Summer 2011 Issue 2

InspiredResearch

RESEARCH NEWS FROM THE DEPARTMENT OF COMPUTER SCIENCE, UNIVERSITY OF OXFORD

QUANTUM PHYSICS ⇓ NATURAL LANGUAGE PROCESSING

GIVING MATHEMATICAL STRUCTURE TO LANGUAGE

UNIVERSITY OF OXFORD
DEPARTMENT OF COMPUTER SCIENCE

Welcome

In*Spired* Research is a twice-yearly newsletter published by the Department of Computer Science at Oxford University.

If you would like to learn more about anything you read in these pages please get in touch:

Department of Computer Science, University of Oxford,
Wolfson Building
Parks Road
Oxford, OX1 3QD

editorial@cs.ox.ac.uk

Editorial Board

Contributors

Andrew Markham
Andrew Martin
Bill Roscoe
Blanca Rodriguez
Craig Tranfield
Dan Olteanu
Dave Parker
Edward Grefenstette
Karo Moilanen
Leanne Carveth
Marta Kwiatkowska
Mehrnoosh Sadrzadeh
Nikos Tzevelekos
Philippa Hopcroft
Samson Abramsky
Stephen Cameron
Tahir Mansoori

Printed on recycled paper.

Computing Laboratory becomes Department of Computer Science

On 1 June 2011, The Oxford University Computing Laboratory changed its name to the Department of Computer Science, University of Oxford.

The annual Showcase day on which the Department invites representatives from industry to come to Oxford to explore our research activities was held to mark the date.

Bill Roscoe, Head of the Department, commented: 'This name change is simply to help the world at large understand our role as the University's department of computer science. I am excited that it gives us the opportunity to reach out more easily and tell everyone what we are doing in both teaching and research.

When the Computing Laboratory was founded in 1957 it literally was somewhere where the University's scientists such as Dorothy Hodgkin came to try things out on this new sort of machine. However we have long since become a large academic department doing world-leading research in many areas related to computing.'

'Killer app' for research launched

New free software, launched by Oxford University scientists, gives researchers the tools they need to collaborate more efficiently and quickly with colleagues scattered around the world and working in a variety of different research areas.

The Colwiz ('collective wizdom') R&D platform manages the entire research lifecycle from an initial idea, through a complex collaboration, to publication of the results.

'At the moment researchers are using a dizzying array of different applications to communicate and collaborate,' said colwiz Chief Scientist Professor David Gavaghan of Oxford University's Computer Science Department. 'These might include Google Apps, Microsoft Live Services, LinkedIn, Yammer and Social Text. But because these are separate applications they don't do everything and don't always talk to each other, and this slows researchers down. Colwiz replaces this hotchpotch with an integrated suite of tools custom-built for fast and efficient management of the research process.'

At the heart of the colwiz platform is a publication library that enables users to manage publications using both a desktop application (for Windows, Linux and Mac) and a version 'in the cloud' that can be accessed from anywhere over the Internet. This is combined with communications and collaboration tools for brainstorming, research tasks and schedule management.

'By breaking down the research process into its key components we have figured out which tools were potentially the most important. We then custom-built each tool from scratch and integrated them seamlessly into a single platform for individual and group productivity', said Tahir Mansoori, an Oxford doctoral student, and co-founder of Colwiz.

Colwiz is being launched through Isis Innovation's Software Incubator – a new programme designed to promote software start-ups from the University of Oxford. Members of academic institutions from US and UK universities can sign up for free and start using the Colwiz platform.

Further information: www.colwiz.com

New Royal Society Fellow

Ian Horrocks, a Professor in the Department has been elected as Fellow of the Royal Society.

His research interests include logic-based knowledge representation and reasoning and the semantic web, with a particular focus on ontology languages and applications. He played a leading role in the design of the World Wide Web Consortium's OWL ontology language standard, and he developed algorithms and optimisation techniques that underpin OWL reasoning systems. Applications of OWL are found in areas as diverse as life sciences, energy and defence systems. Ian is also a Fellow of Oriel College.

New Computer Science and Philosophy degree announced

The Department of Computer Science and Faculty of Philosophy have launched a new joint degree.

Artificial intelligence (AI), robotics, virtual reality: fascinating areas where Computer Science and Philosophy meet. But there are also many others, since the two disciplines share a broad focus on the representation of information and rational inference, embracing common interests in algorithms, cognition, intelligence, language, models, proof, and verification.

Although the combination of Computer Science and Philosophy at undergraduate level is new to Oxford, it is a well-established blend offered by universities across the globe. The programme aims to produce graduates with highly marketable skills. Computer Science teaches students how to program computers, and how to design processes that are effective and efficient. Philosophy teaches how to analyse complex concepts and the interconnections between them and – crucially – how to express this analysis, elegantly and precisely, in written form. Graduates will have the intellectual equipment needed for technical leadership and high-level positions in today's highly complex world. The initial cohort of students will start in October 2012. Computer Science and Philosophy is a three- or four-year course, leading either to a BA or a Masters degree. Students do not need to have previously studied either Computer Science or Philosophy.

The Department recently held a taster day for the degree with talks and demonstrations from staff and students, attended by around 30 prospective students, parents and teachers. Further open days are scheduled for 6 – 7 July and 16 September.

Further information: www.cs.ox.ac.uk/admissions/ugrad

Oxford success in robot football

Reader in Computing Science Stephen Cameron and three of his students have recently returned from RoboCup competitions in Germany and Iran, clutching a trophy and ready to tackle the World Championships in Istanbul in July. Students Helen Flynn, Jie Ma and Yanjun Zhou entered three of the categories, and came away with third place in the Virtual Rescue Robots competition.

Although originally focussed on robot football, RoboCup now encompasses a number of related disciplines which are designed to stretch our understanding of robotics and artificial intelligence, doing so in a competitive environment that pushes the researchers – and their creations – to the limit.

Students from the University's Computer Science Department have been entering such competitions for four years now, but this year sees them entering the competitions with real robots for the first time, thanks to a partnership with the universities of Newport and Crete. Their joint team is using robots from Aldebaran Robotics to compete in a standard football competition. The photograph shows the Oxford group holding their prize from Iran. Organisations interested in sponsoring the Oxford team should contact stephen.cameron@cs.ox.ac.uk.

TheySay – harnessing social intelligence for business improvement

It is hard to imagine an area, industry, issue, or topic that does not benefit from the ability to automatically monitor people's opinions, sentiments, and feelings.

A recently launched text analytics/ social intelligence start-up, TheySay, exploits the deep sentiment analysis technology developed as part of a DPhil research project that was completed recently by doctoral student Karo Moilanen under the supervision of Stephen Pulman.

TheySay, which was awarded an Oxford University Challenge Seed Fund (UCSF) grant in September 2010, applies deep Computational Linguistics methods to tackle what many regard as the 'trickiest' part of sentiment analysis – accurately determining the fine-grained sentiment expressed or implied in text towards specific entities such as mentions of products, people, companies, places, or even abstract issues.

Unlike existing sentiment tools, the algorithm developed by Karo and Stephen doesn't rely on statistical machine learning but rather applies logical affective reasoning to interpret sentiment in a much more human-like manner.

For example, the algorithm is able to reason that the expression 'Enrofloxacin can even kill bacteria' constitutes a positive outcome and that it is positive for 'Enrofloxacin' but negative for 'bacteria'.

TheySay's exhaustive analysis offers a real advance in the sentiment analysis field as a whole and addresses the genuine yearning expressed by end users and analysts for much more accurate capabilities and greater coverage than what is possible with the current state-of-the-art offerings.

Further information: www.theysayit.com

‘How was your day?’ – talking to a virtual companion

Most of us, when we get home after a long day at work, have someone to share the happenings of the day with. A partner, or a flatmate to complain to about the commute, or share our professional triumphs with. But what if the person you come home and ‘off-load’ to is not actually a person, or even your cat, but a computer? Is it possible for a virtual person to satisfy the emotional and practical aspects of our lives that having a human companion, or even a pet does? Is it possible to develop true partnerships between people and computers?

At Oxford’s Department of Computer Science, members of the Computational Linguistics Group have been involved in the Companions project – research into the development of virtual conversational ‘companions’ – computer systems that are intended to change the way people think about the relationships between humans and computers.

Relationships are a basic human requirement: we all need mutual caring and trust to ensure our emotional well-being. It is well documented that the biggest cause of depression amongst the elderly is loss of a human companion. We have relationships with companions, be they pets, friends or partners: to form a relationship, the user needs to care about the interaction, to invest emotion in it. We know that virtual companions can be used as a form of entertainment, but can they have medical and functional applications, for example to assist the elderly in maintaining an acceptable standard of life? Many computer scientists believe the answer is ‘yes’. We already know that people became emotionally dependent even on simple devices like the Tamagotchi, the Japanese digital pets that became a worldwide phenomenon in the 90s. The next step could be virtual companions that have a really profound effect on peoples’ mental and physical well-being.

It is these issues that have inspired COMPANIONS, a four-year, 12.88m Euro project that started in November 2006. The Oxford scientists are part of a consortium of 14 partners from across Europe and the US that brings together experts in a range of cutting-edge technologies including dialogue management, parsing and semantic interpretation, speech recognition and synthesis, embodied conversational agents and human-computer interaction.

The ultimate goal of the project was the creation of a Companion to help its human user at an emotional level by providing fellowship, as well as in practical ways. A Companion that would communicate with the user primarily by using and understanding speech; who could ‘chat’ to an elderly user to relieve their boredom, and perhaps also allow them to use the internet, irrespective of technical ability. It would work across a range of platforms, indoor and outdoor, static and roaming – an ‘agent’ that stays with the human user for long periods of time, allowing it to develop a relationship by ‹knowing› its owners’ preferences and wishes.

By learning its user’s habits, needs and memories, the Companion could become able to assist with carrying out tasks.

> ***‘A Companion might prompt its user to take their medication, or sound an alert if the user has said they are ‘popping out for half an hour’ and haven’t returned after a specified period.’***

An agent that has enough aspects of a human personality to establish loyalty and trust between users and the agents, that was believable, intuitive, and above all acceptable to the users. In a word: a Companion.

The project has been made up of a consortium of 14 partners from across Europe and the US, and funded under the Sixth Framework Programme of the European Commission. On the education side this includes the Oxford Internet Institute, the University of Sheffield's Natural Language Processing Group (where the project was born), the University of Teesside's School of Computing, Napier University's Centre for Interaction Design, the Swedish Institute of Computer Science, from Finland the Tampere University Unit for Computer – Human Interaction, in the US, the Universities of Washington and Albany, and from the Czech Republic, Charles University's Institute of Formal and Applied Linguistics and the Speech Processing Group of the University of West Bohemia.

Industry involvement comes from Loquendo, a spin – off of the Speech and Language Technology Telecom Italia R&D Labs, The Telefónica Group Research Labs in Madrid, and France Telecom in Paris.

Artificial human companion applications, in their most basic form, have been with us for decades. The earliest, such as the 'psychologist' program *ELIZA* from the 60s did little more than identify key words and feed them back to the user. The barrier to Companions so far, beyond very primitive forms, has been lack of progress in the adaptability of speech and language technology.

For this project the embodiment of a Companion is relatively unimportant: it could be engaged with via a screen head, a mobile phone, or other easily portable object. For the Oxford scientists the challenge has been for the Companion, whatever its size or shape, to be a properly functioning conversational entity.

The Computational Linguistics Group has been particularly involved in the *'How was your day?'* scenario part of the project. They have helped to build a conversational Companion – she's called Samuela – that can establish a relationship with a human user through conversations about the user's working day. The conversation is based on turn taking, where shifting of topic is motivated both by the user and by the Companion. Samuela greets the human, and as she engages him in conversation, she asks questions to find out more relevant information about what happened during the day, as you would with a normal human-to-human interaction.

The Companion has to guess the emotional state of the user – angry, upset, happy or excited. The system relies on the analysis of **what** is said – using a 'sentiment analysis' module developed by a DPhil student in the group, which has itself led to a commercial spinout called `TheySay' – see page 3 – and **how** it is said, detecting emotional effects in the speech signal. This emotion detection influences Samuela's facial expression and gesture, which aims to be appropriate to the mood of the speaker, and also the future direction of the conversation, eliciting sympathetic or encouraging behaviour.

As the conversation progresses the Companion is able to align her own 'emotions' with the user's. For example, on discovering the user has had a tough day, she comforts him, empathises with his problems, and attempts to cheer him up. Her words are soothing; her gestures sympathetic; the content of what she says relevant and helpful.

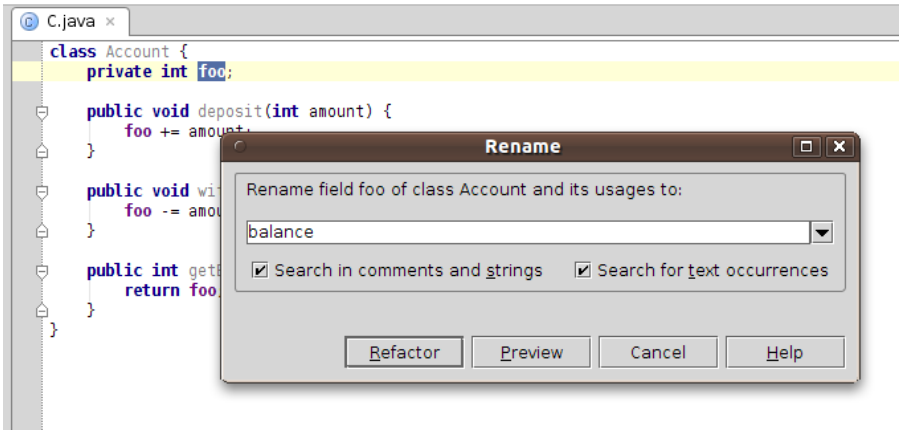
If the user feels that Samuela is talking too much, or has misunderstood, she can be interrupted. Depending on the type of interruption, and the current state of the dialogue, Samuela may keep silent, change the topic, or resume but with a different perspective. By manipulating the emotional polarity of her responses, Samuela is also able to generate some ironical or humorous responses: 'An argument? That will really improve things!'

Although there is still a long way to go, the Oxford team and their colleagues from across the globe have, through the COMPANIONS project, made significant advances to what is possible when it comes to human-computer conversations, and brought the possibility of having virtual Companions in our homes a lot closer to reality. Future possible developments might include experiments with Companions installed in the homes of elderly people living alone, specialised to help with everyday care and tasks, to see whether virtual Companions can improve their quality of life.

A video of a Samuela having a live, unscripted conversation with a human user is at: www.companions – project.org/?p=english.

Getting automated refactorings right

Refactoring is the process of improving the structure of an existing code base by means of behaviour – preserving source-to-source transformations, themselves called refactorings. In recent years, refactoring has become a very popular technique; modern integrated development environments (IDEs) for languages like Java all come with built-in refactoring support.



A refactoring is usually identified by a name, such as Rename Field, a set of preconditions to be checked for ensuring behaviour preservation, and a code transformation. For Rename Field, one has to check that there is no name clash with another field of the new name, and that no uses of other fields are shadowed by the renamed field; the transformation consists of renaming the field's declaration, and consistently updating all references to the field.

In most modern Java IDEs, such refactorings can be executed at the push of a button. The figure shows a screenshot of the popular Java IDE IntelliJ IDEA performing Rename Field: the programmer has selected the field to rename and provided the new name; the tool will then check the preconditions (informing the programmer of any violations), perform the transformation, and offer a preview of the proposed changes.

But, while the technique of refactoring itself is well-established, refactoring tools are not: a recent study found that programmers perform about 90 per cent of their refactorings by hand, even where tool support is available. One reason for this is the very low standard of correctness of current tools, which routinely cause compilation errors or subtle behavioural changes.

Over the past four years Oege de Moor, Torbjörn Ekman, Max Schaefer and Mathieu Verbaere have investigated techniques for improving the quality of refactoring tools in the context of the EPSRC-funded Aspect Refactoring Tools project. One key result of their research has been a dependency-based approach to the specification of refactorings, where refactorings are described in terms of their effects on static semantic dependencies of the program. For instance, the crucial correctness criterion for Rename Field is that it should preserve name binding: names in the refactored program should refer to the same declarations as in the original program.

In a series of publications at high-profile international conferences such as OOPSLA, ECOOP, ESOP and ICSE, the group has shown how this simple approach generalises to most other well-known refactorings and allows formulating high-level specifications of refactorings that are precise enough to be directly implementable, thus laying the foundations for a new generation of more robust refactoring tools.

And in other news...

- Bill Roscoe has been invited by the Ministry of Industry and Information Technology of China to present the payment security project to major mobile companies and banks in Beijing. Demos of the technology produced by this project can be viewed at www.cs.ox.ac.uk/hcbk.

- Congratulations to Nigel Crook of the Computational Linguistics Group, who has been appointed to the position of Head of the Department of Computing and Communication Technologies in the Faculty of Technology, Design and Environment at Oxford Brookes University.

- Linda Northrop of the Software Engineering Institute, Carnegie Mellon University, USA gave the May 2011 Strachey Lecture. Her talk was entitled *'The Impact of Scale'.*

- Bob Coecke obtained a FQXi Large Grant from the Foundational Questions Institute for his project *'a relativistic universe of interacting quantum processes'.*

- Muzammil Hussain has been awarded a Santander Travel Award. The award is for £900 and covers the travel costs for his collaboration work in University of Porto, Portugal, with the group developing localisation algorithms for autonomous underwater vehicles (AUVs).

- Version 2.2 of Chaste – the Cancer Heart And Soft Tissue Environment – has been released.

- Congratulations to Lu Feng, one of four people to reach the final round of the UK ICT Pioneers Competition, 'Technology Everywhere' prize. Organised by the Engineering and Physical Science Research Council (EPSRC), the competition recognises the most exceptional UK PhD students in ICT-related topics, who demonstrate innovation in the exploitation potential of their research. Lu is a third-year DPhil student under the supervision of Professor Marta Kwiatkowska, and is funded by a project studentship from the Large-Scale Complex IT Systems (LSCITS) Initiative.

Going underground: tracking badgers using wireless networks

University Lecturer Andrew Markham has recently been awarded an Engineering and Physical Sciences Research Council (EPSRC) cross-disciplinary fellowship to undertake research into tracking and monitoring badgers in their underground burrows.

Conventional technology such as GPS simply doesn't work underground, as soil heavily attenuates radio signals. Andrew has designed a novel technique that uses low frequency magnetic fields to continually and automatically map the positions of animals in 3-D. Furthermore, the badgers act as mobile mappers and over time reveal the internal architecture of the tunnel system.

The system uses conventional short-range (Zigbee) wireless links to transfer data from the badger collars to a wireless network. The purpose of this study is to completely characterise the behaviour and habits of badgers within their setts, and also to monitor the ambient environmental conditions.

This work has the potential to address open questions about the transfer of bovine TB within a badger population. The core technology is not restricted to use only on animals, and can be used in other environments where radio tracking is hindered, such as indoors or in parking garages.

DIADEM project wins Amazon grant

The DIADEM project, lead by Georg Gottlob, is the recipient of an Amazon Web Services Research Grant worth $7500 for 2011.

DIADEM (Domain-centric Intelligent Automated Data Extraction Methodology) addresses the problem of automating vertical object search in the world wide web using domain-specific knowledge. It is supported by an 'Advanced Investigator Grant' from the European Research Council (ERC). The Amazon research grant recognises the industrial relevance and scientific potential of DIADEM.

The grant supports the experimental evaluation of early DIADEM prototypes including OXPath, DIADEM's formalism for parallelised data extraction. OXPath is an extension of the well- known XPath language with features such as user action simulation for the navigation through menus and search masks, and the highly parallel execution of massive data extraction tasks using cloud computing.

The OXPath research team at Oxford consists of DPhil student Andrew Sellers, post-doctoral research assistants Giovanni Grasso, Tim Furche and Christian Schallhart, and Professor of Computing Science Georg Gottlob.

Designing better and faster DNA circuits

Oxford researchers have been awarded a Microsoft Research PhD Scholarship to fund new research into automated verification techniques for DNA computing.

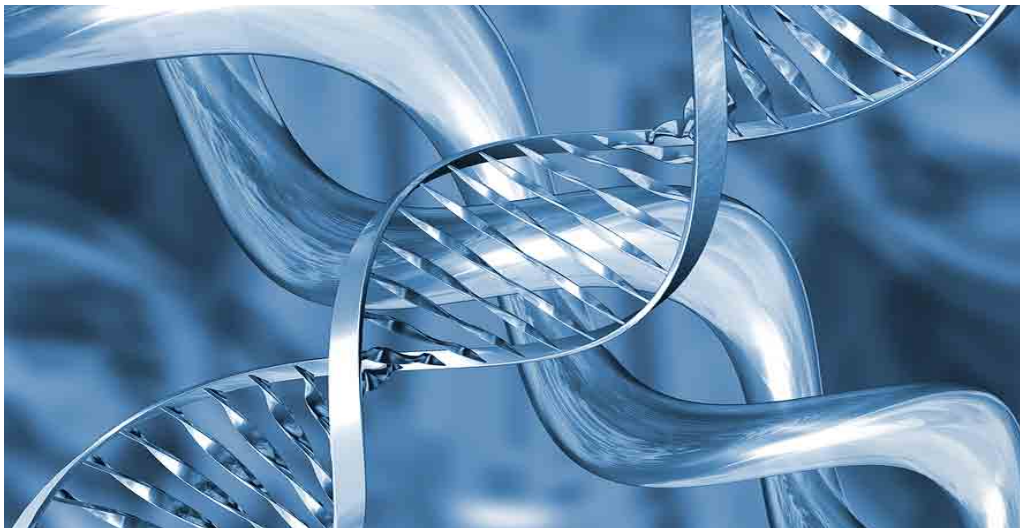
Oxford's Marta Kwiatkowska, in collaboration with Oxford Physics Professor Andrew Turberfield and Microsoft Research Cambridge's Andrew Phillips will develop new techniques and tools to help scientists design correct and efficient DNA computing designs.

DNA computing is a new and fast-growing field that aims to engineer artificial computing devices using bio-molecular materials such as DNA.

Inspired by the pioneering experiments of Leonard Adleman in the mid 90s, who used DNA to solve simple computation problems, researchers have succeeded in assembling increasingly complex nanometre-scale devices. The technology has exciting applications in the development of programmable molecular devices for use in bio-sensing and drug delivery. The new research project will develop techniques for formal verification of DNA computation and assembly designs. Formal verification methods apply rigorous, mathematical reasoning to ascertain the correctness of computerised systems. These are built into efficient software tools such as model checkers, which allow push-button use of verification technology.

The project will have a particular emphasis on quantitative verification techniques, which can be used to analyse properties such as the efficiency and reliability of DNA designs. This will build on PRISM, a software tool developed in Marta's group, for quantitative verification of systems with probabilistic behaviour. PRISM has already been applied to many diverse systems, from wireless communications protocols such as Bluetooth, to quantum cryptography algorithms, to biological signalling pathways.

The project will design new modelling languages and verification techniques for DNA computation designs and build them into new software tools. The techniques developed will be validated against experimental data from state-of-the-art DNA assembly/computation methods, performed in the Turberfield Lab at Oxford.



Quantum physics ⇓ natural language processing

By Edward Grefenstette and Mehrnoosh Sadrzadeh

The nature of language and its mechanics have been at the core of philosophical and scientific enquiry throughout the ages. From a way to persuade peers in political Athens to the medium through which we describe the world around us, language is the primary tool through which humans express their intelligence and rational ability. Some philosophers go so far as to erase or blur the distinction between language and intelligence. Transmitting at least some of our linguistic power, our capacity to understand language, to the computer increases its abilities and range of applications.

The task of giving mathematical structure to language, such that a simple procedure may be followed to reason about it or transform it, is nothing new. The association of philosophical logic, developed since Aristotelean times, with the rational language of the mind, goes back to Gottfried Wilhelm Leibniz and his desire to produce a formal calculus to serve as a perfect language. Further formalisation of this logic by Gottlob Frege, in conjunction with the formal approach to grammatical analysis developed by linguists such as Noam Chomsky, paved the way for a class of mathematical models of language called Formal Semantic Models.

These models, pioneered by Richard Montague, treat natural language as a programming language, where syntactic forms (grammatical rules) are paired with semantic interpretations such that the meaning of a sentence becomes a function of meanings of its parts. Using these models, computers can extract the logical structure from human language and evaluate their truth or falsehood, hence automatically reason inferentially or abductively about them.

The notion that meaning of expressions of natural language reduces to logical structures is not unique. The manner by which we should capture the propositional or predicative qualities of utterances such as '*hello', 'oh hell!' or 'pass me that stapler'* is not obvious. Furthermore, sometimes we wish to compare the meaning of individual words, or of sentences which are connected by topic or some other abstract relation, rather than by truth value. Finally, just as humans learn language without the provision of hard and fast rules or logical models, we'd expect a mathematical model of language

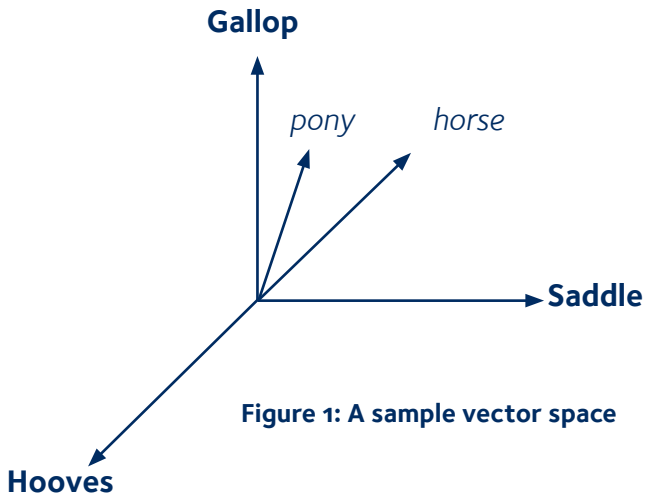


Figure 1: A sample vector space

which emulates our human linguistic abilities to take the learning process into account.

Empirical approaches such as the distributional class of semantic models seek to incorporate the learning process into their description of language. Following the Wittgensteinian creed that language gains meaning from and within the context of its use, these models ascribe meaning to words by calculating the distribution of their contexts of use from a corpus (a collection of written texts).

'Contexts of use' may refer to something as simple as the other words that occur close to the word being learned but can also refer to more complex structures based on grammatical relations, depending on the sophistication of the model. The distribution is generally calculated by counting how many times the word being learned occurs in each possible context, then smoothing this count, based on the probability of the context across the corpus, the probability of the word within the corpus itself, and other such factors.

The meaning of words can be automatically compared using their distributions. For example, 'pony' and 'horse' have similar meanings since they both 'gallop', have 'hooves' and are 'saddled'. These facts are reflected in their contexts, allowing us to determine how close their semantics are. Formally, and as depicted in Figure 1, the distributions can be represented as vectors, grouped as clusters related by topic; the angles between vectors measure the degree of similarity along a continuum rather than according to binary truth value.

These two classes of models both have advantages and limitations. Formal semantic models are syntax sensitive, compositional, reduce language meaning to well known logics, hence benefit from automated inference techniques thereof. However they also commit us to this reduction, which does not fit every notion of meaning and does not scale to naturally occurring language. Distributional semantic models are empirical and versatile, allowing us to learn different meanings of words by training the models against real corpora from different domains. However, they are word-based and do not scale to the sentence level and beyond: they lack compositionality. Both these mathematical models seem to capture some aspect of language meaning, yet are insufficient on their own. The question is:

> ***How can we develop a model of meaning which is both empirical and compositional?***

In a rare collaboration between linguists, logicians, and physicists, we are employing cross-disciplinary techniques to tackle this question, with firm foundations in the categorical formalism for quantum mechanics work by Samson Abramsky and Bob Coecke, from the Department's Quantum group. Their main contribution is the development of the first unified compositional distributional model of meaning, whereby meanings of sentences are built and reasoned about based on their grammatical structure as well as the empirical meanings of their constituent words.

Depicting the grammatical structure

In order to compute the meaning of a sentence consisting of *n* words, meanings of these words must *interact*. Hence the overall meaning should not be decomposed into *n* disconnected entities, as suggested by the empirical models and depicted in Figure 2. Instead, it must be acted upon further, as in Figure 3.

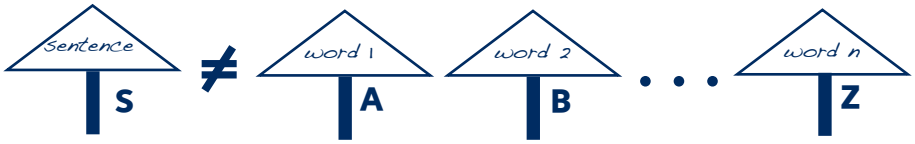


Figure 2: Disconnected meaning

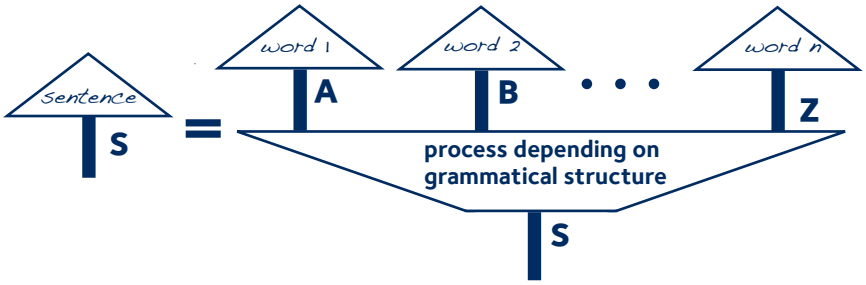


Figure 3: Interconnected meaning

But meanings of words are not sets, as suggested by the logical methods: they are vectors learnt from empirical distributions. The further action is still derived from the grammatical and logical structure of the sentence, but needs to be turned into a formalism compatible with the logic of distributional models, namely vector spaces. Surprisingly, similar logics have been used to analyse grammars of natural languages in structures known as Lambek pregroups, whereby the grammatical structure of a sentence is a network of interactions between the words. The structure of a transitive sentence is depicted in Figure 4, where the verb interacts with the subject and object via the underlying cups, then produces a sentence via the outgoing line.

Continued...

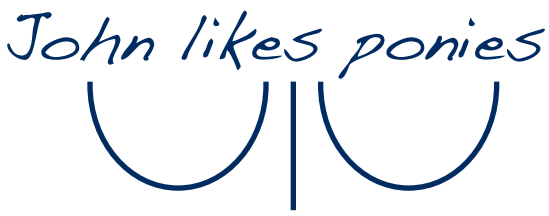


Figure 4: Grammatical diagram of a transitive sentence

Continued...

The structure of a negative sentence is depicted in Figure 5, where the logical words 'does' and 'not' play a major role in the overall inter-sentential interactions.

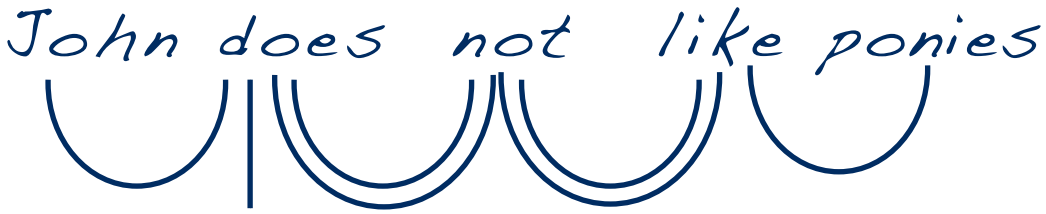


Figure 5: Grammatical diagram of a negative sentence

Although quite young in age, pregroups have been applied to formalising the grammar of a wide range of languages: from English and French to Arabic, Polish, and Persian. The corresponding interlingual interaction diagrams provide a simple way of comparing structures of sentences in different languages. See Figure 6 for the diagrams of 'John bought a pony from the auction' in English – French and Persian – Hindi. Without any knowledge of these languages and just by looking at the diagrams, one observes that the former is verb-in-the-middle and the latter verb-at-the-end (the outgoing lines mark the location of verbs).

Figure 6: Grammatical diagrams of a ditransitive sentence in English – French and Persian – Hindi

Quantising the grammatical structure

It remains to make the grammatical structure of the sentence *act* on the vector meanings of its words. This is done by transforming the pregroup interaction networks into a map that can act on vectors. This passage, often used in quantum field theory, is referred to as 'quantisation'. Due to the above mentioned logical advances in quantum theory, the quantisation of grammar is achieved by simply transforming the pregroup types to word vectors, a process depicted in Figure 7.

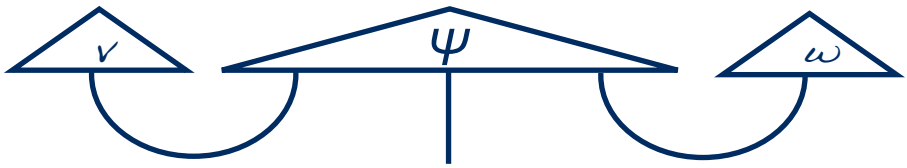


Figure 7: Meaning diagram of a positive transitive sentence

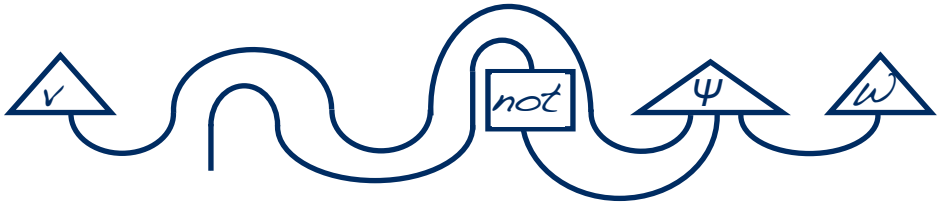


Figure 8: Meaning diagram of a negative transitive sentence

Here v and w are the meanings of subject and object, and Ψ is the meaning of the verb. The diagram defines the meaning of a sentence to be the application of a map corresponding to the grammatical structure of the sentence to the meanings of the words therein. Meanings of sentences can now be compared using the angle between their vectors, even if they have different grammatical structures.

Diagrammatic computations play a crucial role in modeling meanings of logical connectives, for example see Figure 8 for the meaning of a negative transitive sentence.

This diagram tells us that to learn the meaning of a negative sentence, first apply the meaning of its verb to that of its subject and object, then apply the meaning of 'not' to it. The order of this meaning assignment differs from the order in which the words are uttered. The diagram reflects this change of order: subject is still related to the verb but via caps that skip the verbal barriers created by 'does' and 'not'. The mechanism that makes this procedure work is similar to that of entanglement in quantum information. Entanglement models non-local correlations, which cause, as described by Einstein, spooky actions from a distance; it is the building block of protocols such as teleportation, see Figure 9, and quantum key distribution.

Evaluation and experiments

In recent work, the authors implemented the above categorical setting and evaluated the fragment of it that deals with intransitive and transitive sentences. We developed and implemented an algorithm that learns matrices for these verbs, evaluated our setting on a disambiguation task, and performed some experiments. Most verbs have different meanings and the context in which they appear can be used to disambiguate them: 'show' can mean 'express' in 'the table showed the result' or 'picture', in 'the map showed the location'. We built meaning vectors for these sentences compositionally, and used the degrees of synonymity of the sentences to disambiguate the meanings of their verbs. The data set for one of our experiments is sampled in Table 1. For instance, the pair of sentences in row 1 has a high degree of synonymity, whereas the pair in row 2 has a low one.

We judged the performance of our method by asking human subjects to rank the synonymity of each pair. The main measure of success is the correlation between these two

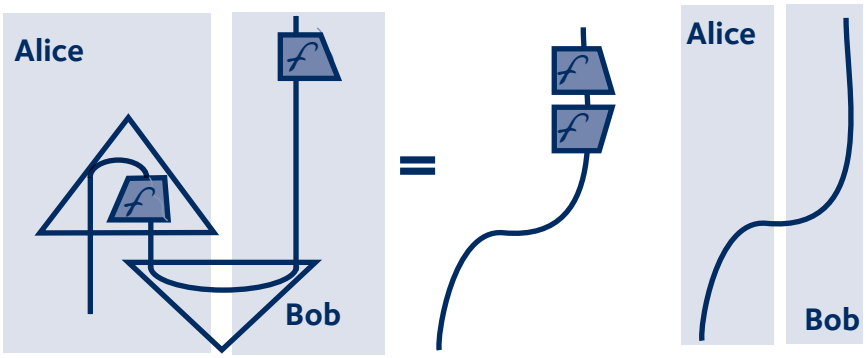


Figure 9: Diagram of quantum teleportation

	Sentence 1	Sentence 2	Degree of synonymy
1	table show result	table express result	6
2	table show result	table picture result	2
3	map show location	map picture location	6
4	map show location	map express location	3

Table 1: Part of the evaluation data set

variables: degree of synonymity as computed by our method and as ranked by people. Our method outperformed the others with a good margin.

Ambition

To be able to work with logical words in Google, one needs to decompose them in the 'advanced search' tab and provide extra context for their different meanings. This is fundamentally non-compositional and goes against the spirit of automated search. It is exactly here that the lack of compositional methods in meaning assignment causes practical problems and where our compositional methods become essential. We are in contact with Google researchers about this line of work. We are confident that our results are also able to tackle related problems in other natural language processing tasks such as paraphrasing and question-answering.

Ancient text in Sanskrit

Verum: creating zero-defect software systems

Robert Howe and Guy Broadfoot founded Verum in 2004 to create a better way to develop embedded software. Verum's establishment and on-going work on verification have been based on links with Oxford Computer Science.

Guy Broadfoot is CTO at Verum and a graduate of Oxford's part-time MSc in Software Engineering. The ASD:Suite, Verum's product, is based on Guy's MSc dissertation in which he explored how one could make the benefits of Tony Hoare's CSP (Communicating Sequential Processes) accessible to software developers working in industry, by combining it with a practical high level specification language. In professional collaboration with fellow Oxford graduate Philippa Hopcroft, Verum's Head of IP and Research, he further developed these initial ideas and jointly authored the resulting patents.

This work became the core of Verum's Automated Software Design (ASD) platform, giving software developers access to fully automated, formally based software design verification.

The ASD:Suite verifies runtime behaviour by generating CSP models and using Oxford's model checker FDR which is integrated into it. The ASD:Suite enables the most complex concurrency and control errors to be discovered and eliminated early in the software development lifecycle and before any implementation is done. As a result, creators of embedded software are able to bring reliable products to market quickly, predictably and cost-effectively.

Current ASD customers include Philips and Ericsson who report that ASD delivers a direct productivity increase from the conventional norm of 1 – 3 delivered lines of code per man – hour to 12 – 24 defect-free lines of delivered code per man-hour. This is equivalent to a reduction in cost from €40 to €5 per delivered line of code. Taking into account other factors, such as engineering overhead and the presence of existing code in a system, Ericsson has reported a net 30 – 50 per cent reduction in software engineering and maintenance costs, a fact that is driving their further adoption of the product.

Verum collaborates with the Department of Computer Science's Bill Roscoe on extending the use of CSP and FDR within the ASD domain in order to push the boundaries of the types and size of industrial systems that can be modelled and verified. Verum has recently sponsored the implementation of the priority operator described in Bill's new book *Understanding Concurrent Systems* because it enables a vital check of the systems ASD generates. Bill explains he 'always enjoys collaborating on novel industrial applications like this of CSP and FDR because they throw up new and important challenges that theoretical work alone would never uncover.'

Algorithmic nominal game semantics

In their paper 'Algorithmic Nominal Game Semantics', which appeared at ESOP (the 19th European Symposium on Programming) in March, Andrzej Murawski and Nikos Tzevelekos presented an algorithmic characterisation of a fragment of the ML language with numerical references. Their construction comes as a culmination of a three-year long collaboration and capitalises on major achievements of the last years in semantics of programs with dynamic generation of resources, and in particular on nominal game semantics and fresh-register automata.

Devising precise descriptions of higher-order programs with dynamic generation of resources is a daunting task, especially if one is interested in focussing just on observable behaviours. The reasons for this complication are due to the fact that dynamic resources combined with higher-order functions give rise to issues of privacy and data flow which cannot be resolved by standard analyses. A major breakthrough in this direction has been made recently with the introduction of nominal game semantics, a strand of the well-established theory of game semantics which addresses generative dynamic behaviours.

Andrzej and Nikos have been instrumental, not only in developing the theory, but also applying it towards giving algorithmic representations of its models. In a paper which appeared in FOSSACS (The International Conference on Foundations of Software Science and Computation Structures) in 2009, they introduced a game model of Reduced ML, a fragment of ML with integer variables. Their model was the first such to provide an explicit characterisation of observable program behaviour – and it did so in a compositional denotational manner. Moreover, the explicitness of the description allowed for algorithmic representations of the model, by means of abstract machines.

Primitive such machines for dynamic resource generation, called fresh-register automata, had been recently introduced by Nikos in a paper which appeared in POPL'11. Andrzej and Nikos further extended these machines to capture game semantics. They were thus able to represent their model of a non-trivial fragment of Reduced ML and thus prove decidability of program equivalence. This is the first such result in languages with dynamic generative features and paves the way for algorithmic descriptions of other languages of this kind, including fragments of Java or C#.

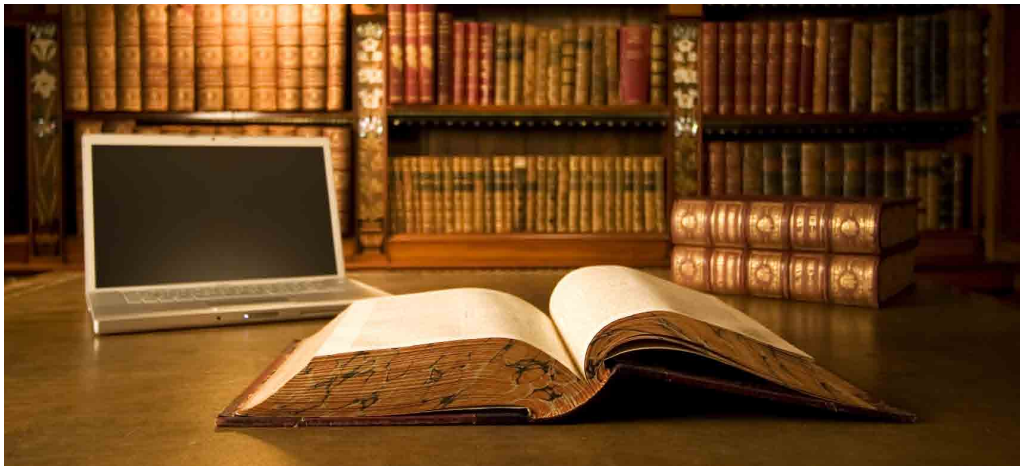
Book news

Bob Coecke has edited a new book entitled New Structures for Physics. This volume provides a series of tutorials on mathematical structures which recently have gained prominence in physics, ranging from quantum foundations, via quantum information, to quantum gravity.

These include the theory of monoidal categories and corresponding graphical calculi, Girard's linear logic, Scott domains, lambda calculus and corresponding logics for typing, topos theory, and more general process structures. Most of these structures are very prominent in computer science; the chapters here are tailored towards an audience of physicists.
New Structures for Physics. **Bob Coecke (Ed.). ISBN 978 – 3642128202**

Foundations of Semantic Web Technologies, a textbook co-written by Oxford's Markus Krötzsch has been recognised in the American Library Association's prestigious list of 'Outstanding Academic Titles'.

The book is one of just seven Information and Computer Science books highlighted in this year's 'Outstanding Academic Titles' – a nomination that brings with it extraordinary recognition amongst the international academic library community.
Foundations of Semantic Web Technologies. **Pascal Hitzler, Markus Krötzsch, and Sebastian Rudolph. ISBN: 978 – 1420090505**



Bill Roscoe has published a new book entitled Understanding Concurrent Systems.

It succeeds his 1997 book Theory and practice of concurrency as the standard text for studying concurrency via Hoare's Communicating Sequential Processes (CSP), and contains much material on the theory and applications CSP developed in recent years.

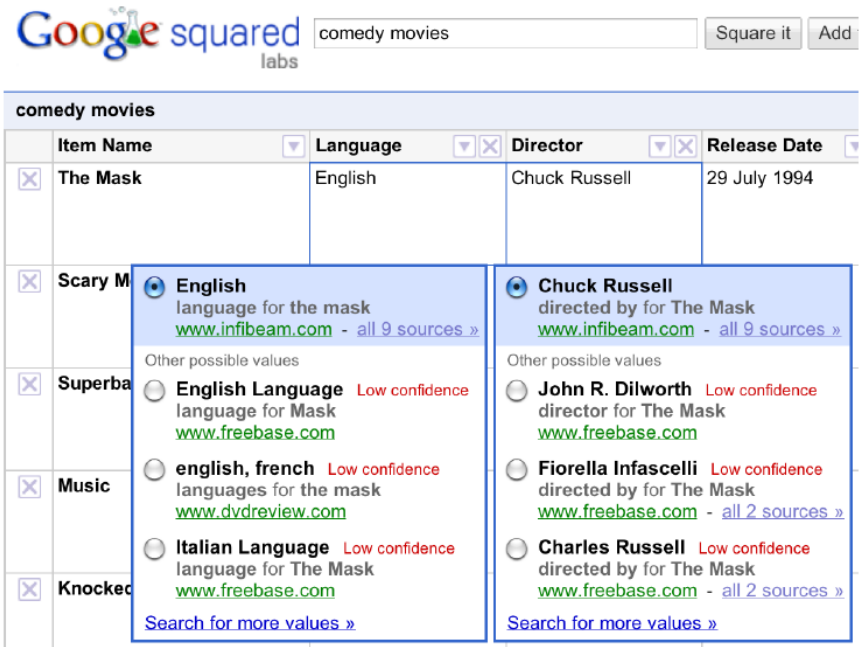
Understanding Concurrent Systems serves as a comprehensive introduction to the field, in addition to providing material for a number of more advanced courses. A first point of reference for anyone wanting to use CSP or learn about its theory, the book also introduces other views of concurrency, using CSP to model and explain these. The text is fully integrated with CSP-based tools such as Failures-Divergences Refinement (FDR), and describes how to create new tools based on FDR.

Most of the book relies on no previous knowledge of the theoretical background other than a basic knowledge of sets and sequences. Sophisticated mathematical arguments are avoided whenever possible.
Understanding Concurrent Systems. **A W Roscoe. ISBN: 978 – 1848822573**

Managing probabilistic data with SPROUT

Traditional databases are deterministic. Every record stored in the database is meant to be present with certainty, and every field in that record has a precise, unambiguous value. Today data management needs to include new data sources, where data are uncertain, and which are difficult or impossible to manage with traditional database systems.

Consider business intelligence (BI), whose goal is to extract and analyse business data by mining a large collection of databases. BI systems can be made more useful by including external data such as twitter feeds, or blogs, or email messages, in order to extract even more valuable business information. For example, by analysing blogs or twitter feeds and merging them with offline databases of products, companies can obtain early feedback about the quality of a new product, or its degree of adoption, such as for a new car model, or a new electronic gadget, or a new movie; such knowledge is very valuable, both for manufacturers and for investors. However, traditional database systems require the data to be precise: for each tweet the system needs to know precisely what product it mentions, and whether the comment is favorable or unfavorable. Also, the data must be cleaned before is can be used.



> ***"The goal of probabilistic databases is to extend today's database technology to handle uncertain data, where uncertainty is expressed in terms of probabilities of records or probability distributions over possible record values."***

When an SQL query is executed, the system returns a set of answers, and annotates each answer with a probability representing its degree of confidence. Typically, the answers are ranked in decreasing order of their output probability, so that users can inspect the top, most credible answers first.

The SPROUT system developed by Dan Olteanu's team understands probabilistic data. In collaboration with the Google Squared team, led by Andrew Hogue, they have developed a web system on top of SPROUT that can integrate uncertain yet dynamic web data with clean offline relational databases and can answer SQL queries over them.

The picture shows a tabular view over unstructured data collected and aggregated from public web pages about comedy movies. Each row represents characteristics of a movie, such as language, director, etc. For each field, only the value with the highest degree of confidence is displayed. However, if the user clicks on that value, then alternative choices are shown. For example, the most likely director of *The Mask* is Chuck Russell, but there are further possible values with lower confidence. Similarly, for the language, English is the most likely, but a few other values exist. Using SPROUT, one can ask for birthplaces of directors of comedy movies with a budget of over $20M by joining the web table for comedy movies (where we can ask for the budget) with some other external dataset(s).

This work is supported by the EU grants FOX and HiperDNO and by the EPSRC grant ADEPT. The SPROUT system will be presented this month in Athens at the SIGMOD conference.

The foundations of probabilistic databases are described for the first time in a book to appear in 2011. ***Probabilistic Databases. Synthesis Lectures on Data Management.*** **Dan Suciu, Dan Olteanu, Christopher Ré, and Christoph Koch. (In press). ISBN: 978 – 1608456802 (paperback) and 978 – 160845681 (ebook).**

And in other news...

- The paper 'Combining answer set programming with description logics for the Semantic Web' by Thomas Lukasiewicz and four former colleagues Thomas Eiter, Giovambattista Ianni, Roman Schindlauer and Hans Tompits from the Technical University of Vienna, Austria is currently one of the top 10 cited articles (published in the last five years) in the journal *Artificial Intelligence.*

- Members of the Department have co-authored nine papers (out of 55 accepted papers) at the upcoming International Conference on Computer Aided Verification. This premier event in formal verification is being held in July 2011 in Utah, USA.

- Oxford authors have given six papers at ICLAP 2011 – the 38th International Colloquium on Automata, Languages and Programming – one of the top conferences in theoretical computer science. 'Regular Languages of Words Over Countable Linear Orderings' co-authored by Gabriele Puppis received the best paper award.

- The paper 'Managing uncertainty and vagueness in description logics for the Semantic Web' by Thomas Lukasiewicz and Umberto Straccia (of ISTI – CNR in Italy) is currently one of the top 10 cited articles (published in the last five years) in the *Journal of Web Semantics.*

David Blunkett lectures on information security and privacy issues

The Department played host to a seminar by Rt Hon David Blunkett MP in March. His talk was part of a series on information security and privacy, organised jointly by Andrew Martin in the Computer Science Department and Dr Ian Brown in the Oxford Internet Institute.

Speakers in the series considered a range of topics from an inter-disciplinary perspective: from historical perspectives on information assurance through to privacy in smart metering, and information warfare. David's own career has given him a unique perspective on the balance of security and privacy in an open society, and in his talk reflected on some of these experiences. He urged us to take a long view and learn from history – starting with using appropriate language as well as the technical measures we adopt.

In common with several other speakers, David referred to Cabinet Office research showing the annual cost of cyber crime being £27bn, falling on individuals, businesses, and government. The figure was disputed and debated during the term, but a new-found emphasis on cyber security seems timely.

The seminar series continues in Trinity Term. Details are at www.softeng.ox.ac.uk/infosec, together with recordings of many of the talks. The objective is to develop a sustained inter-disciplinary programme of activities in information security and privacy – all are welcome.

Logic for Interaction

A workshop on 'Logic for Interaction', organised by Samson Abramsky and Nikos Tzevelekos, was held at the Department on the weekend of April 2–3. The workshop gathered participants from the ESF Eurocores LogICCC program 'Logic for Interaction – LINT', of which Samson's own EPSRC project of the same title is a partner.

LINT is a collaborative research project aimed at developing mathematical foundations for interaction. Intelligent interaction involves agents in complex scenarios like conversation, team-work, or games. Contours of a broad mathematical description are starting to emerge today, based on several individual research developments that now need to be brought together. The project gathers logicians, computer scientists and philosophers from six European countries in an effort to lay the foundations for a unified account of the logic of interaction.

The workshop held in Oxford in April was attended by researchers from affiliated research groups in Amsterdam, Gothenburg, Helsinki, Paris and Tampere. Local speakers comprised Samson and Nikos, who gave talks on high-level categorical methods for contextuality and non – locality, and on game techniques for reasoning about dynamic resources respectively. Moreover, Boris Motik gave an introductory lecture on description logics, while Mehrnoosh Sadrzadeh presented her work on using adjunctions in modal logic to reason about information flow. Other themes covered by talks in the workshop included the very active current work on logics of dependence and independence, with connections to database theory, complexity, generalised quantifiers and linguistics.

Further information: http://sites.google.com/site/oxfordlint/

Computing At School conference hosted

The Department recently welcomed over 50 students and teachers to the Computing At School Sixth Form Conference. The one-day event saw young people attending a series of lectures and practical sessions given by members of the University's Computer Science Department, as well as speakers working in industry.

The conference was one of about a dozen similar events across the country during (or close to) National Science Week that were supported by the Computing at School Group, part of the British Computer Society, The Chartered Institute for IT. The events were all designed to encourage students to find out more about the subject, and to consider studying at university and making a career in this exciting area.

Attendee Oliver Smith from Cirencester College described the Oxford event as 'both inspiring and reassuring.' He added: 'The students and I had a wonderful time and it is likely to have quite an effect on their future.'

To receive emails about similar events sign up at: www.cs.ox.ac.uk/mailinglist

Oxford speaker gives Roger Needham Lecture

The British Computer Society's Roger Needham Lecture was given by Joël Ouaknine at the Royal Society, London. 'Timing is Everything' explored the scientific challenges in devising effective methodologies for accurately modelling and analysing real-time computer systems (such as those in Satnavs and in-car breaking systems), in order to verify and guarantee that they function as they are intended to.

Wanted: more women in Computer Science

Encouraging more young women to enter the field of Computer Science continues to be one of the major foci of the Department. Current statistics indicate that although the numbers seem relatively low in comparison to the number of male students, Oxford is actually on a par with the rest of the UK.

The intake figures for 2011/12 in Computer Science indicate that approximately 18 per cent of places were offered to women applicants, resulting in five places. However, the corresponding statistics for Maths and Computer Science show only 10 per cent of places were offered to women, which equates to three places.

The figures for MSc in Computer Science female candidates have remained fairly static over the past four years with approximately 18–20 per cent of places being offered to women, whereas the number of DPhil female applications is on the increase. However overall the percentage offers to women candidates has remained static at approximately 14–16 per cent, which equates to between seven and nine places.

So what do the global statistics tell us? We know that Computer Science is still a male dominated subject. The major issue is getting girls interested in the subject early enough so that they select the right options for this type of study. Once women are interested in Computer Science they tend to stay within the subject.

What can we do to encourage women? One crucial way is to hold access and outreach events such as open days and regional conferences both within Oxford and around the UK that specifically target potential female candidates. This includes getting a level of interest from schools, teachers and parents alike.

The Department is currently involved in a range of such activities including the Women in Computer Science Day, which is being held in June. One of the key aims of this event is to dispel the various myths and stereotyping that surround Computer Science, in particular that is it only for 'geeks with no social life'. This event is extremely popular and is already over–subscribed. Is this an indication of a change in trend?

We've also recently offered an academic session as part of the University's Women in Science Easter Residential course specifically for 16–17 year-old females from Oxfordshire and the neighbouring counties.

The department is about to submit a case for an Athena Swan award. Athena Swan awards recognise that a department is working to promote gender equality within a particular discipline. In particular the award demonstrates the level of commitment that a department has to attracting women into science and also how best to support female academics.

preDiCT: advancing drug safety

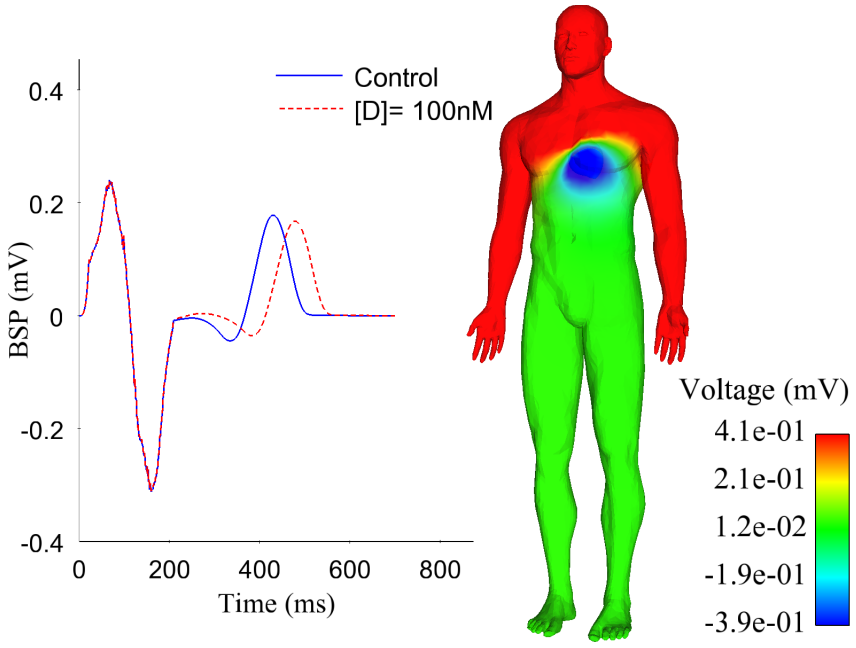
Regulators, academia and industry discuss the future of computational cardiac modelling to predict the safety of drugs.

The preDiCT project started in 2008, funded by the European Commission. Its main aim is to develop the computational infrastructure necessary for 'in silico' (performed on computer or via computer simulation) simulation, and ultimately accurate prediction, of the impact of drug candidates on cardiac electrophysiology.

A major objective has already been met – to enable the simulation of drug – induced effects on the human body surface ECG by performing simulations thousands of times faster than was possible at the start of the project.

A workshop was convened on 7–8 February 2011 in Oxford, bringing together the predict academics, regulators and industry researchers. The objectives were to discuss the feasibility of in silico modelling approaches, the achievements of the preDiCT project and to gather their advice towards assimilation of simulation into routine regulatory activities.

Further information: www.vph–predict.eu/



Multiscale computer simulations of drug-induced effects on the heart: from ion channels to electrocardiogram

Cloud computing project launched

A European consortium including computer scientists from Oxford University has announced Trustworthy Clouds or TCLOUDS, a new virtual infrastructure project. The goal of the project is to prototype an advanced cloud infrastructure that can deliver a new level of secure, private and resilient computing and storage that is cost-efficient, simple and scalable.

To evaluate the feasibility of the TCLOUDS infrastructure, scientists and engineers will use two scenarios including a cloud that is resilient and steers a smart energy grid and a cloud that stores and protects the privacy of medical data.

The first is a smart energy grid with Portugal's leading energy and solution providers Energias de Portugal and EFACEC: A combination of smart metering and a web-based real-time status and energy consumption control system enables public utility providers to monitor and efficiently control a public lighting network. TCLOUDS will show how such energy-preserving systems can be migrated to a cloud infrastructure whilst increasing their resilience, privacy protection and tolerance, from both hackers and hardware failures.

The second is a patient-centric home healthcare service with San Raffaele Hospital in Milan, Italy, that will remotely monitor, diagnose and assist patients outside of a hospital setting. The complete lifecycle from prescription to delivery to intake to reimbursement will be stored securely in the cloud and thus be accessible by the patient, doctors and pharmacy staff. The goal is to demonstrate how the quality of in-home healthcare can be improved cost-efficiently without reducing privacy.

Andrew Martin, the James Martin Research Fellow, and his team from Oxford are part of a £8.9 million EU-funded consortium of organisations including IBM, and Sirrix AG security technologies, Portuguese energy and solution providers Energias de Portugal and EFACEC and San Raffaele (Italy) Hospital.

'Today, data can be gathered everywhere and accessed by anything, but doing so doesn't come without some risk, including security and data loss',comments Matthias Schunter, technical leader for TCLOUDS and computer scientist at IBM Research – Zurich.

> ***'With TCLOUDS we aim to demonstrate that the rewards in terms of both cost efficiencies and smarter services, such as healthcare and energy, can be achieved by using advanced cloud technology to reduce or, in some cases, eliminate those risks.'***

Data is stored on remote hardware via the internet instead of being kept on a local server or computer.

To achieve the security, resiliency and scalability needed when outsourcing critical IT-systems to a cloud, scientists from the consortium will build an advanced *'Cloud of Clouds'* framework for the project. As the name implies, this framework will provide multiple back-ups of the TCLOUDS data and applications in case of a hardware failure or intrusion.

Newly designed security mechanisms will also be developed to remotely verify the security and resiliency of the cloud infrastructure, guaranteeing the integrity of a hardened cloud computing platform to users of cloud services.

Besides advanced technology, TCLOUDS will also study the legal, business and social aspects of cross-border cloud computing, such as country-specific privacy laws; writing cloud computing service agreements; and user-centric requirements, including languages and accessibility.

The TCLOUDS project is scheduled to be completed by September 2013.

And in other news...

- Bob Coecke's paper 'Quantum Picturalism' is in the top 10 for the most downloaded papers of the journal *Contemporary Physics.*

- As part of the Department's commitment to becoming increasingly environmentally friendly, the Computing Support team are applying for energy conservation grants to replace the last few remaining high-current-drawing equipment with more modern low-consumption devices and to expand the current virtual machine infrastructure.

- Congratulations to Christopher Broadbent, who has been selected as one of twelve 'laureate' postdocs of the Foundation Sciences Mathématiques de Paris. The Foundation brings together researchers in mathematical sciences, including fundamental and applied mathematics and fundamental computer science.