Research Bytes

Welcome to the Department of Computing. We aim to produce this periodical once a term. Its frequency depends on *you!*. We are always looking for good copy, so please let Androulla have your ideas and material. In this issue we have featured two of our Research Groups: VIP and BioInformatics. We hope that you enjoy reading about these and it inspires you to contribute.

Chris Hankin, Director of Research

To all PhD students and staff
Did you know that PhD students have weekly meetings with tea and biscuits, on Thursdays at 15:30 in the staff common room? The meeting is open to staff as well, and all sorts of topics are discussed.

Also, there is a PhD online discussion forum at
http://www.doc.ic.ac.uk/%7Etwh1/discus/ **and an online notice board at**
http://www.doc.ic.ac.uk/~mcs98/SLURP.cgi?PhDWiki

Sophia and Julie

Dear new PhD students
Welcome to Imperial! As PhD Admissions Tutor, I have seen your file and have probably corresponded with you by email, although I may not know who you all are! I'll be at the PhD party on 28th November, so do say 'hi'. I look forward to seeing you there and hope your research with us will be enjoyable, exciting and rewarding.

Best wishes
Philippa

The Summer Science Exhibition

BACKGROUND

The Royal Society is the UK academy of science. Founded in 1660, its mission is to recognise excellence in science, to encourage research and its application, to further the role of science, engineering, and technology in society, and to promote the public's understanding of science. The Summer Science Exhibition is an annual showcase of scientific innovation, attracting thousands of people to the Royal Society every year. The origins of the exhibition lie in the mid-nineteenth century, but its focus has always been fixed firmly on the future. All of the research projects on display are competitively selected for their scientific quality, creativity, and public appeal.

The exhibition offered researchers a unique opportunity to

- raise the profile of their research with senior scientists, industrialists, government and the media,

- enhance their communication skills by discussing their research and its potential applications with a variety of visitors ranging from school groups to Nobel Prize Laureates,

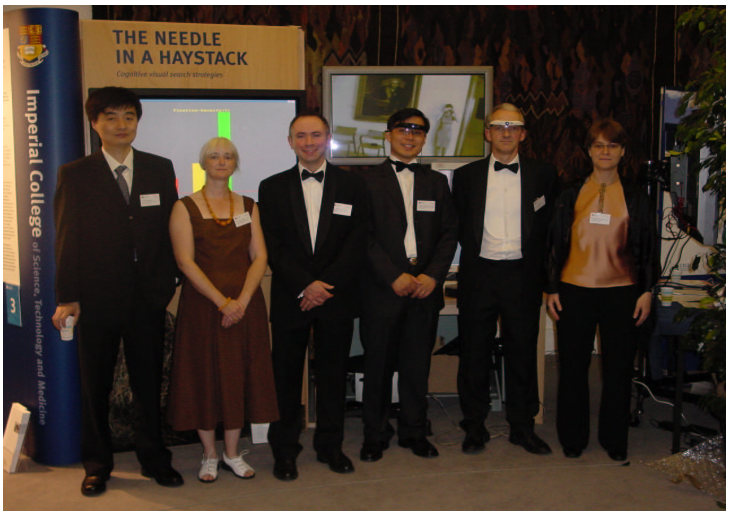
- have educational resources based on their research developed by consultants, and paid for by the Society,

- promote their institution/company to post-16 students looking at career choices,

- network with other research teams from across the country.

Twenty entries were selected for 2002 Royal Society Summer Science Exhibition (out of 120 submissions). ***That Needle in the Haystack***, led by Professor Guang-Zhong Yang of Department of Computing, explained cognitive visual search strategies involved in human vision. It was voted as one of the most popular exhibits for the 2002 event and has attracted a significant amount of media interests including BBC, Radio 4, and a number of national and international newspapers and IT magazines.

Each year over 2500 people visit the exhibition during the public opening times or as specially invited VIP guests to the Soirées. Thousands of students have been to the exhibition in previous years and enjoyed finding out about new research - how it relates to what they are studying in school or college and how it might affect their lives. Many have been inspired with new ideas about further study and found their assumptions about how scientists work have been changed forever.



***That Needle in the Haystack* – cognitive visual search strategies**

Visual Search is the act of searching for a target item within a scene. The eyes move across the scene, stop for a few hundred milliseconds at given points (the fixations) and a certain strategy is used to find the target. Search patterns differ with the observer and the task at hand. Many of the myriad of visual search tasks we perform every day of our lives are subconsciously undertaken, but specialists in fields such as medicine need to identify specific targets correctly and rapidly, often with life-saving consequences.

One way of understanding more about Visual Search is to make use of a biomedical tool, the eye tracker, which measures precisely the location of fixations in time and space. This allows one to follow and play back the eyes' path across a scene, and hence provides insight into an observer's strategy and its underlying cognitive processes. For a visitor to the exhibition seeing his, or her, scan path played back can be a fascinating experience for more reasons than one. First comes the realization that, although the impression had been one of a smooth search movement, the eyes were actually darting very rapidly from one spot to another in a seemingly disorganized way. This is followed by the question of whether the scan path suggests that a strategy, and hence, a mind process, was used in the search. Finally, watching the scan paths of other visitors emphasizes the fact that we all use our eyes slightly differently from each other. Within the context of an enjoyable experiment the visitor has experienced some of the fundamental principles of Visual Search. For the more advanced visitor, this exhibit will also outline cutting edge research into how medical experts assess X-ray and CT images, and how this expertise may be transmitted from the consultant to the trainee student.

Guang-Zhong Yang

Report on 12th Mini Euro Conference

The 12th Mini Euro Conference was held in Brussels in April 2002. Although it sounds a likely venue for Gordon Brown, it was actually organized by the Association of European Operational Research Societies. "Mini" because it only focused on a small number of topics within the full range of Operations Research concerns, this year: Decision Support Systems, Electronic and Mobile Commerce, Multicriteria Decision Aid, Human Centred Processes and Ethical Dilemmas in Decision Making. Quite a ragbag really – the streams on eCommerce seemed a poor fit with the rest of the proceedings.

I was there to present a paper written jointly with Berc Rustem: "Interactive Decision Making through Quadratic Goal Programming", one of three papers to be presented within the Goal Programming session in the "Multicriteria Decision Aid – Theory" steam. Most real world decisions involve considering more than one aspect of any possible solution, for example in choosing a new public transportation system, the number of passengers carried, capital and running costs, safety and pollution are just some of the issues that would need to be addressed. Hence the term "multicriteria".

When using Goal Programming, the decision maker starts off by specifying a set of bliss values which represent his idea of perfection – whether completely achievable or not. Using quadratic programming, our system attempts to find a way of achieving this solution. However, if this is not possible, the decision maker's sights may have to be set slightly lower and compromises will need to be made, involving trade-offs between the different dimensions. Our system offers an interactive method for arriving at an acceptable solution without any need for the decision maker to consider trade-offs explicitly. Instead, a dialog is carried out between the computer and user: the system proposes a feasible solution – the user responds with a modification, closer to his aspirations – the computer replies with the nearest feasible solution to what the user just proposed – if not happy with this, the user responds with another modification, and so on. In this way, the user finally achieves a feasible solution, as close as possible to his desires. The language in which the dialog is carried out is purely graphical – and uses the "parallel coordinate" approach to designating points in n dimensions, both for displaying solutions and for specifying desires.

Sadly, the audience was very sparse – five or six people, including the chairman and the other speaker. However, they came up with a least one interesting question, which Berc and I subsequently addressed in the final version of the paper.

Most of the far better attended plenary sessions deal with "Ethical Dilemma" issues, which amounted to pleas that OR solutions should embody concerns for sustainable development, lack of exploitation of the Third World, etc. Hardly revolutionary stuff!

Perhaps my strongest impression was that with ninety separate sessions, spread over ten parallel streams, there is a real danger that many of the sessions do not achieve a critical mass – both in terms of the number of papers presented and in the size of the audience. More concentrated and focussed workshops are probably a better way of presenting and sharing new ideas.

Frank Kriwaczek

Computational Bioinformatics Laboratory

The CBL is physically situated in the Department of Computing in what is presently Room 407 Huxley (an area containing 3 offices and a large open area for post-docs and PhD students). The remit of the Centre is the study of the theory, implementation and application of computational techniques to problems in Biology and Medicine. The Centre is developing strong and growing links with other related groups and centre across Imperial College.

Research

Scientific Knowledge Discovery using ILP

The pharmaceutical industry is increasingly overwhelmed by large-volume-data. This is generated both internally as a side-effect of screening tests and combinatorial chemistry, as well as externally from sources such as the human genome project. On the other hand the industry is predominantly knowledge-driven. For instance, knowledge is required within computational chemistry for pharmacophore identification, as well as for determining biological function using sequence analysis. From a computer science point of view, the knowledge requirements within the industry give higher emphasis to ``knowing that'' (declarative or descriptive knowledge) rather than ``knowing how'' (procedural or prescriptive knowledge). Mathematical logic has always been the preferred representation for declarative knowledge and thus knowledge discovery techniques are required which generate logical formulae from data. Inductive Logic Programming (ILP) provides such an approach.

ILP algorithms take examples *E* of a concept (such as a protein family) together with background knowledge *B* (such as a definition of molecular dynamics) and construct a hypothesis *H* which explains *E* in terms of *B*. For example, in protein fold domains, *E* might consist of descriptions of molecules separated into positive and negative examples of a particular fold (overall protein shape).

Discovery of biological function

Biological functions are regulated by the docking of small molecules (ligands) with sites on large molecules (proteins). Drugs, such as beta-blockers, mimic natural small molecules, such as adrenaline. Effectiveness of drugs depends on the correct shape and charge distribution of ligands. Thus beta-blockers block the binding of adrenaline, and so stop over-stimulation of heart muscle in patients prone to heart attacks.

Results on scientific discovery applications of ILP are separated below between those related to small molecules (such as ligands) and those related to proteins.

Small molecules

Structure-activity prediction

The majority of pharmaceutical R&D is based on finding slightly improved variants of patented active drugs. This involves laboratories of chemists synthesising and testing hundreds of compounds almost at random. The average cost of developing a single new drug is around $300 million. The ILP system Golem has been shown to be capable of constructing rules which accurately predict the activity of untried drugs. Rules were constructed from examples of drugs with known medicinal activity. The accuracy of the rules was found to be slightly higher than traditional statistical methods. More importantly the easily understandable rules provided insights which were directly comparable to the relevant literature concerning the binding site of dihydrofolate reductase.

Mutagenesis

The ILP system Progol has been used to predict the mutagenicity of chemical compounds taken from a previous study in which linear regression had been applied. Progol's predictive accuracy was equivalent to regression on the main set of 188 compounds and significantly higher (85.7% as opposed to 66.7%) on 44 compounds which had been discarded by the previous authors as unpredictable using regression. Progol's single clause solution for the 44 compounds was judged by the domain experts to be a new structural alert for mutagenesis.

Pharmacophores

In a series of ``blind tests'' in collaboration with the pharmaceutical company Pfizer UK, Progol was shown capable of re-discovering a 3D description of the binding sites (or pharmacophores) of ACE inhibitors (a hypertension drug) and an HIV-protease inhibitor (an anti-AIDS drug).

Carcinogenicity

Last year Progol was entered into a world-wide carcinogenicity prediction competition

run by the National Toxicology Program (NTP) in the USA. Progol was trained on around 300 available compounds, and made use of its earlier rules relating to mutagenicity. In the first round of the competition Progol produced the highest predictive accuracy of any automatic system entered.

Proteins

Protein secondary structure prediction. Golem was applied to one of the hardest open problems in molecular biology. The problem is as follows: given a sequence of amino acid residues, predict the placement of the main three dimensional sub-structures of the protein. The problem is of great interest to pharmaceutical companies involved with drug design. For this reason, over the last 20 years many attempts have been made to apply methods ranging from statistical regression to decision tree and neural net learning to this problem. Published accuracy results for the general prediction problem have ranged between 50 and 60%, very close to majority-class prediction rates. In our investigation we found the ability to make use of background knowledge from molecular biology, together with the ability to describe structural relations boosted the predictivity for a restricted sub-problem to around 80% on an independently chosen test set.

Discovery of fold descriptions

Protein shape is usually described at various levels of abstraction. At the lower levels each family of proteins contains members with high sequence similarity. At the most abstract level folds describe proteins which have similar overall shape but are very different at the sequence level. The lack of understanding of shape determination has made protein fold prediction particularly hard. However, it is intriguing that although there are around 300 known folds, around half of all known proteins are members of the 20 most populated folds. Progol was applied to discover rules governing these 20 most populated protein folds. Average in class cross-validated prediction was around 70% and many of the rules were judged to be good characterisations of the fold classes by Michael Sternberg, a world-class protein prediction expert at the Imperial Cancer Research Fund in London. (He is now Professor of Structural Bioinformatics in the Department of Biological Sciences at Imperial).

Conclusion

In his statement of the importance of this line of research to the Royal Society Sternberg emphasised the aspect of joint human-computer collaboration in scientific discoveries. Science is an activity of human societies. It is our belief that computer-based scientific discovery must support strong integration into the existing social environment of human scientific communities. The discovered knowledge must add to and build on existing science. We believe that the ability to incorporate background knowledge and re-use learned knowledge together with the comprehensibility of the hypotheses, have marked out ILP as a particularly effective approach for scientific knowledge discovery.

Stephen Muggleton
http://www.doc.ic.ac.uk/~shm/cbl.html

IFIP World Computer Congress Montreal August 2002

I attended the Theoretical Computer Science stream of this year's WCC. The theme was Foundations of Information Technology in the Era of Network and Mobile Computing (TCS 2002). There were two tracks – Track 1: Algorithms, Complexity and Models of Computation; Track 2: Logic Semantics, Specification and Verification. There were parallel sessions for the submitted papers and joint sessions for the invited speakers.

I presented a paper (joint work with Maria Grazia Vigliotti) on the Ambient Calculus (of Cardelli and Gordon) within the semantics track. As well as other papers on process calculi, there were papers on types and on verification.

I also attended talks on the complexity track. In many cases there was little or no connection with mobility, but the current hot topic is "Ad Hoc Wireless Networks", which are networks of transmitter/ receivers with no central control and limited ranges. The application areas may be limited, but would include setting up a communications network after the normal infrastructure had been destroyed in e.g. an earthquake.

Carl Gunter gave an invited talk on "micro-mobile code", such as programs written in square bar-codes on food packaging which could be used to instruct microwave ovens to cook the food appropriately. I spoke to Carl about his work in mobile computing. He has done work on the PLAN programming language for active networks, but is skeptical about practical take-up. The client-server model still has a lot of mileage left.

It was generally agreed that the conference organisation was not good, and the attendance was disappointing. I thought that organisation was over elaborate in some areas (over-designed website with flash) and lacking in other areas - no lunch organised and large gaps between sessions. The registration fee and banquet were both expensive, and participants may have been put off by the cost. The proportion of European participants versus North Americans was quite high, but there were few people attending TCS who did not have a paper to present.

On a more positive note, the atmosphere in the talks was friendly and questions were constructive rather than attacking.

Iain Phillips

http://local.cips.ca/info2002/en/index.html

Aesop Group

In April, Pete Harrison's Aesop Group hosted the 12th International Conference on Modeling Tools and Techniques for Computer and Communication System Performance Evaluation (Tools 2002), which was a roaring success. Shortly afterwards, the majority of the group had work published in the proceedings of the Eighteenth Annual UK Performance Engineering Workshop (UKPEW 2002), and we have been invited to host the workshop at Imperial in 2004. Pete travels to Orlando Florida later this month to give a plenary presentation at ISCIS 2002, then to India in December to address the International Conference on Stochastic Modeling in Cochin, India, to round off a successful year. For more details, and an explanation of the group's new name, visit http://www-aesop.doc.ic.ac.uk/autumn2002.

Dave Thornley
Aesop Group Research Liaison
Autumn '02

Recent Grant Announcements

☺ *S Muggleton,* **"Studying Biochemical Networks using Probabilistic Knowledge Discovery", £217,031, started April 2002 for 3 years.**

☺ *S Rueger,* **"Cultural Heritage Language Technologies (CHLT)", £63,643, started in June 2002 for 3 years**

☺ *M Sloman,* **Platform Grant, "AEDUS: Adaptable Environments for Distributed Ubiquitous Systems", £409,681, started in July 2002 for 5 years.**

☺ *J Darlington,* **"A distributed pipeline for structural-based proteome annotation using grid technology", £208,463, started in July 2002 for 3 years.**

☺ *J Darlington & Y Guo,* **"Biological Atlas of Insulin Resistance (BAIR)", £104,618(JD), £538,504 (YG), started in October 2002 for 5 years.**

☺ *J Darlington,* **"GENIE: Grid ENabled Integrated Earth System Model", £351,828, started in October 2002 for 3 years.**

☺ *S Muggleton,* **"Metalog - Integrated machine learning of metabolic networks applied to predictive toxicology", £657,510, started in October 2002 for 3 years.**

Recent announcements:

☺ *P Gardner,* **"Spatial Logistics for Querying Data on the Web", £172,157, starting in April 2003 for 3 years.**

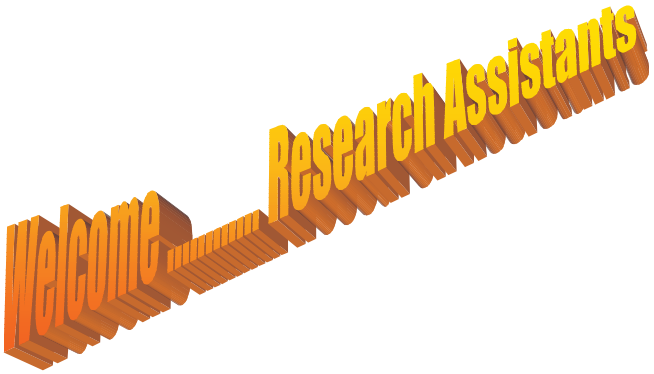
☺ *D Rueckert,* **"Modelling and Understanding of Immature Brain Development using Statistical Models of Shape and Appearance", £257,204, starting in April 2003 for 3 years.**

Welcome ... new PhD students

Chan Ahn
Tal Caplin
Rob Craven
Neil Datta
Fani Deligianni
Hywel Dunn Davies
Andreas Fidjeland
Howard Foster
Ali Hafiz
Jeffrey Hau
Altaf Hussain
Su-Lin Lee
Ting Ting Lee
Salman Marvasti
Michelle Osmond
Dhruv Pandya
Theodore Papatheodorou
Irene Papatheodorou
Panayiotis Parpas
Matthew Smith
Alexander Trifunovic
George Tzallas Regas
Luke Urquhart
Huy Vu
Uri Zarfaty

We are expecting more PhD students to register shortly!!!

Adrian Chung
Robert Chatley
Gary Kong
Danny Lee
William Lee
Nicholas Maudet
George Mylonas
Shay Seng
Sebastian Uchitel
Zheng Ye
Alex Yip
Yong Zhang



Congratulations to
Yike Guo who has been promoted to ***Professor***

Congratulations to
Sophia Drossopoulou who has been promoted to ***Reader***

Welcome.... New Lecturers

Dr Simon Colton

Dr Steven Newhouse

Dr Stefen Rueger

Dr Nobuko Yoshida

Welcome..... Staffing

Nilufer Betik - **Projects Secretary**
Steve Ingram -**Teaching support officer**
Oliver Jevons - **Operational Manager, e-Science**
Androulla Pieri - **Research Secretary**

Welcome..... Teaching Assistants

Nikolaos Rizopoulos

Our best wishes and thanks go to ex staff who have moved onto 'pastures new'….

Jane Bright
Manny Lehman
Katherine Piggott

TERM DATES:

Session 2002-2003

28 September – 13 December 2002
4 January - 21 March 2003
26 April – 27 June 2003

Session 2003-2004

4 October – 19 December 2003
10 January – 26 March 2004
24 April – 25 June 2004

Hold the back page

Graduate School of Engineering and Physical Sciences

On Monday 28th October 2002 the GSEPS was formally launched. The aim of the new School will be to enhance the postgraduate student experience and environment for its constituent faculties. The main activities of the new School will fall within the areas of academic training and quality assurance. It will bring together ten Departments of the Faculty of Engineering (including DoC!) and four Departments of the Faculty of Physical Sciences. **For more details, visit www.imperial.ac.uk/gradeps**

Research Bytes…. Research Bytes…… Research Bytes… Research Bytes… Research Bytes … Research Bytes….

British Computer Society (BCS)

Based in Swindon, Wiltshire this is one of the leading Chartered Engineering Institutions in the field of information systems. **For more details, visit to www.bcs.org/membership**

Research Bytes…. Research Bytes…… Research Bytes… Research Bytes… Research Bytes … Research Bytes….

Engineering and Physical Sciences Research Council (EPSRC)

The EPSRC has been given a mission by government to stimulate, facilitate and manage research and skill development in engineering and the physical sciences for the benefits of industry in the United Kingdom and the research base. It also has the role of promoting public awareness of science, engineering and technology. For more details, visit **www.epsrc.ac.uk**

Research Bytes…. Research Bytes…… Research Bytes… Research Bytes… Research Bytes … Research Bytes….

Research Bytes is ***the Research Newsletter for the Department!***

Our aim is to publish 4 times a year – end of October, January, April and July. We aim to include articles such as events for the diary, new funding opportunities, examples of innovative research, visits to conferences etc etc etc. Please send any contributions to Androulla Pieri, Research Secretary – ap15@doc.ic.ac.uk or phone on 020 7594 8220. We reserve the right to edit any contributions!

****DEADLINE FOR JANUARY ISSUE – 10th January 2003****