and Design and Analysis of Algorithms. Advanced courses and seminars are organized on diverse modern research topics in algorithmics such as String Processing Algorithms, Machine Learning, Computational Geometry, Computational Biology and Bioinformatics, Computer Graphics, Signal Processing, Cryptography etc. Mathematics beyond the obligatory curriculum is suggested to all students of the specialisation area. Other suggested minor subjects include, e.g., Physics, Theoretical Physics and Statistics.

On the research front the current main focus areas are combinatorial pattern matching and its applications in computational biology, theory and practice of machine learning with interest in the application of the methods to data mining, data structures for biological databases, and computational geometry. An analytical approach to diverse problems is the common factor in the many different research projects within the specialisation area.

The contact persons in Algorithms are Professor Tapio Elomaa and Academy Professor Esko Ukkonen.

Education (in Finnish): http://www.cs.helsinki.fi/u/elomaa/opetus/algoritmit/
Research (in English): http://www.cs.helsinki.fi/research/fdk/

3.1.2 Intelligent Systems

The specialisation area of Intelligent Systems studies issues related to the design and analysis of computational methods for adaptive and intelligent systems. The area covers several fields in computer science such as artificial intelligence, computational intelligence, artificial life and search methods. Future information technology will include more and more adaptive components both in hardware and software and therefore many of the methods studied are core components for future systems.

The curriculum of the Intelligent Systems area contains a wide variety of special courses, which typically require as prerequisites good programming and problem solving skills and a solid analytical background. Studies of automating intelligent behaviour are by necessity drawing ideas from many sciences, and thus the studies in many cases are multidisciplinary in nature. In addition to the methodological courses in the "Three Concepts" series, other suitable courses can be selected either from more applied topics (robotics, string matching, data mining) or from the general introductions such as artificial intelligence and machine learning. A research seminar in Intelligent Systems is intended for postgraduate level studies and multidisciplinary studies in cognitive science, neurocomputing (in cooperation with Helsinki University of Technology) and theoretical philosophy are strongly encouraged.

The research in the Intelligent Systems area is focused on issues related to deep foundational issues in modelling and learning , and their applications to various fields from engineering to social sciences and medicine. Most of the research is associated with the Complex

Systems Computation Group (CoSCo), which studies theoretical and applied topics in probabilistic and information-theoretical modelling and stochastic search methods. Current ongoing projects have applications e.g. in user profiling and personalisation, adaptive learning environments, autonomous satellite diagnosis systems and telecommunications.

The contact person in Intelligent Systems is Professor Henry Tirri.

Education (in Finnish): http://www.cs.helsinki.fi/u/tirri/alykaslinja.html
Research (in English): http://cosco.hiit.fi

3.1.3 Software Engineering

The specialisation area of Software Engineering studies methods, techniques, and tools to systematically develop large and complex software systems of high quality. The area covers not only technical issues, but also managerial and co-operative aspects of team and project work as well as the business dimension of software products.

The curriculum of the specialisation area is rather broad, containing courses and seminars in all of the dimensions of software engineering mentioned above. There are four obligatory basic courses in the curriculum: Software Engineering, Software Engineering Lab, Software Architectures, and Software Processes and Quality. On top of these, special courses and seminars can be selected on topics such as programming languages, compilers, software testing, software performance, user interfaces, and software project management. The importance of the software business perspective has been taken into account by offering a minor subject on this topic, in co-operation with Helsinki University of Technology and Helsinki School of Economics and Business Administration.

The research in software engineering is focused on one of the cornerstones of software

quality, the software design phase and the related technical artefact of software architectures. Currently ongoing research projects develop techniques and tools to early measurement and assurance of the quality of a software system in its design phase, before entering the implementation and testing phases. Modern object-oriented techniques, most notably design patterns, application frameworks and the UML modelling language, play an important role in the techniques and tools developed in the research projects.

The contact persons in Software Engineering are Professor Jukka Paakki and Professor Inkeri Verkamo.

Education (in Finnish): http://www.cs.helsinki.fi/u/verkamo/selinja.html
Research (in English): http://www.cs.helsinki.fi/research/roosa/

3.1.4 Distributed Systems and Data Communication

The specialisation area of Distributed Systems and Data Communication studies how systems can be divided into independently working parallel parts, and how these parts communicate with each other. The research concentrates on examining the basic components and the protocols of such systems. The goals of the group are twofold. On one hand, tools are developed for analysing and modelling systems based on theory. On the other hand, more powerful services are developed for the application platform based on operating systems, data communication, distribution algorithms and effective information management solutions.

The curriculum of the specialisation area is broad, containing courses and seminars on all the research dimensions. Obligatory courses in the curriculum are Operating Systems I—II, Data Communication I—II, Introduction to Specification and Verification, Computer Orga-

nization I—II, Concurrent Systems, and Distributed Systems. Additional special courses and seminars can be selected on varying topics in mobile computing, agent technology, security, formal methods, and electronic commerce.

The research in Distributed Systems and Data Communication has four focus areas:

The mobile and nomadic computing group develops wireless data communication systems and applications. The basic requirements for such systems include real-time processing, capability for multimedia transfer, applicability of changing conditions on data transfer and adaptability to heterogeneous environments.

The open distributed software systems group develops software architectures and services for inter-organizational systems. Openness requires capabilities for automatic negotiation of new cooperation networks amongst independently developed and autonomously managed systems.

The operating system enhancements group currently focuses on real-time systems, covering fast data storage, fast data transfer and high availability requirements. In real-time databases optimistic strategies, memory-based storage, and replication-based fault-tolerance are under experimentation.

Formal specification and verification of distributed systems is studied in the modelling of concurrent systems group. The theoretical results of the group are based on process algebras and temporal logic. The results are applied to software engineering tools.

The contact persons in Distributed Systems and Data Communication are Professor Timo Alanko and Professor Kimmo Raatikainen

Education (in Finnish): http://www.cs.helsinki.fi/u/alanko/ht_linja/ Research (in English): http://www.cs.helsinki.fi/u/alanko/nodes.html

3.1.5 Information Systems

The specialisation area of Information Systems studies methods, techniques, and tools needed in manipulating large volumes of data, often implemented as a database. The central areas of the respective ACM CR category are covered in the education: database management, information retrieval, applications and interfaces. Especially data mining as well as methods of text and document management in current digital media are emphasised in our research. The last area is more comprehensively described in part 3.7.3 (Language Technology).

The core of the curriculum of Information Systems contains obligatory courses on database management and database applications: Introduction to Application Design, Introduction to Databases, Database Application Project, Database Management, Database Modelling, and Database Structures and Algorithms. Besides these, special courses are offered on data warehousing, data mining, information extraction, spatial information systems, structured documents, information retrieval, and user interfaces. Specific subareas are covered mainly by seminars with topics varying yearly. In the Software Engineering Lab, design and implementation of a database system or an interface are common topics.

The research in information systems is strongly focused on data mining and document management (language technology). Techniques of gathering and filtering information from heterogeneous sources, and reuse and analysis of information to form suitable knowledge bases are essential research topics. Methods for at least semiautomatic management of information, often using XML, are being developed. In applied research, problems of gene technology, ecology, telecommunication, and publication business are dealt with. Research in the database area is focused on recovery and concurrency control. Some development work has also been made in preparing environments for database teaching.

18

The contact persons in Information Systems are Professors Hannu Toivonen, Helena Ahonen-Myka, and Seppo Sippu.

Education (in Finnish):http://www.cs.helsinki.fi/u/erkio/info_linja.html
Research (in English): http://www.cs.helsinki.fi/research/doremi/
http://www.cs.helsinki.fi/research/gist/

3.2 TEACHER IN COMPUTER SCIENCE

The subprogramme educates qualified computer science teachers for all levels of the educational system (lower and upper secondary schools, colleges and polytechnic institutes) and for business training. The curriculum contains a pedagogical component (35 cu), which lessens the required advanced computer science studies. The studies in pedagogy are carried out alongside the studies in the subjects to be taught. A course on Computer Uses in Education is obligatory to the students in the teacher subprogramme. Lecture courses, seminars and the subject for the Master's thesis can be chosen from another subprogramme. Mathematics is the most popular minor subject.

The contact person in Teacher in Computer Science is Professor Tapio Elomaa.

Education (in Finnish): http://www.cs.helsinki.fi/u/elomaa/opetus/opettaja

3.3 APPLIED COMPUTER SCIENCE

The subprogramme Applied Computer Science is meant for students who want to specialise in some application area and study that more than is possible in the other subprogrammes. Every student has an individual study programme. The research activities are pursued in several of the other research divisions, e.g. within the algorith-

mics, machine learning, biocomputing, computational linguistics and data mining groups.

Professor Eero Hyvönen is in charge of the subprogramme.

Education (in Finnish): http://www.cs.helsinki.fi/u/eahyvone/sv/svsov.html

3.4 COMPUTER MATHEMATICIAN

The subprogramme Computer Mathematician is a joint programme with the Department of Mathematics. The students must include computer science and mathematics to the same amount in their degrees. Every student has an individual study programme. The mathematically oriented research problems for the students come from other research areas.

Professor Eero Hyvönen is in charge of the subprogramme.

Education (in Finnish): http://www.cs.helsinki.fi/u/eahyvone/sv/svmat.html

19

3.5 SOFTWARE BUSINESS

The Department of Computer Science, Helsinki University of Technology and the Helsinki School of Economics offer their major subject students the possibility to take software business as a minor subject. This minor subject module consists of general courses in software engineering, software production, business skills, marketing and computer legislation, as well as a project in the field. Of the students at the department, 15 were chosen for this minor subject module in 2001.

Professor Eero Hyvönen is in charge of the minor subject module.

More information (in Finnish): http://www.cs.helsinki.fi/u/eahyvone/ohli/

3.6 GRADUATE SCHOOLS

3.6.1 HeCSE

The Helsinki Graduate School in Computer Science and Engineering is a postgraduate programme jointly offered by the Helsinki University of Technology (HUT) and the University of Helsinki (UH). It is mainly intended for students wishing to pursue their graduate studies full time, aiming for a PhD degree in four years. For these students, HeCSE offers the possibility of financial support and some other benefits such as special courses.

HeCSE incorporates the following laboratories and departments: Laboratory of Information Processing Science (HUT), Laboratory of Telecommunications Software and Multimedia (HUT), Laboratory for Theoretical Computer Science (HUT), Neural Networks Research Center (HUT) and Department of Computer Science (UH). The collaboration with Helsinki Institute for Information Technology (HIIT) has started recently.

Students can choose any area represented by the professors as their specialisation area. However, the main research areas in HeCSE are software systems, software engineering, telecommunication software and distributed systems, as well as learning and intelligent systems. There are about 50 active students in HeCSE at the moment. HeCSE can finance 28 students, 10 from the University of Helsinki and, 18 from the Helsinki University of Technology. The other students are financed by projects or teaching.

More information (in English): http:// www.cs.helsinki.fi/hecse/

3.6.2 ComBi

The Graduate School in Computational Biology, Bioinformatics, and Biometry (ComBi) is a postgraduate program jointly offered by the University of Helsinki, the University of Turku and the University of Tampere. The Department of Computer Science at the University of Helsinki coordinates the school.

The research goal of ComBi is to develop computational, mathematical, and statistical methods for biological sciences. To that end, ComBi will educate PhDs with high-quality methodological expertise. In their thesis work the students are expected to apply this expertise to computational, data analysis or modelling problems in biology or in some related field. The thesis projects are carried out in cooperation with one or more research groups in the application area (such as biochemistry, molecular biology, microbiology, ecology, evolution research and systematics, geography, public health).

Computational biology is a new field of research which develops methods and software implementations for computational problems in molecular biology, biotechnology, and genetics. Bioinformatics refers to the development and use of (molecular) biological databases. Biometry investigates statistical modelling problems related for example to genetic mapping, 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 of ecological and genetic phenomena.

More information (in English): http:// www.cs.helsinki.fi/combi/

3.6.3 KIT Graduate School

The KIT graduate school is a multidisciplinary, nation-wide graduate school for human language technology-related PhD studies. The graduate school is part of the emerging Nordic cooperation in the field of language technology researcher training. The school starts at the beginning of year 2002.

The contact person is Professor Helena Ahonen-Myka.

3.7 OTHER EDUCATION

3.7.1 Upgrading Studies

A temporary programme for upgrading studies in computer science started in the autumn of 1999. It aims at teaching IT knowledge to persons with a degree in another subject, to persons who have no degree but a certain amount of computer science studies from before, or to students who want to change their major subject into computer science. In total, 78 participants were admitted in 2001 based on a separate selection process. The programme continues in the year 2002, when the last group of students will be admitted to the programme.

The students are divided into tutor groups. Each group has a supervisor who guides the students to set up individual study plans. The students usually start their studies with second year courses according to the model study schedule. There is also a basic course that aims at refreshing basic knowledge in programming and software implementation in a modern environment. Additional courses and study circles are organized in the evenings and during the weekends, which provides more flexibility in scheduling the studies.

3.7.2 The Co-operation with Upper Secondary Schools

The Department of Computer Science started a trial cooperation with upper secondary schools in the spring of 1999, with six schools in Helsinki joining in. At the moment, there are 12 schools participating in the programme.

The pupils of the cooperating schools are able to participate in the first major courses in computer science after being accepted to the programme through an entrance exam. In spring 2001, 25 pupils out of 34 applicants were accepted.

The programme gives the Department a chance to develop the academic skills of the upper secondary pupils before they start their actual academic studies. It is also a way of recruiting new undergraduates and motivating them.

The goal of this cooperation is to even out the difference between being an undergraduate and being a pupil, and to make the transition from school to university studies easier. From the point of view of the students, the main goals are to clarify study interests and career goals, to motivate upper secondary studies, and to take responsibility for their own studies. For the schools, the objective is to vary and enhance the upper secondary education, promote the profile of the schools, to make use of the expertise of scientists, and to increase the cooperation between school and university teachers.

21

3.7.3 The Language Technology Network

The term human language technology or natural language processing means using linguistics for information technology applications processing natural languages. Some fields of application are the use of spoken language for computer user interfaces, computer-aided language education, writing accessories (spell-checkers, grammar checking, etc), data mining and document management, automatic translation, speech recognition and production, interactive network applications, mobile communications technology and electronic dictionaries.

The Department of Computer Science is a member of the nation-wide language technology

education network (Kieliteknologian opetus, KIT). This network enables extensive and multidisciplined language technology studies in several universities. Students may take language technology as a large minor subject (with one other minor) in the Applied Computer Science subprogramme, or as a normal minor in the Computer Science subprogramme.

The contact person is Professor Helena Ahonen-Myka.

Education (in Finnish): http://www.cs.helsinki.fi/u/hahonen/ktekno/
Research: http://www.cs.helsinki.fi/research/doremi/

4. Research

4.1 RESEARCH UNITS

4.1.1 FDK

The From Data to Knowledge Research Unit (the FDK unit, for short; www.cs.helsinki.fi/research/fdk) develops computing methods for forming useful knowledge from large masses of data. The unit is multidisciplinary, combining in its research groups expertise in algorithmics, statistical methods and application fields such as bioinformatics and human language processing. The unit was appointed a Centre of Excellence of the Academy of Finland for a six year period starting from the beginning of 2002.

The FDK unit is a cooperation between the University of Helsinki and the Helsinki University of Technology (HUT). The main site of activities is the Department of Computer Science of the University of Helsinki. Academy Professor Esko Ukkonen is the director of the unit. The personnel of the unit consists of about forty researchers and post graduate students. At the end of 2001 the following researchers with a doctorate were members of the unit: Professor Esko Ukkonen, Professor Heikki Mannila (the Basic Research Unit of HIIT), Professor Tapio Elomaa, Professor Helena Ahonen-Myka, Professor Jaakko Hollmen (HUT), Professor Hannu Toivonen, Kimmo Fredriksson, Juha Kärkkäinen, Kjell Lemström, Greger Lindén, Juho Rousu, Marko Salmenkivi.

The core competence area of the unit is algorithmics for data analysis, within which two special fields are emphasized. The first is the so-called combinatorial pattern recognition and string matching. It considers pattern recognition and matching in various discrete structures. The other main pillar is machine learning and, in particular, its application-oriented branch known as data mining. The goal is to find interesting and useful patterns, dependencies, and other hidden structures in the data. By means of these findings, different abstract models will be built for the phenomena described by the data. When developing data mining algorithms, one has to expand the traditional tool box of algorithmics: the concepts and methods from statistics and stochastics are also needed.

The unit emphasizes in its activities the interaction between theory development and practical applications. Our goal is to find research problems, whose conceptual basis and solution algorithm have a wider application potential. When developing algorithms we are not content with just tailoring some general-purpose approaches (as, for example, the so-called genetic algorithms) for the problem in hand. Rather, the goal is to develop and analyse original algorithms using the wide technical arsenal of algorithmics.

The unit functions as four main projects, which are closely connected and partially overlapping. The same persons are active in several projects. This facilitates internal communication and the utilisation of expert knowledge.

23
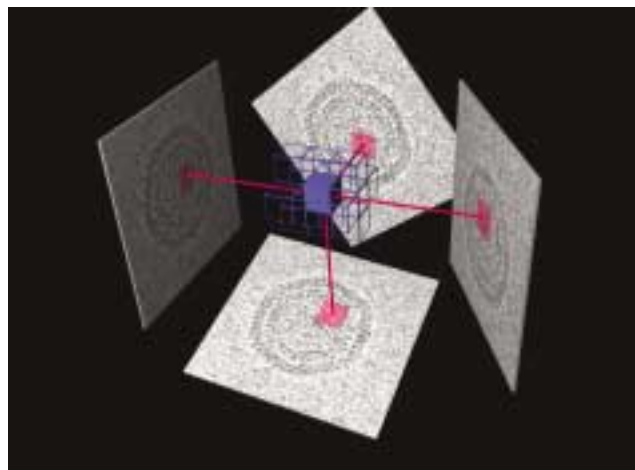
The theme of the main endeavour, Project I, is data mining and machine learning. The project develops original concepts and algorithms to strengthen a core area of the unit. We aim at results in theoretical basic research. The relevance of the results is tested in various applications. Text databases and document collections as well as event sequences in telecommunication networks are examples of the data we use. Information filtering from the Internet and other human language technology belong to the area of this project as well as using machine learning in image analysis.

Project II specialises in computational biology by studying the methods for medical genetics and for analysing data on genomics, proteomics, and metabolics. The spectrum of collaborators is wide ranging from UCLA to the European Bioinformatics Institute and to several top-class groups in Finland. We develop computational methods for example for locating the genetic basis of multifactorial diseases as well as for synthesizing gene regulatory networks from data. The intensive progress in this field is boosted by the growing number of entirely sequenced genomes and the rapid expansion of the new high-throughput measurement techniques that are able to follow the behaviour of the whole genome. The new era of 'systems biology' is about to start, with computation in a central role.

Project III deals with combinatorial pattern recognition and information retrieval. Again we combine algorithm theory and applications. Central research questions include approximate pattern matching, efficient index structures, and the synthesis of patterns from data. Applications are several. One subproject considers the retrieval and analysis of symbolically encoded music, for example.

Project IV focuses on computational structural biology. This area is somewhat different
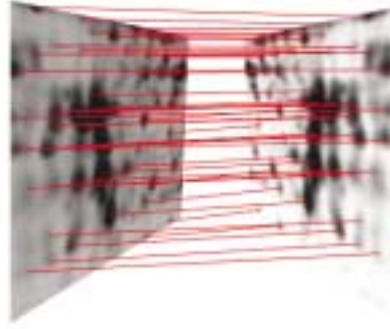
from the DNA sequence-driven mainstream computational biology. The main task is to construct three-dimensional computer models for biological macromolecular complexes, from electron micrographs, for example. Technically, one has to generalise the standard computer tomography here. Comparing the reconstructed models is also a challenging task, offering another application for combinatorial pattern-matching methods. In this project we collaborate with Professor Dennis Bamford's Structural Biology Group, located at the Division of Biotechnology at Helsinki University.

In addition to the basic research and doctoral education, the FDK unit also wants to serve as an algorithm 'atelier' which develops computational solutions to new problems in different fields. We are searching for new partners who could pose open computational problems in the forefront of the research.

4.1.2 The Basic Research Unit at HIIT

The Helsinki Institute for Information Technology (HIIT) is a joint research unit of the University of Helsinki and Helsinki University of Technology. The unit aims at significantly enhancing the strategic research in the field of computer science, while improving the co-operation between the University of Helsinki and the Helsinki University of Technology. HIIT consists of two units: the Advanced Research

Unit (headed by Martti Mäntylä), which works in co-operation with the industries and functions primarily at the University of Technology, and the new Basic Research Unit (headed by Heikki Mannila) that started operations at the beginning of the year 2002. The Basic Research Unit functions primarily at the Computer Science Department in Vallila.

The Basic Research Unit (BRU) carries out first-rate basic research in the field of computer science, in close co-operation with end-users in other disciplines and in the industries. BRU consists of research groups with national and international senior researchers, researchers and post-graduates. Their basic funding is part of the long-term funding of BRU, but the groups are active in procuring outside funding, as well. The activities of BRU are arranged around certain blanket themes. At first, the research will be focused on two main themes: proactive computing (ubiquitous computing, calm/invisible computing) as well as the basic problems and applications of data analysis and data mining.

In the field of proactive computing, the activities at BRU will focus on a few areas, as follows.

1. The basic problems with linking computation and mobile communications. In this field, we will examine the rules of distributed computing. The purpose is to examine computing that is energy-aware, models of adiabatic computing, etc.

2. Adaptive user interfaces. All (successful) human communication must be set in context, i.e. what is communicated and how it depends on the situation. Data processing software is typically free of context, and cannot adapt to the user's situation at the moment of use, or to user history, for example. Research is carried out on methods that can enhance the usability of the software.

BRU is involved in developing the theories and applications of data analysis and data mining in close co-operation with different end-users. During its first year, the unit focuses on the following research themes, among others.

1. Gene expressions, problems concerning the regulation and inheritance of expressions. The researchers will develop methods for analysing gene expression data, which can be linked with other biological material.

2. The theoretical analysis of DNA sequences. The exact function of the non-coding areas in the human genome are unknown so far. This research develops methods with which divergently structured areas can be quantitatively identified in a DNA string.

3. Methods of analysis for large discrete data arrays. Many applications allow the user to gather multidimensional data arrays, where each observation value is typically a small integer. This kind of material may be approximated effeciently by using the method of frequent sets.

In its entirety, the unit will focus on 4-6 large themes that will overlap each other and on the research carried out at institutions for other disciplines. The Basic Research Unit works in co-operation with research groups at the department, especially with the group From Data to Knowledge, a Centre of Excellence led by Academy Professor Esko Ukkonen. The staff of BRU act as teachers at both the Department of Computer Science and Engineering at the Helsinki University of Technology and at the Department of Computer Science at the University of Helsinki. In addition, intensive courses are arranged in co-operation with the graduate schools in the field.

Please see http://www.cs.helsinki.fi/hiit_bru/ and http://www.hiit.fi for more information. Queries may be sent to heikki.mannila@cs.helsinki.fi.

25

4.2 RESEARCH GROUPS

4.2.1 CoSCo

The CoSCo research group investigates computational issues related to complex systems, focusing on prediction and model selection tasks. The research areas addressed include stochastic modelling and data analysis - especially with Bayesian networks and related probabilistic model families, such as finite mixture models and Bayesian multinets – information theoretical approaches to inference (MDL/MML), theoretical and empirical validation of on-line prediction algorithms, case-based reasoning (CBR), and stochastic optimisation algorithms, such as simulated annealing and genetic algorithms. The work has both a strong basic research component, being at the intersection of computer science, information theory and mathematical statistics, and an applied component where the results are applied in multiple disciplines from social sciences, criminology, ecology and medicine to industrial engineering. Recent focus areas in the applied research include personalisation for the Internet, diagnostics for space satellites, and modelling for location-aware services.

In general the CoSCo team benefits from a rare combination of theoretical competence with top-level programming skills. As a concrete demonstration of this, CoSCo has recently opened the unique B-Course service (http://b-course.cs.Helsinki.fi), a sophisticated tool for state-of-the-art data analysis on the web. For performing the extensive empirical tests required for evaluating the theoretical methods developed, the group has developed software that allows the researchers to use several dozens of Linux-workstations as a single "virtual supercomputer", which has made it possible to study empirically many interesting exponential-time problems.

In two recent (1997,1999) evaluations in National Technology Agency (TEKES) research

programmes, the group got excellent remarks both for theoretical and more applied work. Two more recent acknowledgements of the high level of the scientific work done in the group were 2nd place in CoIL Challenge 2000, a large international prediction competition with 147 contestants consisting of academic and commercial research groups all over the world, and an honorable mention at the Knowledge Discovery and Databases prediction competition (KDD Cup'01) for being 2nd out of 114 participating international groups in the pharmaceutical prediction task concerning molecular bioactivity. The CoSCo group has also established excellent research contacts to all major probabilistic modelling research groups in the world, with frequent visiting scientists.

CoSCo is affiliated with Helsinki Institute for Information Technology's (HIIT) Advanced Research Unit (http://www.hiit.fi).

Contact person: Professor Henry Tirri
Research group homepage: http://cosco.hiit.fi

Selected publications:

P.Myllymäki, T.Silander, H.Tirri, P.Uronen, Bayesian Data Mining on the Web with B-Course. Pp. 626-629 in Proceedings of the 2001 IEEE International Conference on Data Mining, edited by N. Cercone, T.Y. Lin and X. Wu. IEEE Computer Society Press, 2001.

P.Kontkanen, P.Myllymäki, H.Tirri, Classifier Learning with Supervised Marginal Likelihood. Pp. 277- 284 in Proceedings of the 17th International Conference on Uncertainty in Artificial Intelligence (UAI'01), edited by J.Breese and D.Koller. Morgan Kaufmann Publishers, 2001.

M.Paakko, N.Holsti, P.Myllymäki, H.Tirri, Bayesian Networks for Advanced FDIR. Pp. 311-318 in Proceedings of the ESA Work-

Vision for enhanced middleware
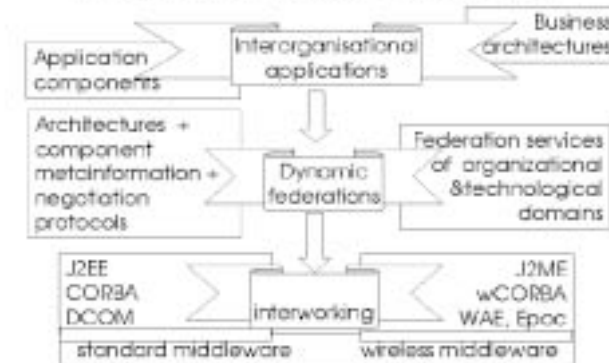
shop on On-Board Autonomy (Noordwijk, The Netherlands, October 2001). WPP-191, European Space Agency, 2001.

J.Kurhila, M.Miettinen, M.Niemivirta, P.Nokelainen, T.Silander, H.Tirri, Bayesian Modeling in an Adaptive On-Line Questionnaire for Education and Educational Research. Pp. 194-201 in Proceedings of the 10th International PEG2001 Conference, Tampere, June 2001.

4.2.2 Networking in Open Distributed Environments — NODES

The NODES group has been formed to stimulate and coordinate research on distributed and networked systems and applications. The group's scope of interests includes concurrency, data communication software, and distributed systems.

The current focus is on the design and development of infrastructure for mobile computing, on performance evaluation of distributed and networked systems, and on open distributed software architectures and services. In 2001, there were four main research group active:

The mobile and nomadic computing group

works with wireless data communication systems and applications. The work started in the mid-90s, and the original goal was to create convenient access from a portable computer through the wireless WAN (e.g., GSM) to the Internet services. Subsequent research has diverted from this platform in various directions. A group of projects examines data communication protocols in heterogeneous networks (IIP-Mobile, wCORBA, MIND). One project aims at developments in mobile applications (CRUM-PET).

The open distributed software systems group develops software architectures and services for inter-organisational systems. Openness requires capabilities for automatic negotiation of new cooperation networks amongst independently developed and autonomously managed systems. The results of this research are both conceptual and constructive (PILARCOS).

In the operating system enhancements group there were two active projects. One project was developing a fault-tolerant real-time database system; special attention was given to concurrency control and transaction scheduling (RTD-pilot). The second project investigates the capabilities of Linux in environments where high availability and real-time characteristics are required (nsrtLinux).

Formal specification and verification of

27

distributed systems is studied in the Modelling of concurrent systems group. The theoretical results of the group are based on process algebras and temporal logic. The results are applied to software engineering tools.

Project list:

Improving Internet protocols for mobile computing (IIP-Mobile)
Wireless CORBA (wCORBA)
Mobile IP based network developments (MIND)
Tourism-related value-added services for nomadic users (CRUMPET)
Production and integration of large component systems (PILARCOS)
Fault-tolerant real-time database systems (RTD-pilot)
Non-stop real-time Linux (nsrtLinux)
Modelling of concurrency (MOCO)

28

**Contact person: Professor Timo Alanko
Research group homepage: http://
www.cs.helsinki.fi/u/alanko/
nodes.html**

Selected publications:

M. Kojo, A. Gurtov, J. Manner, P. Sarolahti, T. Alanko, K. Raatikainen: Seawind: A Wireless Network Emulator. In Proc. of the 11th GI/ITG Conference on Measuring, Modelling and Evaluation of Computer and Communication Systems (MMB 2001), RWTH Aachen, Germany, September 11-14, 2001.

J. Manner, M. Kojo, A. Laukkanen, M. Liljeberg, T. Suihko, K. Raatikainen: Exploitation of Link Layer QoS Mechanisms in IP QoS Architectures, ITCom 2001, Denver, Colorado, 20 - 24 August, 2001. Published in "Quality of Service over Next-Generation Data Networks", Proceedings of SPIE Vol. 4524, pp. 273-283.

J. Manner, K. Raatikainen: Extended Quality-of-Service for Mobile Networks, IEEE/IFIP Ninth International Workshop on Quality of Service (IWQoS 2001), Karlsruhe, Germany, June 6 - 8, 2001. Published in the Springer LNCS Series, Vol. 2092, pp. 275-280. (© Springer-Verlag)

J. Manner, M. Kojo, A. Laukkanen, K. Raatikainen: A QoS Architecture Framework for Mobile Networks, International Conference on Third Generation Wireless and Beyond (3Gwireless'01), San Francisco, USA, May 30 - June 2, 2001.
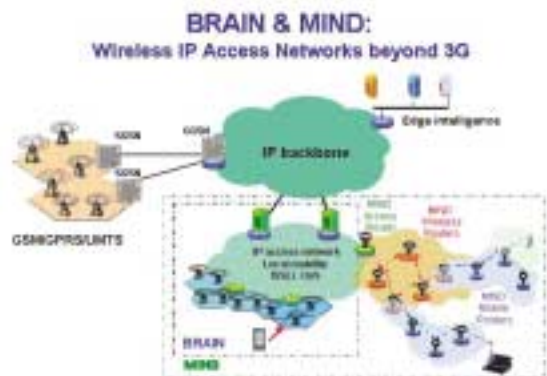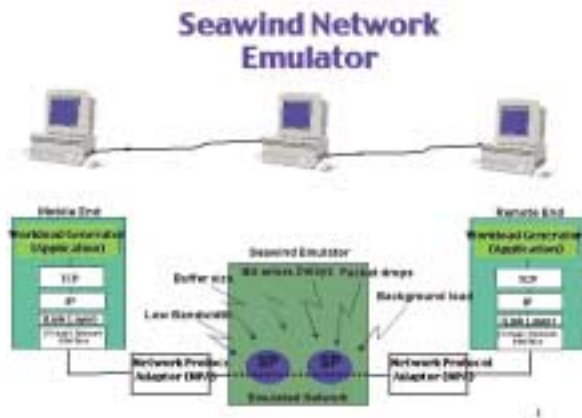
J. Border, M. Kojo, G. Montenegro, J. Griner, Z. Shelby: Performance Enhancing Proxies Intended to Mitigate Link-Relate Degradations, RFC

3135 (Informational), Internet Society, June 2001.

S. Dawkins, G. Montenegro, M. Kojo, V. Magret: End-to-end Performance Implications of Slow Links, RFC 3150 (Best Current Practice No: 48), Internet Society, July, 2001.

S. Dawkins, G. Montenegro, M. Kojo, V. Magret, N. Vaidya: End-to-end Performance Implications of Links with Errors. RFC 3155 (Best Current Practice No: 50), Internet Society, August 2001.
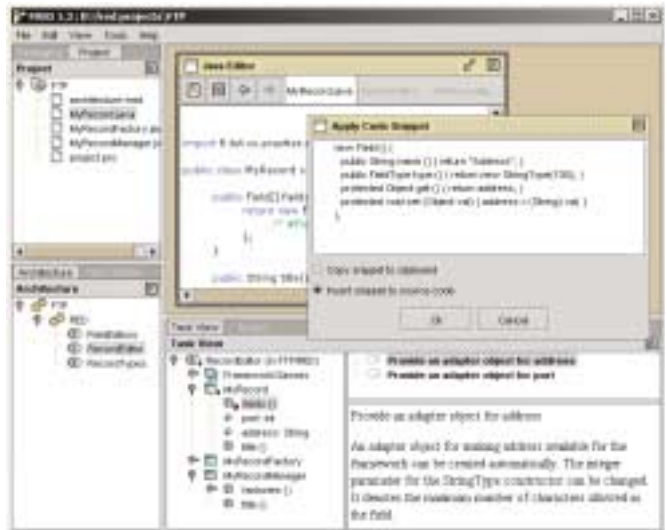
4.2.3 Research Group on Object-Oriented Software Architectures (ROOSA)

The ROOSA research group at the Department of Computer Science studies software architectures, and especially their object-oriented paradigms. The research focuses on conceptual issues as well as on constructive methods. The methods and tools developed provide support for systematically forward-engineering software in an architecture-centric way, for reverse-engineering software into its architectural views, and for assuring the quality of a software system in terms of its architecture.

In 2001, the following research projects and groups were active:

Framework Editor for Java (Fred).
The project aims at increasing the knowledge of application frameworks and design patterns in software construction. The main result of the project is a software tool which supports a design pattern-based method of building reusable application frameworks in Java, and guides the systematic construction of final applications from the frameworks. By including a substantial part of the functional behaviour of a family of software systems in the same domain, applica-

tion frameworks are currently considered as a main approach to large-scale systematic software reuse. The follow-up project for Fred, James, will show the practical benefits of the developed method and tool by a number of industrial case studies.

Software Architecture Analysis, Recovery and Assessment (Saara).
The project develops methods and tools for (semi-)automatically creating architectural views over a software system family by extracting different kinds of patterns from the source code of the family. Technically, the patterns conform to the notion of "specialisation pattern" developed in the Fred/James project, so the method makes it possible to bring an application framework under the management of the Fred environment.

Metrics for Analysis and Improvement of Software Architectures (Maisa) & Techniques for UML-Based Software Development (UML++).
The projects develop methods for the measurement of software quality at the design level. The metrics are computed from the system's architectural description (given in UML), predicting the quality attributes of the system derived from it. Most notably, size and performance metrics are addressed. A novel technique that has been developed is design analysis and quality prediction by automated search of design patterns and anti-patterns from the architecture. Typically, a large number of design pattern occurrences is a sign of

29

good quality, whereas finding occurrences of anti-patterns probably is an indication of poor quality. The method is implemented in a software tool that computes the metrics, visualises the pattern occurrences that are found and, based on this information, predicts the quality of the software system to be built from the architecture.

Framework Integration and Testing Application (Rita).

Testing object-oriented frameworks is difficult because a framework, when considered as code, is an incomplete program whose behavior cannot be studied until actual applications have been derived from it. The Rita project develops methods and tools for the testing of framework-based software systems and for verifying the quality of testing with new forms of code coverage criteria. The project is part of the large European Eureka/ITEA project CAFÉ (From Concept to Application in System-Family Engineering) that develops techniques for systematically building system families.

Selected publications:

Verkamo A.I., Gustafsson J., Nenonen L., Paakki J.: Measuring Design Diagrams for Product Quality Evaluation. In: Proc. 12th European Software Control and Metrics Conference, London, England, 2001, 357-366.

Hakala M., Hautamäki J., Koskimies K., Paakki J., Viljamaa A., Viljamaa J.: Annotating Reusable Software Architectures with Specialization Patterns. In: Proc. Working IEEE/IFIP Conference on Software Architecture (WICSA 2001), Amsterdam, The Netherlands, 2001. IEEE Computer Society, 2001, 171-180.

Hakala M., Hautamäki J., Koskimies K., Paakki J., Viljamaa A., Viljamaa J.: Generating Application Development Environments for Java Frameworks. In: Proc. Third International Conference on Generative and Component-Based Software Engineering (GCSE 2001), Erfurt, Germany, 2001. Springer, 2001, 163-176.

Ferenc R., Gustafsson J., Müller L., Paakki J.: Recognizing Design Patterns in C++ Programs with the Integration of Columbus and Maisa. In: Proc 7th Symposium on Programming Languages and Software Tools (SPLST'2001) (T. Gyimothy, ed.), Szeged, Hungary, 2001. University of Szeged, 2001, 58-70.

4.2.4 Document Management, Information Retrieval and Data Mining (Doremi)

The Doremi research group is active in the areas of document management, information retrieval, data mining and human language technology. The fields of study include document assembly, automatic content management, event detection and tracking, text mining, and question-answering systems. The research is both theoretical and applied.

The research in document assembly concentrates on constructing tailored documents from existing documents and document fragments. The assembly process uses different methods in finding, clustering, reorganising, and modifying relevant fragments to be included in the new document. The assembled document may then be published in different media and platforms such as on paper or on the web. Typical applications include product manuals where a manual may be tailored to correspond to a certain customer-specific product configuration.

The more general field of automatic content management includes gathering informa-

tion from heterogeneous sources; monitoring several information sources continuously; filtering, combining and modifying information; creating new material by reusing content fragments; and storing and archiving content in databases. The group studies how to best help the knowledge worker automatically in tedious content tasks.

In event detection and tracking the group studies how to find interrelated stories in several independent news feeds. Related stories are organised according to topic and chained together forming events while repeated stories are filtered out. Techniques such as natural language technology, document clustering and text categorisation are used in this task. In a typical application, the user is able to follow a selected news story as it develops through time.

In text mining, the group develops data mining techniques for finding regularities in texts. The research has resulted in methods for finding all maximal word sequences with at least a given frequency in a document collection. The set of frequent sequences that occur in a document can then be used as a condensed content descriptor for the document.

A question-answering (Q/A) system provides answers to a user's written natural language questions from a database. If no answers can be found, a human expert writes a new answer, which is also stored in the database. The techniques used are based both on statistical and linguistic methods.

Project list:

Knowledge Worker's Workstation (TYTTI)
Question-Answering for Processing of Natural Language Queries (APPA)
ACME Media Convergence Core (MC Core)
Doremi Text Mining

Contact persons: Professor Helena Ahonen-Myka, Greger Lindén

Research group homepage: http://www.cs.helsinki.fi/research/doremi/

Selected publications:

Makkonen J. and Piitulainen J., Expanding document vectors in text categorization. In Proceedings of IR2001, Oulu, Finland, 2001, 53-60.

Ahonen-Myka H., Heikkinen B., Heinonen O., and Klemettinen M., Printing structured text without stylesheets in XML Scandinavia 2000, May 2-4, Gothenburg, Sweden, 2000.

Ahonen-Myka, H., Heinonen, O., Klemettinen, M., and Verkamo, A. I. (1999), Finding co-occurring text phrases by combining sequence and frequent set discovery, in R. Feldman, ed, Proceedings of IJCAI-99 Workshop on Text Mining: Foundations, Techniques and Applications, Stockholm, Sweden, pp. 1-9.

Heinonen, O. (1998), Optimal multi-paragraph text segmentation by dynamic programming, in Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and the 17th International Conference on Computational Linguistics, COLING-ACL, Vol. II, Université de Montréal, Montreal, Quebec, Canada, pp. 1484-1486.

31

4.3 RESEARCH PROJECTS

Algorithms
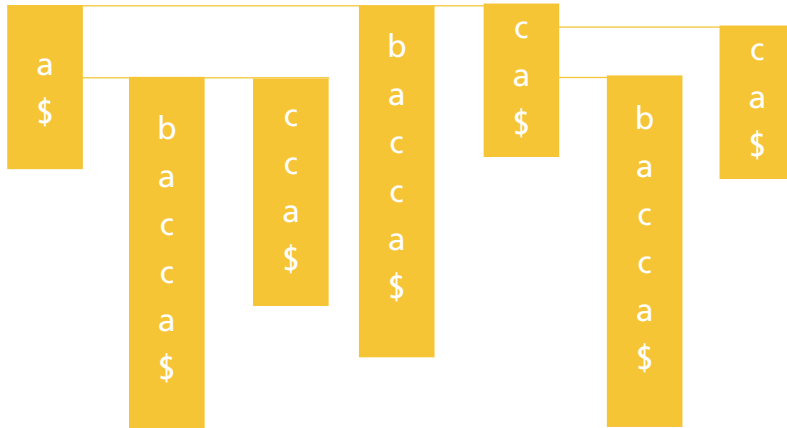
Project: Algorithmic Machine Learning
Researchers: Elomaa, Tapio; Rousu, Juho; Kääriäinen, Matti
Funding: Department of Computer Science

Summary:
Machine learning is concerned with the construction of computer programs that automatically improve with experience. We aim at mathematically rigorous analysis and development of learning programs and understanding of the underlying theory of the learning process(es). The group has studied extensively the properties of evaluation functions that are used to partition numerical attribute domains in decision tree learning algorithms.

Project: Robotics and Active Machine Vision
Researchers: Elomaa, Tapio; Autio, Ilkka; Kurppa, Teemu
Funding: University of Helsinki
Summary:
Robotics presents a challenging application area for computer science. We have a Nomad Super Scout II research robot with Vision. The project studies the control software architecture of an autonomous office robot. In particular, we are interested in uses of active machine vision and applications of machine learning methods therein.

Project: A Global Molecular Approach in the Study of Microbial Stress
Researchers: Ukkonen, Esko; Rantanen, Ari; Pesonen, Antti (VTT)
Funding: Academy of Finland (Life2000 Programme)
Summary:
This is a subproject of a larger consortium co-ordinated by Professor Hans Söderlund (VTT). The goal is to develop computer software and related theoretical concepts and algorithms for the data analysis, visualization and modelling of the transcriptional and translational data produced by the biology partners of the consortium. The following computational problems will be considered: clus-

tering of transcriptional and translational profiles; correlating such profiles against each other and against patterns in DNA sequences; synthesis of gene regulatory networks from data; computational support of a novel transcriptional profiling technique; analysis and alignment of 2D proteomics gels; development of data management infrastructure.

Project: Integrated Computational Methods for Genomic, Proteomic and
              Metabolic Modelling
Researchers: Ukkonen, Esko; Rousu, Juho; Mäkinen, Veli
Funding: Academy of Finland (MaDaMe Programme)
Summary:

The goal of this project is to develop computational methods and computer software for integrated analysis and mathematical modelling of genomic, proteomic, and metabolic data. The tools will aid to uncover and visualize the regulatory patterns and networks hidden in the data. Various computational approaches - such as pattern matching, machine learning, data mining, knowledge representation and sequence analysis - will be used.

Project: Logic-Based Query Languages for Molecular Biology Databases
Researchers: Ukkonen, Esko; Nykänen, Matti; Hakli, Raul; Tamm, Hellis
Funding: Academy of Finland, Department of Computer Science
Summary:

New application domains for database management systems demand that systems adapt to domain-specific data needs and not vice versa. In particular, databases of molecular biology data such as DNA strands encoded as strings must offer flexible tools for their manipulation. Thus we have extended the well-known relational database model to encompass strings as an independent data type, which permits user-defined string manipulation predicates within the query language. The extension draws its motivation from the concept of multiple sequence alignment employed in molecular biology, while its mathematical and computational aspects stem from temporal logic and automata theory, respectively.

33

Project: Neural and Computational Learning (NeuroCOLT)
Researchers: Ukkonen, Esko; Tirri, Henry; Elomaa, Tapio; Kivinen, Jyrki; Myllymäki, Petri
Funding: European Commission
Summary:

A European working group of more than ten partners on the theory and applications of machine learning.

Project: Pattern Matching and Machine Learning - Algorithms and
              Biocomputing Applications
Researchers: Ukkonen, Esko; Lemström, Kjell; Tamm, Hellis; Fredriksson, Kimmo; Suhonen,
              Heikki; Palin, Kimmo; Pienimäki Anna
Funding: Academy of Finland
Summary:

The project researches new algorithms for machine learning, pattern matching, and string matching, and applies them in various data analysis and information management problems. Special emphasis will be given to algorithmic problems in molecular biology and bioinformatics. Another application area is information retrieval in computer networks, for example, and in image and music databases.

Project: Structure, Assembly and Dynamics of Biological Macromolecule Complexes
Researchers: Ukkonen, Esko; Kivioja, Teemu; Ravantti, Janne; Verkhovsky, Anatoly; Mielikäinen,
                Taneli; Lamberg, Peter
Funding: Academy of Finland (RAKBIO Programme)

Summary:
The project develops original algorithms and software for the tomographic reconstruction of 3D mo-
dels for biological macromolecule complexes such as viruses. The computational problems in the
modeling of the assembly processes as well as the dynamic behaviour of such complexes will also
be considered. A joint project with Professor Dennis Bamford (Division of Genetics).

Intelligent Systems

Project: Computationally Efficient Probabilistic Learning and Reasoning
Researchers: Myllymäki, Petri; Kontkanen, Petri
Funding: Academy of Finland

Summary:
The main motivation for this work is the need for developing methods for constructing computatio-
nally intelligent models from sample data and expert domain knowledge. The constructed models
will be used for solving predictive (regression, classification) and explorative (data mining, visuali-
zation) modelling tasks. Although the advantages of probabilistic models have been largely recogni-
zed, the approach has often been neglected as a theoretically correct, but computationally infea-
sible methodology. The purpose of this research is to develop computationally efficient methods for
probabilistic modelling.

34

Project: Advanced FDIR
Researchers: Tirri, Henry; Myllymäki, Petri; Valtonen, Kimmo; Silander, Tomi
Funding: Space Systems Finland

Summary:
The motivation for this work is the need to increase the autonomy and reliability of spacecraft. This
is necessary partly to reduce operational costs, partly to enable missions where ground contact is
intermittent (e.g. low earth orbit) or not real time (e.g. planetary missions). Traditionally, failures de-
tected on-board a spacecraft trigger a transition to a "safe mode", in which the spacecraft tries to
ensure its survival, while waiting for ground control to analyse the failure and issue commands to
isolate it and recover from it. In the Advanced FDIR project, the goal is to avoid the "safe mode"
approach and instead develop on-board recovery procedures that can be used without interrupting
mission operation, even if recovery leads to degraded spacecraft performance.

Project: Computationally Efficient Methods for Deep Computing (DeepC)
Researchers: Tirri, Henry; Myllymäki, Petri; Silander, Tomi; Lahtinen, Jussi; Perttu, Sami
Funding: Academy of Finland

Summary:
Deep Computing is a term for methods solving complex and large-scale modelling and analysis pro-
blems with emerging computer systems that combine ultrafast processing with sophisticated analy-
tical software. Deep Computing can be seen to consist of three intertwined research areas: Deep
modelling (prediction and data mining with very large data sets), Deep optimization (computational-

ly efficient optimization of complex multivariate cost functions) and Deep view (interfaces for understanding high-dimensional data). The methodological research objective of the project is to develop the theory and methods required for obtaining very large-scale computational, data and communications capabilities that can be used to solve "grand challenge"-level Deep Computing problems in business and science. The applied research objective is to demonstrate solutions to previously intractable business and scientific problems by exploiting the advances in Deep Computing research in areas such as data modelling and analysis, high-end computing, search and optimization algorithms, and high-dimensional visualisation.

Project: Personalized Adaptive Interfaces (PAI)
Researchers: Tirri, Henry; Myllymäki, Petri; Nokelainen, Petri; Silander, Tomi; Ojanpää, Sami;
                   Misikangas, Pauli; Miettinen, Miikka; Lepola, Tuomas; Wettig, Hannes
Funding: TEKES, Alma Media, Ekahau, BayesIT
Summary:

The main objective of the PAI project is to develop methods for applying probabilistic modelling techniques, such as Bayesian network models, in building and using personalised, adaptive user interfaces. Specific research problems include user data segmentation, user profiling and user identification, and location-aware computing. The associated pilotprojects focus on problems related to intelligent educational technologies, adaptive WWW services and adaptive mobile services.

Project: PROMOS - Probabilistic Modelling of Baltic Salmon Stocks
Researchers: Tirri, Henry; Myllymäki, Petri; Valtonen, Kimmo; Mononen, Tommi
Funding: European Commission
Summary:

The objectives of the project are: 1. To develop a rigorous probabilistic assessment methodology for Baltic salmon stocks. 2. To undertake an application of the developed methodology by estimating the status of salmon stocks in the Gulf of Bothnia through a rigorous evaluation of Swedish and Finnish river data sets. 3. To develop a Bayesian decision analysis methodology for fisheries management of Baltic salmon and to give advice for the practical management. 4. To analyse the combined use of TAC management and effort regulation of passive gears in the coastal trap-net fishery of a mixed stock fishery and the information requirements of these management methods. 5. To identify the key sources of uncertainty and most effective management control measures given the uncertainties. In this study, preliminary estimates will be obtained for the historical development and present status of the case study baltic salmon stocks by combining various old and new data sets from rivers. This information will improve the empirical and scientific foundation for the ongoing Salmon Action Plan and for the yearly quota decision-making of the Baltic Sea fisheries.

Software Engineering

Project: Tool Support for Industrial Java Frameworks (James)
Researchers: Paakki, Jukka; Viljamaa, Antti
Funding: Tampere University of Technology (as a joint Tekes project)
Summary:

The project develops a software tool that supports a design pattern-based method of building reusable application frameworks in Java and guides the systematic construction of final applications from

the frameworks. The practical benefits of the resulting method and tool will be assessed in a number of industrial case studies.

Project: Software Architecture Analysis, Recovery and Assessment (Saara)
Researchers: Paakki, Jukka; Viljamaa, Jukka
Funding: Academy of Finland
Summary:
The project develops methods and tools for (semi-)automatically creating architectural views over a software system family by extracting different kinds of patterns from the source code of the family. Technically, the patterns conform to the notion of the "specialisation pattern" developed in the James project, so the method is integrated with the framework tool under development in James.

Project: Metrics for Analysis and Improvement of Software Architectures (Maisa)
Researchers: Paakki, Jukka; Verkamo, A. Inkeri; Gustafsson, Juha; Nenonen, Lilli
Funding: Tekes, Kone, Nokia Mobile Phones, Nokia Research Center, Space Systems Finland
Summary:
The project develops methods for the measurement of software quality at the design level. The metrics are computed from the system's architectural description (given in UML), predicting the quality attributes of the system derived from the description. A novel technique under development is design analysis and quality prediction by automated search of design patterns and anti-patterns from the architecture. The method is implemented in a software tool that computes the metrics, visualizes the pattern occurrences that are found and, based on this information, predicts the quality of the software system to be built from the architecture.

Project: Techniques for UML-Based Software Development (UML++)
Researchers: Paakki, Jukka; Gustafsson, Juha
Funding: Academy of Finland
Summary:
The project continues the research started in the Maisa project in a consortium with the University of Tampere and Tampere University of Technology as the other partners. The goal is to develop a comprehensive software development environment with UML as the central technology. One service of the environment will be architectural quality assurance, to be studied at the Department.

Project: Framework Integration and Testing Application (Rita)
Researchers: Paakki, Jukka; Kauppinen, Raine; Taina, Juha
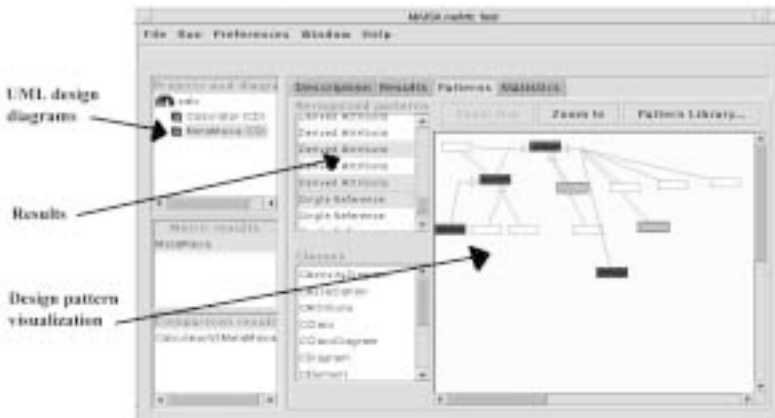
Funding: Nokia Research Center (as a joint Tekes project)

Summary:
The project develops methods and tools for the testing of framework-based software systems and for verifying the quality of testing with new forms of code coverage criteria. The project is part of the large European Eureka/ITEA project CAFÉ (From Concept to Application in System-Family Engineering) that develops techniques for systematically building system families.

Distributed Systems and Data Communication

Project: Improving Internet Protocols for Mobile Computing (IIP-Mobile)
Researchers: Alanko, Timo; Kojo, Markku; Raatikainen, Kimmo; Kuhlberg, Panu; Kyrö, Jaakko; Sarolahti, Pasi
Funding: TEKES, Nokia Research Center, Nokia Hungary, Sonera

Summary:
The objective of the IIP-Mobile project is to measure the TCP performance implications of those link characteristics that are typical for wireless wide-area links as well as to design and implement new experimental TCP performance enhancements and analyse the impact of the enhancements. The results are contributed to the Internet Engineering Task Force (IETF).

In addition, a real-time software emulator called Seawind is developed further in the project. Seawind enables measurements of real protocol implementations in modelled networking environments. Seawind provides a rich set of ways to define transfer characteristics typical for wireless links, including delays and errors.

The software also has the means to conduct large sets of experiments in an automatic fashion. Tools of analysing measurement data have been integrated into the Seawind software.

Project: Modelling of Concurrency (MOCO)
Researchers: Tienari, Martti; Karvi, Timo; Luukkainen, Matti; Kuuppelomäki, Päivi
Funding: Department of Computer Science

Summary:
Recently, research on the theoretical aspects of concurrency has concentrated on three areas. The theory of partially defined specifications and their refinement relations has been transferred from the bisimulation semantics to the decorated trace semantics. Secondly, the modelling and verification of timed systems has been studied. Thirdly, liveness verification with software tools has been under investigation. Besides these theoretical studies, experimental work on specification-related algorithms has been carried out.

Project: Production and Integration of Large Component Systems (Pilarcos)
Researchers: Kutvonen, Lea; Alanko, Timo; Vähäaho, Markku; Haataja Juha; Silfver, Egil
Funding: Tekes, Nokia, SysOpen, Tellabs

Summary:
The Pilarcos project develops middleware solutions for the automatic management of interorganizational applications. The application behaviour is governed by contracts between sovereign components; these contracts define the forms and rules of cooperation.

The federated application architecture under development presents an application as an architecture model, the roles in which are populated by traded service components. The services can be selected either at the development, startup or run-time of the application. The service component providers are independent of each other in the selection of implementation technology, for example.

The federated architecture requires advanced middleware services for contract negotiation, for distributed startup of services, and for establishment of heterogenous communication channels. The Pilarcos project develops prototypes of these services, mostly in a CORBA environment.

The project results are applicable to mobile computing or electronic commerce applications, for example.

Project: RTD-Pilot
Researchers: Kutvonen, Lea; Raatikainen Kimmo; Niklander, Tiina; Lindström, Jan; Raatikka,
               Vilho; Lyytinen, Ilpo
Funding: TEKES, Solid Information Technology, Nokia Networks
Summary:
RTD-Pilot continues the series of RODAIN projects in the study and prototype development of a real-time database system for the needs of telecommunication. The database solutions under study are distributed, fault tolerant and optimistic.

Project: TranSat
Researchers: Raatikainen, Kimmo; Kojo, Markku; Daniel, Laila
Funding: Alcatel
Summary:
The Transat project was launched at the end of 2001. The aim of the project is to develop the operation and performance of Internet transport protocols in a satellite environment. The project focuses on two main fields. The first is the enhancement of TCP protocols by improving the TCP sender and receiver algorithms, and to improve the support offered to the transport level at link level. The second is the enhancement of resource management for the satellite link, and on the IP level, the coupling of differentiated package treatment (IP QoS) to the differentiated services offered by a satellite link. A prototype implementation will be made based on the suggested solutions, and finally, experiments will be made to test it.

Project: BRAIN
Researchers: Raatikainen, Kimmo; Kojo, Markku; Manner, Jukka; Laukkanen, Aki
Funding: (EC) Nokia Research Center
Summary:
The main objective of the project is to propose a system architecture, which combines broadband radio access systems with UMTS and GSM to enable full coverage of seamless IP-based services for users in hot spot areas and on the move. Emerging fixed network services will be made available also for mobile users with comparable Quality of Service (QoS). Based on this technological development, business opportunities will be enabled and improved for companies as manufacturers, network operators, SMEs and content providers within the entire coverage area. People on the move will be connected to their office, the Internet and thereby to the market. The speed of information exchange, competition and business will increase to support the growth of the economy.

Project: CRUMPET
Researchers: Raatikainen, Kimmo; Tarkoma, Sasu; Leinonen, Jani; Styrman, Avril
Funding: European Commission
Summary:

The overall aim of CRUMPET is to implement, validate, and trial tourism-related value-added services for nomadic users (across mobile and fixed networks). In particular the use of agent technology will be evaluated (in terms of user-acceptability, performance and best-practice) as a suitable approach for fast creation of robust, scalable, seamlessly accessible nomadic services. The implementation will be based on a standards-compliant open source agent framework, extended to support nomadic applications, devices, and networks.

Project: Mobile IP Based Network Developments (MIND)
Researchers: Raatikainen, Kimmo; Kojo, Markku; Manner, Jukka
Funding: (EU) Nokia Research Center
Summary:

Driven by the increasing demand for mobile and cordless broadband services in hot spot areas like airports, campus, and conference centres, the MIND project will provide a true broadband multimedia IP-based radio technology. Starting from the results of its predecessor project (BRAIN), the MIND project will offer the integration of end-to-end services over IP and evolve IP towards mobility. Furthermore, it will enable the interworking of private, corporate, and public networks. It will also support a wide range of services (point-to-point, point-to-multi-point, symmetric, and asymmetric) and allow roaming as well as inter- and -intra-system handover with GSM/GPRS and UMTS networks. MIND will also include new network topologies like Ad Hoc, self-organising and meshed networks, enhanced QoS in IP-based mobile networks and self-organisation at all layers of HIPER-LAN/2.

39

Project: Non-Stop Real-Time Linux
Researchers: Raatikainen, Kimmo; Häkkinen, Auvo; Karlstedt, Mika; Vähäkangas, Taneli
Funding: Nokia
Summary:

nsrtLinux (Non-Stop Real-Time Linux) studied the real-time properties and error tolerance of Linux. With the help of testing and simulation programmes implemented in replicate server environments, the group examined delays that arise in the kernel and applications. Special focus was laid on the changes to time-slice management and to alternation, and on the effect of open source code CORBA platforms on delays. In addition, the group studied the real-time performance of open source databases.

Project: ROBOCOP QoS System Design
Researchers: Raatikainen, Kimmo; Kyrö, Jaakko
Funding: Nokia
Summary:

The Robocop project is in the process of specifying the middleware component architecture of embedded systems. Embedded systems vary from mobile phones and PDAs to digital TVs and network interface facilities. The core of the Robocop architecture consists of a component model framework and examples of software components on various abstraction levels. This approach allows the use of software components from any producer as part of the embedded software environments. It is also

possible to exchange hardware components without changes to the application program.

Project: Wireless Corba (wCORBA)
Researchers: Raatikainen, Kimmo; Kangasharju, Jaakko
Funding: TEKES
Summary:

The aim of the wCORBA project is to support the UHE/DCS in the preparations of joint submissions to the OMG Telecom Wireless RFP. By participating in joint submissions the UHE/DCS tries to affect the contents of the proposed technology so that the forthcoming standard nicely fits the visionary framework of future mobile systems by the mobile computing research group in the UHE/DCS. In addition, the project aims to implement an open source (GNU LGPL) prototype of the Wireless CORBA extension to an open source ORB, called MICO.

Information Systems

Project: Workflow Management (WorkMan)
Researchers: Laine, Harri; Puustjärvi, Juha
Funding: Department of Computer Science
Summary:

Workflow is a collection of tasks organized to accomplish a business process. The specification of a workflow presents the tasks, their mutual dependencies, as well as possible transactional demands on them. The tasks included in the workflow may be carried out in different locations. The WorkMan project implements the prototype of a workflow system. With the help of the prototype, we are hoping to ensure a reliable execution of transactional workflows in a distributed environment. In implementing the system, we are striving to use the services offered by database systems based on the SQL language as much as possible.

Project: Graphical Interface Solutions and Techniques (GIST)
Researchers: Laakso, Sari A.; Laakso, Karri-Pekka; Saura, Asko
Funding: Department of Computer Science, Interacta Design Oy
Summary:

To design a successful user interface, we need use cases based on the realistic goals of users. We have developed a tentative design process that shows how the features and the user interface of an application are derived from goal-based use cases. We define use cases with field observations and contextual interviews. Because each user observation session produces a great deal of data, we have developed a discount method for recording qualitative field observations. The report includes a sequential overview of the whole session as well as task-based sequential views of specific events.

Project: Automatic Content Management (ACME)
Researchers: Ahonen-Myka, Helena; Lindén, Greger; Örström-Koch, Siru
Funding: TEKES, Alma Media, Elisa, Ericsson, Sonera, YLE, Tietoenator
Summary:

Content management includes gathering information from numerous sources; monitoring several information sources continuously; filtering, combining and modifying information; creating new material based on or reusing content fragments; and storing and archiving content in databases. In this

project content management processes are analysed in order to recognise sub-processes that can be automated. The project is a subproject of the Core project of the Media Convergence programme of the Helsinki Institute for Information Technology (HIIT).

Project: The Knowledge Worker's Work Station (TYTTI)
Researchers: Ahonen-Myka, Helena; Heinonen Oskari; Makkonen, Juha; Piitulainen, Jussi;
                Lindén, Greger; Lehtonen, Miro; Fluch, Martin; Popescu, Andrei
Funding: TEKES, Alma Media, Nokia Networks, Vaisala, WSOY, Lingsoft
Summary:
The project aims at building tools for a professional knowledge worker (e.g. a journalist), including tools for creating, accessing, and assembling structured (XML) documents, and tools for detecting and tracking events in newsfeed.

Project: Question-Answering for Processing of Natural Language Queries (APPA)
Researchers: Ahonen-Myka, Helena; Heinonen, Oskari; Aunimo, Lili; Kuuskoski, Reeta;
                Makkonen, Juha; Viljanen, Kim; Virtanen, Otso
Funding: Companies
Summary:
A question-answering (Q/A) system provides answers to a user's written natural language questions from a database. If no answer can be found, a human expert writes a new answer, which is also stored in the database. The techniques used are based both on statistical and linguistic methods.

41

Project: Doremi Text Mining
Researchers: Ahonen-Myka, Helena; Doucet, Antoine; Kostiainen, Hanna-Kaisa; Kuuskoski,
                Reeta; Viljanen, Kim
Funding: Academy of Finland
Summary:
We attempt to extract from text a small set of word sequences that describe the contents of a document. Moreover, the methods should be able to handle hierarchically structured documents, e.g., XML. This means studying how to attach structural information to textual phrases. We also consider extracting generalised phrases from annotated texts, e.g., from texts annotated with morphological features (part-of-speech, case, number). Generalisation means replacing infrequent words with more general features, such as the part-of-speech.

Project: Data Mining Methods for Medical Genetics
Researchers: Mannila, Heikki
Funding: University of Helsinki
Summary:
Many common diseases have a clear genetic component. Different biological methods exist for finding information about the genomes of affected and control individuals, but using this information to actually locate the genes that predispose to the diseases is quite difficult. This project is collaboration with the Finnish Genome Centre, and it has developed data mining algorithms methods for finding susceptibility loci. The project has also developed simulation-based tools and methods that help in study design in isolated populations, and MCMC methods for finding possible genetic models on the basis of population risks. The tools and methods have been used in concrete studies.

Applied Computer Science

Project: Urban Biodiversity: a Multitaxon Model in Multiple Scales for Urban Planning
Researchers: Lokki, Heikki; Mäkelä, Matti; Nyqvist, Tommy
Funding: Academy of Finland
Summary:

As a part of an interdisciplinary team headed by Professor Juhani Lokki (Finnish Museum of Natural History, University of Helsinki) a valuation model is to be constructed as part of a general model of urban biodiversity.

5. Supporting Activities

5.1 ADMINISTRATION

Department Steering Committee in 2001

Members of Steering Committee	Deputies
Professor Timo Alanko, chairman (1st Jan – 31st July 2001)	
Professor Jukka Paakki, chairman (1st Aug 2001 ->)	Professor Kimmo Raatikainen
Professor Seppo Sippu	Professor Timo Alanko
Professor Henry Tirri	Professor Matti Mäkelä
	(1st Jan – 31st Oct 2001)
	Professor Tapio Elomaa
	(1st Nov 2001 ->)
Lecturer Auvo Häkkinen	Lecturer Teemu Kerola
Lecturer Harri Laine	Senior Assistant Jan Lindström
Researcher Tiina Niklander	Lecturer Markku Kojo
Student Sampo Lehtinen	Student Seija Törmälehto
Student Antti Mattila	Student Nuutti Rintala
Student Marko Saaresto	Student Joonas Muhonen
Amanuensis Jaana Heino, secretary (1st Jan – 31st July 2001)	
University Lecturer Pirjo Moen, secretary (1st Aug 2001 – 28th Feb 2002)	

5.2 LIBRARY

The Department maintains a library with large collections of literature in computer science. The library is mainly used by the staff and advanced students of the Department.

It is located in the same premises as the Department, but administratively it is part of the new Kumpula Science Library founded in March 2001 by the Faculty of Science. The move of the complete Department to new premises in Kumpula is expected to take place in 2004.

The library now holds about 49,000 volumes of literature, making it the largest computer science library in Finland. The annual cumulation is about 350 monographic titles and

43

200 journal subscriptions. In 2001 the library used about 485,000 FIM for acquisition of new material.

In 2000 and 2001 a major part of older material was deposited at the National Repository Library in Kuopio. Also, a number of journal subscriptions in paper were cancelled and replaced by electronic versions.

The library is open to everybody and collections are freely available for reading and browsing in the premises. Home loans, however, are not normally granted to outside visitors.

To search and locate the required literature, the library maintains a web database of its holdings. In this database the material is classified using the ACM Computing Classification System (CCS). Collections are also catalogued in the national library system (HELKA/LINDA) which will in a few years replace the library's own database.

The library is also responsible for the distribution and sale of departmental reports, including PhD theses. Paper copies of reports 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.

5.3 OVERVIEW OF COMPUTING FACILITIES

The department is dedicated to providing a wide range of advanced high-quality computing facilities for use by computer science faculty and students. The number of users of these facilities is about 3,800. The facilities are operated by a technical staff who are not only responsible for the installation and maintenance of the systems, but who also assist faculty and students in the use and development of software systems for research projects.

Our workstation network consists of more than 500 PCs (mostly Pentium III, most of them with flat TFT monitors) running Linux. Win-dows 2000 can be used as an alternative to Li-nux. More than 100 of the workstations are mobile laptops that can join and leave the network dynamically.

The general computing facilities include a farm of servers: general-purpose computers, a computing cluster, file servers and other functionally dedicated servers (mail, WWW, FTP etc.), and servers for different user groups. Linux is used almost entirely as the operating system for the servers. The centralised file servers utilise RAID technology and currently offer over 1.5 Tbytes storage space. Together, these systems support a wide variety of services, languages and software tools including electronic mail and news, graphics and visualisation tools, several typesetting systems, and relational database systems. Special attention has been paid to security and reliability.

Networking is based on switched 100 Mbit/s Ethernet with an optical backbone. The mobile laptops can also utilise a departmental IEEE 802.11b type radio network which currently has 15 base stations. In the Linux (and UNIX) environment NFS is used to share common resources. On the Windows side Samba (a Linux-hosted Lan Manager Server) is utilized. The workstations are used as tools for software development, in research and on all levels of teaching.

The network of the department is connected through a firewall to the university backbone network, giving access to general-purpose UNIX computers at the University IT Department as well as to the FUNET wide area network that links Finnish universities and research establishments. The IT Department also offers a large modem and ISDN pool for remote access.

In addition, the department has access to a number of supercomputing facilities at the Center for Scientific Computing.

The national FUNET network is further connected to the Nordic University Network, Nordunet, with a 5 Gbit/s connection. The Nordunet has a 2.5 Gbit/s connection capacity to the United States as well as many high-capacity connections to the European network infrastructure.

Appendices

ALGORITHMS

Journal articles

1. M. Nykänen: Using acceptors as transducers. *Theoretical computer science* 267, (2001) : 1/2, 83-104.

2. I. Shmulevich, O. Yli-Harja, E. Coyle, D-J. Povel and K. Lemström: Perceptual issues in music pattern recognition: complexity of rhythm and key finding. *Computers and the humanities* 35, (2001) : 1, 23-35.

3. T. Elomaa and J. Rousu: On the computational complexity of optimal multisplitting. *Fundamenta informaticae* 47, (2001) : 1/2, 35-52.

4. J. Rousu and R. Aarts: An integrated approach to bioprocess recipe design. *Integrated computer-aided engineering* 8, (2001) : 4, s. 363-373.

5. T. Elomaa and M. Kääriäinen: An analysis of reduced error pruning. *Journal of artificial intelligence research* 15, (2001), 163-187.

45

Peer-reviewed conference and collection articles

6. V. Mäkinen: Using edit distance in point-pattern matching. *Eighth Symposium on String Processing and Information Retrieval*, November 13-15, 2001, Laguna de San Rafael, Chile, 153-161.

7. V. Mäkinen, G. Navarro and E. Ukkonen: Approximate matching of run-length compressed strings. *Proceedings: Combinatorial pattern matching : 12th annual symposium, CPM 2001*, July 1-4, 2001, Jerusalem, Israel, 31-49.

8. K. Fredriksson and E. Ukkonen: Faster template matching without FFT. *2001 International Conference on Image Processing*, 7-10 October 2001, Makedonia Palace Hotel, Thessaloniki, Greece, 678-681.

9. T. Mielikäinen and E. Ukkonen: The complexity of maximum matroid-greedoid intersection. *Fundamentals of computation theory : 13th international symposium, FCT 2001*, August 22-24, 2001, Riga, Latvia, 535-539.

10. T. Elomaa and M. Kääriäinen: On the practice of branching program boosting. *Machine learning: ECML 2001: 12th European Conference on Machine Learning*, September 5-7, 2001, Freiburg, Germany, 133-144.

11. V. Mäkinen: Trade off between compression and search times in compact suffix array. *Algorithm engineering and experimentation : third international workshop, ALENEX 2001*, January 5-6, 2001, Washington, DC, USA, 189-201.

12. I. Autio, T. Elomaa and T. Kurppa: Support vector classification of landmarks for a mobile robot. *IC-AI'2001: proceedings of the International Conference on Artificial Intelligence*, Las Vegas, Nevada, USA, June 25-28, 2001, 151-157.

13. T. Elomaa and J. Rousu: Preprocessing opportunities in optimal numerical range partitioning. *2001 IEEE International Conference on Data Mining*, 29 November - 2 December 2001, San Jose, California, 115-122.

14. J. Pöllänen, J. Rousu and J. Kronlöf: A neural network tool for brewery fermentations. *Automaatio verkostuvassa maailmassa – Navigation for the new automation : Automaatio 2001*, September 4-6, 2001, Helsingin Messukeskus, 246-251.

15. K. Lemström, G. A. Wiggins and D. Meredith: A three-layer approach for music retrieval in large databases. *ISMIR 2001, 2nd Annual International Symposium on Music Information Retrieval*, October 15-17, 2001, Indiana Memorial Union, Indiana University, Bloomington, Indiana, USA, 13-14.

16. K. Lemström and E. Ukkonen: Including interval encoding into edit distance based music comparison and retrieval. *AISB-00 Symposium on Creative & Cultural Aspects and Applications of AI & Cognitive Science*, University of Birmingham, England, August 17-18, 2000, 53-60.

17. I. Autio, T. Elomaa and T. Kurppa: Robot landmark learning with SVMs. *Seventh Scandinavian Conference on Artificial Intelligence*, *SCAI'01*, 157-158.

18. D. Meredith, G. A. Wiggins and K. Lemström: Pattern induction and matching in polyphonic music and other multidimensional datasets. *Proc. the 5th World Multi-Conference on Systemics, Cybernetics and Informatics (SCI '2001)*, Orlando, Florida, July 22-25, 2001, 61-66.

PhD theses and technical reports

19. J. Rousu: Efficient range partitioning in classification learning. PhD Thesis, Series of publications / Department of Computer Science, University of Helsinki. A-2001-1.

20. K. Fredriksson: Rotation invariant template matching. PhD Thesis, Series of publications / Department of Computer Science, University of Helsinki. A- 2001-3.

INTELLIGENT SYSTEMS

Peer-reviewed conference and collection articles

21. A. Nevgi, P. Nokelainen, T. Silander, H. Tirri and K. Tirri: Modeling students' views on the advantages of web-based learning with Bayesian networks. *Intelligent computer and communications technology - learning in on-line communities : proceedings of the Tenth International PEG Conference*, PEG 2001, Tampere, Finland, June 23-26, 2001, 101-108.

22. P. Kontkanen, P. Myllymäki and H. Tirri: Classifier learning with supervised marginal likelihood. *Uncertainty in artificial intelligence: proceedings of the Seventeenth Conference (2001)*, August 2-5, 2001, University of Washington, Seattle, Washington, 277-284.

23. P. Myllymäki, T. Silander, H. Tirri and P. Uronen: B-course: a web service for Bayesian data analysis. *13th IEEE International Conference on Tools with Artificial Intelligence, ICTAI 2001*, November 7-9, 2001, Dallas, Texas, 247-256.

24. M. Paakko, P. Myllymäki, N. Holsti and H. Tirri: Bayesian networks for advanced FDIR. *ESA Workshop on "On-Board Autonomy",* 17-19 October, 2001, ESTEC, Noordwijk, The Netherlands, 311-318.

25. P. Myllymäki, T. Silander, H. Tirri and P. Uronen: Bayesian data mining on the web with B-Course. *2001 IEEE International Conference on Data Mining*, 29 November - 2 December 2001, San Jose, California, 626-629.

26. P. Kontkanen, P. Myllymäki and H. Tirri: Comparing prequential model selection criteria in supervised learning of mixture models. *Artificial intelligence and statistics 2001 : proceedings of the eighth international workshop*, January 4-7, 2001, Key West, Florida, 233-244.

27. P. Nokelainen, H. Tirri, J. Kurhila and T. Silander: Bayesian modeling approach to implement an adaptive questionnaire. *Proceedings of ED-MEDIA 2001, World Conference on Educational Multimedia, Hypermedia & Telecommunications*, Tampere, Finland, June 25-30, 2001, 84.

28. P. Ruohotie, P. Nokelainen, H. Tirri and T. Silander: Modeling individual and organizational prerequisites of professional growth: [papers presented at international conferences 1999-2001], 2001.

29. H. Tirri: Designing B-Course: observations from educational point of view. *Effective interactive artificial intelligence resources,* 40-43.

Other articles

30. R. B-B. Levy, M. Ben-Ari and P. A. Uronen: An extended experiment with Jeliot 2000. Proceedings of the First Program Visualization Workshop, Porvoo, Finland, July 7-8, 2000, 131-140.

31. E. Hyvönen (eds.): Inhimillinen kone, konemainen ihminen. *Suomen tekoälyseura ja Suomen automaatioseura*, 2001.

32. E. Hyvönen: Inhimillistyvä kone. *Inhimillinen kone, konemainen ihminen*, 1-15.

33. E. Ukkonen: Tietokone elämän ohjelman jäljillä. *Inhimillinen kone, konemainen ihminen*, 70-80.

34. E. Hyvönen: Tekoelämän arvoitus. *Inhimillinen kone, konemainen ihminen*, 81-103.

SOFTWARE ENGINEERING

Peer-reviewed conference and collection articles

35. M. Hakala, J. Hautamäki, K. Koskimies, J. Paakki, A. Viljamaa and J. Viljamaa: Architecture-oriented programming using FRED. *Proceedings of the 23rd International Conference on Software Engineering, ICSE 2001*, 12-19 May 2001, Toronto, Ontario, Canada, 823-824.

36. R. Ferenc, J. Gustafsson, L. Müller and J. Paakki: Recognizing design patterns in C++ programs with the integration of Columbus and Maisa. *Seventh Symposium on Programming Languages and Software Tools, SPLST'2001*, University of Szeged, Szeged, Hungary, June 15-16, 2001, 58-70.

37. M. Hakala, J. Hautamäki, K. Koskimies, J. Paakki, A. Viljamaa and J. Viljamaa: Annotating reusable software architecture with specialization patterns. *Working IEEE/IFIP Conference on Software Architecture*, Royal Netherlands Academy of Arts and Sciences, Amsterdam, The Netherlands, 28-31 August, 2001, 171-180.

38. M. Hautala, J. Hautamäki, K. Koskimies, J. Paakki, A. Viljamaa and J. Viljamaa: Generating application development environments for Java frameworks. *Generative and component-based software engineering: Proceedings of third international conference, GCSE 2001*, Erfurt, Germany, September 9-13, 2001, 163-176.

39. I. Verkamo, J. Gustafsson, L. Nenonen and J. Paakki: Measuring design diagrams for product quality evaluation. *Project control: satisfying the customer - proceedings of the 12th ESCOM (European Software Control and Metrics Conference)*, 2-4 April 2001, London, United Kingdom, including the SCOPE track on software product quality, 357-366.

Other articles

40. J. Eskola and J. Tarhio: Animation of flowcharts with Excel. *Proceedings of the First Program Visualization Workshop*, Porvoo, Finland, July 7-8, 2000, 59-68.

41. M. Mäkelä: Predictive software agents in pervasive computing. Series of publications / University of Petrozavodsk, 2001, 94-106.

42. J. Eskola and J. Tarhio: Ikoniohjelmointia Excelillä. *Dimensio : matemaattis-luonnontieteellinen aikakauslehti* 65, (2001) : 6, 52-55.

43. J. Paakki: Turha testaaminen on turhaa. *Systeemityö 2001 : 1*, 10-12.

PhD theses and technical reports

44. M. Hakala, J. Hautamäki, K. Koskimies, J. Paakki, A.Viljanmaa and J. Viljanmaa: Task-driven specialization support for object-oriented frameworks. Report / Tampere University of Technology, Software Systems Laboratory, 2001.

45. J. Paakki, M. Sakkinen, J. Koskinen: Ohjelmistotekniikka TJT C25: luentomoniste. Lecture notes / Department of Computer Science, University of Jyväskylä, 2001.

DISTRIBUTED SYSTEMS AND DATA COMMUNICATION

Peer-reviewed conference and collection articles

46. J. Manner and K. Raatikainen: Extended quality-of-service for mobile networks. *Quality of service - IWQoS 2001 : 9th international workshop*, Karlsruhe, Germany, June 2001, 275-280.

47. K. Raatikainen: Middleware. *MITA*, (2001), 153-157.

48. K. Raatikainen: Object Management Group. *MITA*, (2001), 383-384.

49. J. Manner, M. Kojo, A. Laukkanen, M. Liljeberg, T. Suihko and K. Raatikainen: Exploitation of wireless link QoS mechanisms in IP QoS architectures. *Quality of service over next-generation data networks*, 21-22 August 2001, Denver, USA, 273-283.

50. V. K. Shanbhag, K. Gopinath, M. Turunen, A. Ahtiainen and M. Luukkainen: EASN: integrating ASN.1 and model checking. *Computer aided verification: 13th international conference, CAV 2001*, Paris, France, July 18-22, 2001, 382-386.

51. M. Luukkainen: Verification of bounded retransmission protocol with a process algebra based method for dense time. *Workshop on Real-Time Tools*, Aalborg University, Denmark, August 20, 2001.

52. M. Luukkainen: Verification of dense time properties using theories of untimed process algebra. *Formal techniques for networked and distributed systems : FORTE 2001, IFIP TC6/WG6.1 - 21st International Conference on Formal Techniques for Networked and Distributed Systems*, August 28-31, 2001, Cheju Island, Korea, 353-368.

53. A. Lopéz, H. Velayos, J. Manner and N. Villaseñor: Reservation based QoS provision for mobile environments. *Applications and services in wireless networks : ASW'2001*, July 25-27,

2001, Institut National des Télécommunications, Evry, France, 125-138.

54. A. Lopéz, J. Manner, A. Mihailovic, H. Velayos, E. Hepworth and Y. Khouaja: A study of QoS provision for IP-based radio access networks. *Evolutionary trends of the Internet : 2001 Tyrrhenian International Workshop on Digital Communications, IWDC 2001*, Taormina, Italy, September 17-20, 2001, 139-157.

55. M. Kojo, A. Gurtov, J. Manner, P. Sarolahti, T. Alanko and K. Raatikainen: Seawind: a wireless network emulator. *Tagungsband : 11.GI/ITG Fachtagung über Messung, Modellierung und Bewertung von Rechen- und Kommunikationssystemen*, September 11-14, 2001, Aachen, 151-166.

56. J. Manner, A. Laukkanen, M. Kojo and K. Raatikainen: A quality of service architecture framework for mobile networks. *2001 International Conference on Third Generation Wireless and Beyond*, May 20 - June 2, 2001, San Francisco, USA, 66-71.

57. L. Burness, E. Hepworth, A. López and J. Manner: Architecture for providing QoS in an IP-based mobile network. *Expanding the wireless universe : IST Mobile Communications Summit 2001*, Barcelona, Spain, 9-12 September 2001, Universitat Politècnica de Catalunya.

58. K. Granlund: Resource locking in packet-switched wireless networks. *Algorithms and software for mobile communications: 10th Aachen Symposium on Signal Theory*, Aachen University of Technology (RWTH), September 20-21, 2001, 189-194.

59. J. Lindström: Benchmark for real-time database systems for telecommunications. *Databases in telecommunications II : VLDB 2001 international workshop*, Rome, Italy, September 10, 2001, 88-101.

60. T. Niklander: Increasing availability of real-time main-memory database. *Supplement to the International Conference on Dependable Systems and Networks*, 1-4 July 2001, Göteborg, Sweden, B-48...B-49.

61. J. Lindström and K. Raatikainen: Dynamic adjustment of serialization order using timestamp intervals in real-time databases. *Sixth International Conference on Real-Time Computing Systems and Applications, RTCSA '99*, December 13-15, 1999, Hong Kong, China, 13-20.

Other articles

62. G. Montenegro, J. Border, M. Kojo, J. Griner and Z. Shelby: Performance enhancing proxies intended to mitigate link-related degradations. *Internet Society*, 2001.

63. G. Montenegro, S. Dawkins, M. Kojo and V. Magret: End-to-end performance implications of slow links. *Internet Society*, 2001.

64. G. Montenegro, S. Dawkins, M. Kojo, V. Magret and N. Vaidya: End-to-end performance

implications of links with errors. *Internet Society*, 2001.

65. A. Gurtov: TCP performance in the presence of congestion and corruption losses. Series of publications / University of Petrozavodsk, 2001, 107-131.

PhD theses and technical reports

66. T. Alanko and Y. A. Bogoyavlenskiy: Modernin informaatioteknologian menetelmätutkimusta: STKV '2000 konferenssijulkaisu = Advances in methods of modern information technology : proceedings of FDPW '2000. Petroskoin yliopiston vuosittainen suomalaisen tietojenkäsittelyn viikko 2001. Petroskoi Finnish Data Processing Week at the University of Petrozavodsk 2001.

67. A. Gurtov: Making TCP robust against delay spikes. Series of publications / Department of Computer Science, University of Helsinki. C-2001-53.

68. M. Vähäaho, E. Silfver, J. Haataja, L. Kutvonen and T. Alanko: Pilarcos demonstration prototype: design and performance. Series of publications / Department of Computer Science, University of Helsinki. C, 2001, 64.

69. J. Haataja, E. Silfver, M. Vähäaho and L. Kutvonen: CORBA component model: status and experiences. Series of publications / Department of Computer Science, University of Helsinki. C-2001-65.

70. L. Kutvonen, J. Haataja, E. Silfver and M. Vähäaho: Evaluation and exploitation of OMG CORBA and CORBA component model. Series of publications / Department of Computer Science, University of Helsinki. C- 2001-11.

71. L. Kutvonen, J. Haataja. E. Silfver and M. Vähäaho: Pilarcos architecture. Series of publications / Department of Computer Science, University of Helsinki. C-2001-10.

INFORMATION SYSTEMS

Journal articles

72. P. Sevon, V. Ollikainen, P. Onkamo, H. T. T. Toivonen, H. Mannila and J. Kere: Mining associations between genetic markers, phenotypes, and covariates. *Genetic epidemiology* 21, (Suppl 1) (2001), S588-S593

Peer-reviewed conference and collection articles

73. P. Sevon, H. T. T. Toivonen and V. Ollikainen: TreeDT: gene mapping by tree disequilibrium test. *KDD-2001: proceedings of the Seventh ACM SIGKDD International Conference on Knowledge*

Discovery and Data Mining, San Francisco, CA, USA, August 26-29, 2001, 365-370.

74. H. Mannila and M. Salmenkivi: Finding simple intensity descriptions from event sequence data. *KDD-2001: proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, August 26-29, 2001, San Francisco, CA, USA, 341-346.

75. S. Sippu and E. Soisalon-Soininen: A theory of transactions on recoverable search trees. *Database theory - ICDT 2001: 8th international conference*, London, UK, January 4-6, 2001, 83-88.

Other articles

76. M. Lehtonen: Document assembly with XML structured source data. *XML Finland 2001 conference proceedings "Surviving the XML (r)evolution"*, November 14-15, 2001, Tampere, Finland, 52-61.

77. M. Salmenkivi: Gene expression: a new challenge for data mining. Series of publications / University of Petrozavodsk, 2001, 81-93.

PhD theses and technical reports

78. M. Salmenkivi: Computational methods for intensity models. PhD Thesis, Series of publications / Department of Computer Science, University of Helsinki. A, 2001, 2.

79. G. Lindén: The metalanguage XML: lecture notes - spring 2001. Series of publications / Department of Computer Science, University of Helsinki. C-2001-1.

80. P. Sevon, H. T. T. Toivonen and V. Ollikainen: TreeDT: gene mapping by tree disequilibrium test. Series of publications / Department of Computer Science, University of Helsinki. C-2001-32.

TEACHER IN COMPUTER SCIENCE

Other articles

81. P. Kommers, J. Kurhila and E. Sutinen: Mental programming by visual imagination. *Proceedings of the First Program Visualization Workshop*, Porvoo, Finland, July 7-8, 2000, 207-219

APPLIED COMPUTER SCIENCE

Peer-reviewed conference and collection articles

82. P. Kauppi, K. Lindblad-Toh, P. Sevon, H. T. T. Toivonen, J. D. Rioux, A. Villapakkam, L. A. Laitinen, T. J. Hudson, J. Kere and T. Laitinen: A second-generation association study of the 5q31 cytokine gene cluster and interleukin-4 receptor in asthma. *Genomics,* 77 (2001), 35-42

83. E. Hyvönen: Interval input and output. *Scientific computing, validated numerics, interval methods*, 41-51.

Other articles

84. E. Hyvönen: Tietojenkäsittely ja elämän arvoitus. *Tietoviikko* 7, 2001, 2.

GENERAL

Other articles

85. J. Paakki: It-koulutusrahaa on, mutta ei opettajia. *Helsingin Sanomat* 13.2.2001, A5

86. T. Niklander and J. Paakki (eds.): Muistiviitteitä Timosta: 60 vuoden aikaikkuna. University of Helsinki, Department of Computer Science, 2001, B-2001-1.

87. T. Rusanen and H. Tirri: Internet on vielä lapsenkengissä. *Johtoporras*1/2001, 6-10

88. J. Paakki: Suomeen kansainvälinen IT-yliopisto - ei kai nyt sentään! *Tietojenkäsittelytiede*, June (2001), 4-6.

89. J. Paakki: Reino Kurki-Suonio: akateeminen tietotekniikkavaikuttaja. *Tietojenkäsittelytiede*, June (2001), 9-12.

90. H. Tirri: Paikanna itsesi!: mobiililaitteiden paikantamisesta. *Tietojenkäsittelytiede*, June (2001), 20-23.

91. J. Paakki: TKT-päivät 2001: tietotekniikka muuttaa yhteiskuntaa - miten muuttuvat tietoteollisen alan yliopisto-opetus ja tutkimus? *Tietojenkäsittelytiede*, June (2001), 26- 29.

92. H. Lokki: Pakkotuutorointi: vieri- vai vieroitushoito? *Peda-Forum : yliopistopedagoginen tiedotuslehti* 8 (2001) : 2, 6-8.

93. J. Paakki: Yliopistoihin approbatur-tutkinto. *Helsingin Sanomat* 24.12.2001, A5.

94. E. Ukkonen: Metakääntäjistä Linuxiin ja bioinformatiikkaan: kolme vuosikymmentä tietojenkäsittelytiedettä yliopistossa. *Työvälineitä tietoyhteiskuntaan : Suomen matematikko-, fyysikko- ja tietojenkäsittelytieteilijäliitto ry (SMFL) 40 vuotta*, 22-31.

95. M. Tienari: Utvärdering av datavetenskapliga/datalogiska utbildningar i Sverige, Högskoleverkets rapportserie, 2001.

Other reports

96. T. Alanko and A. Pienimäki (eds.): Annual report 2000. University of Helsinki, Department of Computer Science, 2001, B-2001-2.

ACCEPTED THESES

PhD Theses

1. J. Rousu: Efficient range partitioning in classification learning. PhD Thesis, Series of publications / Department of Computer Science, University of Helsinki. A-2001-1.

2. M. Salmenkivi: Computational methods for intensity models. PhD Thesis, Series of publications / Department of Computer Science, University of Helsinki. A-2001-2.

3. K. Fredriksson: Rotation invariant template matching. PhD Thesis, Series of publications / Department of Computer Science, University of Helsinki. A-2001-3.

PhLic Theses

1. Jan Lindström : Optimistic concurrency control methods for real-time database systems. (C-2001-9).

2. Päivi Hurri : Hypertekstien samankaltaisuuden tunnistaminen. (C-2001-12).

3. Antti Viljamaa: Pattern-based framework annotation and adaptation: a systematic approach. (C-2001-52).

4. Heikki Helin: Supporting nomadic agent-based applications in FIPA agent architecture. (C-2001-63).

Master's Theses

Algorithms

1. Hämäläinen, Wilhelmiina: Principia conclusionis incertia: Epävarman päättelyn mallintaminen Bayes-verkoilla, Dempster-Shafer -teorialla ja TM-järjestelmillä (C-2002-23).

2. Iso-Markku, Perttu: Turvaverkot tietoverkoissa - miten suomalainen sosiaaliala sopeutuu tietoyhteiskuntaan? (C-2001-23).

3. Lepola, Tuomas: Pulssineuraaliverkon laskentavoimasta (C-2001-38).

4. Malinen, Tuomo: Reiluuden huomioiminen automaattiteoriaan perustuvissa rinnakkaisjärjestelmien verifiointialgoritmeissa (C-2001-60).

5. Pöllänen, Jermu: Oluen käymisnopeuden ennustaminen neuraaliverkon avulla (C-2001-31).

6. Rintala, Toni: Matkustajien kohdekerrosten ennustaminen Bayes-menetelmällä liikennetilastoja käyttäen (C-2000-66).

7. Suhonen, Heikki: Täplien etsiminen, luokitteleminen sekä kehityksen seuraaminen silmän verkkokalvolta otetuissa kuvasarjoissa (C-2001-3).

8. Taipale, Eero: Korjausluvun ehdotusmekanismin kehittäminen tekstinkäsittelysovelluksiin (C-2001-25).

9. Vähä-Erkkilä, Ari: POV-Ray-sovelluksen säteenseurannan hajautus (C-2001-19).

Intelligent Systems

10. Mäkinen-Okubo, Sami: Geneettinen ohjelmointi ja lokerointiongelma (C-2001-62).

11. Nieminen, Mikko: Sigmoidi neuroverkkojen aktivaatiofunktiona (C-2001-56).

12. Silvonen, Paula: Correcting and unifying domain-specific texts (C-2001-37).

13. Tonteri, Teemu: A Statistical modeling approach to location estimation (C-2001-26).

Distributed Systems and Data Communication

14. Asplund, Marko: Kolmikanta-arkkitehtuuri elektronisessa dokumentinhallintajärjestelmässä (C-2001-41)

15. Head, Teemu: Techniques for application integration (C-2001-13).

16. Hyytinen, Meri: Avoimet liitännät DXT:hen (C-2001-75).

17. Hämäläinen, Timo: Paikannetun mobiilimarkkinoinnin tietotekniikka (C-2001-70).

18. Kuhlberg, Panu: Effects of delays and packet drops on TCP-based wireless data communication (C-2001-7).

19. Laukkanen, Jussi: An Evaluation of IPv6 transition mechanisms in implementation of UMTS Internet access (C-2001-50).

20. Maaniemi, Kirsti: DX200-keskuksen suorituskyvyn mittaaminen ja mallintaminen (C-2001-91).

21. Pennanen, Mika: Agents in virtual home environment (C-2001-45).

22. Poikela, Toni: OSA framework implementation in 3G IN (C-2001-18).

23. Rinta-Aho, Teemu: Anonymous mobile Internet server (C-2001-57).

24. Sarolahti, Pasi: Performance analysis of TCP enhancements for congested reliable wireless links (C-2001-8).

25. Toivanen, Mikko: Käyttäjäagenttiprofiilien turvallisuus WAE-palveluarkkitehtuurissa (C-2001-74).

26. Tolvanen, Jarkko: Device security (C-2001-66).

27. Tolvanen, Mika: X.509-sertifikaattien elinkaari ja sertifikaattien hallintaprotokollat (C-2001-16).

28. Vanhala, Anne: Prepaid services in GPRS (C-2001-88).

Information Systems

29. Ahtiainen, Sari: Vaatimusten hallinnan mittaaminen (C-2001-58).

30. Ehrlund, Sampo Heikki: Tietokannan käsittelytavan vaikutus ohjelmiston suorituskykyyn (C-2001-55).

31. Haarala, Mia: Gathering and managing information for centralised user profiles for utilisation in third generation mobile services (C-2001-36).

32. Hakkarainen, Kari: Toipuminen työnkulkujärjestelmissä (C-2001-24).

33. Hellfast, Katja: Asiakkaan profilointi assosiaatiosääntöjen avulla (C-2001-79).

34. Juslin, Jari Sakari: Navigointi WAP-sovelluksissa (C-2001-35).

35. Jäppinen, Jani: Multimedian synkronointi Internetissä (C-2001-44).

36. Kariola, Jukka: Verkon julkaisujärjestelmät (C-2001-86).

37. Knekt, Johanna: Hajautetut sovellukset sekä niiden suorituskyky (C-2001-76).

38. Koponen, Pekka: Mediaattoritoimitusten työmäärän arviointi kauppasopimusta varten (C-2001-89).

39. Kraft, Janne: Toisinteiden laiska päivittäminen (C-2001-6).

40. Kulmala, Erja: Intranetin sisällön vaikutus intranetin käyttöön (C-2001-28).

41. Kuuskoski, Reeta: Kysymys-vastaus -järjestelmät (C-2002-1).

42. Lehti, Harri: Heterogeenisistä tietojärjestelmistä kerättävien tietojen valmistelu varastointia ja analysointia varten (C-2001-5).

43. Lehtinen, Anniina: Verkkopohjaisen oppimisympäristön arviointikriteerit (C-2001-27).

44. Lehtinen, Markus: Olemassa olevien relaatiotietokantojen ja XML:n yhteiskäyttö (C-2001-47).

45. Lehtonen, Miro: Semi-automatic document assembly with structured source data (C-2001-30).

46. Makkonen, Piia Marita: Optimaalisen assosiatiosäännön löytäminen yhdelle tai useammalle numeeriselle attribuutille (C-2001-4).

47. Marttila, Outi: Query optimization in main-memory databases (C-2001-72).

48. Muurinen, Anu: Älykortin valintakriteerit joukkoliikenteessä (C-2001-73).

49. Pärnänen, Antti: Toimialakohtaisen tuotetietosiirron hallinta yritysten arvoketjussa XML:n avulla (C-2001-81).

50. Siltanen, Olli: Digitaalinen aineistonhallinta (C-2001-84).

51. Tamminen, Tiina: Käyttöliittymäsuunnittelu projektin osana (C-2001-2).

52. Teinilä, Janne: Performance analysis of a large database in a customer relationship management system (C-2001-43).

53. Wallenius, Jukka: Applying neural networks in information retrieval (C-2001-80).

Software Engineering

54. Aarnio, Lauri: Automatic testing as a part of extreme programming methodology (C-2001-71).

55. Hendren, Jens: Software development reneval from customer specific projects to product development (C-2001-22).

56. Klemetti, Antti: Oliopohjaisen ohjelmiston laadun mittaaminen ohjelmistoprosessin varhaisissa vaiheissa (C-2001-77).

57. Kräkin, Toni: Komponenttiviitemalli ohjelmistotuotteen teknologiastrategian ohjausvälineenä (C-2001-40).

58. Laine, Juha-Pekka: Aspektiohjelmointi (C-2001-15).

59. Laitinen, Juha: Suunnittelumallit graafisessa käyttöliittymässä (C-2001-61).

60. Majuri, Minna: Suunnittelumallien visualisointi (C-2001-48).

61. Nikula, Isto: Log analysis as a part of a distributed system's debugging process (C-2001-69).

62. Nordlund, Lasse: Organisational development and IT service management in a medium-sized IT organisation (C-2001-54).

63. Paasiala, Liisa: Estimating software project effort (C-2001-46).

64. Riikonen, Nikolai: Organisaation ja johtamisen vaikutus ohjelmistoprojektin tehokkuuteen (C-2001-34).

65. Ruokonen, Minna: UML:n metamalli (C-2001-33).

66. Valtakari, Kasper: Billing systems for Internet service providers (C-2001-67).

67. Vuolasto, Jaakko: A framework for electronic dictionaries (C-2001-49).

68. Vänttinen, Sanni: Ohjelmistoprosessin kypsyyden arviointi (C-2001-14).

69. Yliruusi, Talvikki: Ohjelmistotuotantoprojektin riskienhallinta (C-2001-42).

70. Zheng, Hui: Runtime memory usage estimation from UML diagrams (C-2001-21).

Teacher in Computer Science

71. Elers, Virpi: Tietokoneohjelmien saavutettavuus ja neuropsykologisten testien tietokonetoteutus (C-2001-20).

72. Karvonen, Antti: Itsearviointi ja motivointi verkko-oppimisympäristössä (C-2001-59).

73. Rautama, Erkki: Käsitekarttojen mallintaminen ja käsittely tietokoneella (C-2001-17).

74. Silander, Pasi: Adaptiivisen opetushypermedian taksonomia (C-2001-68).

Applied Computer Science

75. Örström-Koch, Siru: Transformaatiot sisällönhallinnassa (C-2001-29).