Oxford Computer Scientists Win Prestigious ERC Awards

Two professors at the Oxford University Computing Laboratory have been selected to receive highly competitive European grants for research projects that promise major breakthroughs in how we use technology.

The European Research Council has made Advanced Investigators Grants worth a total of €4.4m for the Diadem project, headed by Georg Gottlob, and for Veriware under Marta Kwiatkowska. Fewer than ten of these grants, intended to support pioneering research by leaders in their fields, have been made this year to computer scientists across Europe.

The projects will both last five years, and address challenging but vital research questions that will be central to our relationship with the technology around us over the coming decades. One will help ensure the electronic devices we rely on work properly; the other will help us extract the information we need from the ever-increasing volume of data on the internet by enabling computers to understand its structure on their own.

With **Veriware**, Kwiatkowska aims to make the theoretical and practical breakthroughs that will let us be sure that our technology is functioning as it should as we enter a new era of ubiquitous computing, in which information processing moves out of desktop computers and into the everyday objects all around us.

This kind of computing without computers has been called 'everyware'. The technology is still in its infancy, but examples are starting to appear – from Bluetooth mobile phones that automatically sense each others' presence and exchange information to fridges that can tell when the food in them is running out and automatically order more from the internet.

In future, there will be far more of this – we'll be surrounded by countless tiny computers monitoring their environment with electronic sensors and wirelessly sharing information with each other. The applications are manifold, from giving us access to healthcare and banking to controlling the environment in our homes.

But as things stand there is no rigorous way to make sure these embedded systems work as they should. Already there have been high-profile cases where expensive products have had to be recalled due to unforeseen flaws in the software embedded in them. Some of these faults have merely been expensive for the manufacturer; others, such as problems with cars' built-in computers controlling engine function or braking, could be fatal.

"The pace of technological change is accelerating,' says Kwiatkowska. 'We need a paradigm shift in software verification to let us deal with challenges posed by complex communities of 'everyware'." She specialises in what is known as 'model checking' – a technique that uses dedicated software to analyse a computer programme, model all the possible states it can enter and prove whether or not it will ever reach a specified situation, or how long it could take to reach it. The technology has been used to solve problems like determining the worst-case scenario for the time taken to transfer a given amount of information over a wireless network.

But model-checking a computer programme once before it is installed is a very different proposition to continuously and autonomously model-checking everyware that exists in a complex situation of continuous interaction with the

other devices around it, in an uncertain environment and without human input. Major theoretical breakthroughs will be needed to give us the confidence to depend on these devices, and, with Veriware, Kwiatkowska hopes to make them.

Meanwhile **Diadem** – Domain-centric Intelligent Automated Data Extraction Methodology – sets out to solve the problem of extracting complex, structured information from large numbers of websites. 'If we succeed, Diadem will be the next major step forward in web search technology,' says Gottlob. 'It will boost individual and corporate web users' ability to get the information they need from the internet.'

Traditional web search engines rely on looking for keywords on web pages. They work well when looking for some kinds of information, but struggle with more complex queries – typing 'restaurants near me serving pasta al pesto as today's special' isn't likely to produce useful results.

With Diadem, Gottlob aims to create software that can trawl through every website in a particular field – the property market, for example, or restaurants, or air travel – and pull out the information they contain in structured form. Equipped with a basic knowledge of the general principles its domain works on, it will be able to analyse each web page's low-level structure as it goes in order to extract the information it contains.

Humans find it easy to visit a new website and immediately grasp its structure and what the different elements on each page mean – which of the numbers visible is an item's price, for example, or how to interpret a timetable. But computers struggle with this kind of semi-structured content – they don't understand how websites are structured.

It's possible to teach computers to pull information out of particular websites, but this involves a human spending time showing it which parts of a page do what. This can work on a small scale but takes too much human time and effort when dealing with large numbers of websites; it is also confounded if the website being studied changes even slightly.

By the end of the Diadem project, Gottlob hopes to have built a system that can deal with a specified country's property market, analysing tens of thousands of estate agents' websites and presenting the properties discovered to users. The result won't simply be a web page with links to other pages that may contain relevant information, as with traditional search engines; it will be a structured dataset drawn from the data objects found on sites within the domain, which can easily be searched or further processed by other software applications.

Companies like Google, Microsoft and Yahoo! have already expressed interest in Diadem's results, which could lead to the next generation of search engines, going beyond the limitations of keyword searching.

Notes for editors:

Marta Kwiatkowska is Professor of Computing Systems at Oxford University Computing Laboratory, and a Fellow of Trinity College. Prior to this she was Professor in the School of Computer Science at the University of Birmingham, Lecturer at the University of Leicester and Assistant Professor at the Jagiellonian University in Cracow, Poland. Her research is concerned with modelling and analysis methods for complex systems, such as those arising in computer networks, electronic devices and biological organisms. The analysis methods that she investigates include simulation and formal verification, with particular emphasis on quantitative verification of probabilistic systems. Her work spans the whole spectrum, from theory through algorithms to software implementation and applications. The PRISM model checker (www.prismmodelchecker.org) implemented by her group is the leading software tool in the area.

Georg Gottlob is Professor of Computing Science at Oxford University Computing Laboratory, and a Fellow of St Anne's College. He moved to Oxford from TU Vienna, Austria, in 2006, where he worked since 1988 as a Professor of Computer Science. His move from Vienna to Oxford was facilitated by a Royal Society Wolfson Research Merit Award. He has been helping to build up the Information Systems Group at Oxford University's Computing Laboratory. Gottlob's current research deals with databases, web information processing, artificial intelligence and computational logic. He has done pioneering academic research on web data extraction that gave rise to algorithms and software now used by the Vienna-based company Lixto (www.lixto.com), which he co-founded. He is a founding member of the Oxford-Man Institute of Quantitative Finance.

The Oxford University Computing Laboratory is a leading research institute with an outstanding reputation in computer science, and has been pre-eminent in programming language research since the 1970s. Christopher Strachey, the first Oxford Professor of Computation, together with Dana Scott founded the field of denotational semantics of programming languages. Sir Tony Hoare, second holder of the chair, invented the axiomatic approach to programming language semantics, and developed the formal language CSP to specify the interactions of concurrent processes. Today the Laboratory is a vibrant research environment with expertise in many areas of computer science.

To find out more about the Diadem and Veriware projects, please click DIADEM and VERIWARE.

For more information, contact:

Georg Gottlob
georg.gottlob@comlab.ox.ac.uk
+44 1865 283504

Marta Kwiatkowska
Marta.Kwiatkowska@comlab.ox.ac.uk
+44 (0)1865 283509