

## From the Director

Well, it's over!

The Advanced Knowledge Technologies (AKT) Interdisciplinary Research Collaboration (IRC) began in 2000, and ran until 2006 with an extension to 2007. It's been an adventure, and an achievement that all concerned, our principal investigators, researchers and students, can I think be very proud of.

Some details: AKT was funded by the UK Engineering and Physical Sciences Research Council, and comprised the Dept. of Computer Science, University of Aberdeen, the School of Informatics and the Artificial Intelligence Applications Institute at the University of Edinburgh, the Knowledge Media Institute of the Open University, the Dept. of Computer Science, University of Sheffield, and the School of Electronics and Computer Science, University of Southampton. I am the AKT Director, while the Principal Investigators are Derek Sleeman (Aberdeen), Dave Robertson, Austin Tate (Edinburgh), Enrico Motta (OU), Fabio Ciravegna, Yorick Wilks (Sheffield) and Wendy Hall (Southampton). AKT research has been a genuinely interdisciplinary effort involving leading figures in the worlds of multimedia, natural language and computational linguistics, agents, artificial intelligence, formal methods, machine learning and e-science. Applications and related projects included researchers in areas such as bioinformatics, medical systems, aerospace and the environment.

The initial proposal was written in 1999, and much has changed since then. The motivation for AKT has remained constant – to enhance information and knowledge management in the age of the World Wide



Web. Information is a key resource of the 21<sup>st</sup> century, and the WWW is assuming a pre-eminent role as its global repository. Information is being amassed at such a rate that the challenge is how to publish it in a useful way – not just documents but underlying data sets themselves. However, the massive amount of information released must then present challenges: how to index and search, retrieve and reuse it, curate and maintain it. AKT has engendered a community that has been and is continuing to produce tools and techniques, systems and infrastructure to acquire, exchange, share, filter and visualise content, often but not exclusively, on the WWW.

AKT's research excellence is reflected in a wide range of indicators. The programme of research conducted under the IRC has been an important contributor to the development of the Semantic Web (SW), and to the creation of an interdisciplinary network of influential researchers. Each site has brought specific skills to AKT, and each site has evolved along with the research community within and without the IRC.

AKT pioneered the use of intelligent systems to support the SW, and the use of the SW to prime, build and drive intelligent systems. This research led to several intelligent tools and services for creating and maintaining content on the Web, and for collaborating in the creation of documents, datasets and other knowledge-based resources.

Another important line of work is that of creating the marked-up content in the form that the SW requires, using methods ranging from Natural Language Processing (NLP), ontology-mediated knowledge harvesting, detecting duplicates and ensuring referential integrity, automatic and semi-automatic markup and tools for aiding annotation (including multimedia).

AKT has also developed important pieces of the infrastructure for the SW, at all levels, from scalable data warehousing technologies to interfaces, including methods and tools for searching and browsing on the SW, and tools to find Web content and also to make sense of the content retrieved. The methods developed have and will also be applied to other large knowledge stores, including multimedia and organisational archives, facilitating the reuse of knowledge bases.

One particular strength of AKT's involvement with the SW is the development of tools, methods and techniques for working with ontologies – often regarded as a potential bottleneck for the development of the SW. AKT's work on ontologies includes methods for search, evaluation, mapping, merging, maintenance, modularisation, fragmentation and pruning.

The use of NLP techniques on the SW, and to manage large knowledge repositories, has been central. Tools have been developed to mediate analysis of texts using ontologies to produce filled templates and services such as annotation for information extraction. The scale of the content available on the Web has allowed us to go beyond user-centred annotation, towards unsupervised and semi-supervised extraction.

Our increased understanding of the dynamics of information has led to the development of methods and tools for describing and manipulating the information

lifecycle, notably a lightweight coordination calculus, which have fed through into various real world applications. This strand of work has also led to notations separating control, process and domain knowledge.

AKT members are central to the development of SW services, with for instance an AKT-developed system (IRS) as a reference implementation for the emerging Web Service Modelling Ontology (WSMO) standard for SW services.

Another important strand of AKT research involves using SW information to induce human, institutional and organisational relationships. AKT tools have been developed to identify communities of practice of individuals, to locate expertise and to associate individuals with their expertise.

One key measure of AKT's success is the gearing on the funding. 36 other projects have been attracted to AKT partners to perform work in the space of advanced knowledge technologies. The total amount of funding secured for AKT partners is over £25m. The average funding for AKT partners per project is around £700,000. We have amassed some 89 partners in these projects. 25 of these are UK institutions; 81 are situated in the EU. 21 of these projects are follow-on projects from AKT, carrying on AKT's work beyond the lifetime of the IRC, which together total £19.5m for AKT partners. Six of these projects realised £1m+ for AKT partners, and a further eight realised between £0.5m and £1m.

AKT has put a considerable amount of effort into making links with industrial partners in order to demonstrate the relevance of the technologies which AKT has developed. There has been a strong relationship with Rolls-Royce throughout the project which led to 3 significant testbed systems being implemented. The deep relationship fostered by the success of these testbeds led directly to the DTI-IPAS

project jointly funded by Rolls-Royce and the DTI involving Aberdeen, Sheffield and Southampton. The project is applying natural language, ontology management, and Web services techniques developed in AKT to help make information produced during the routine processes of engine repair and maintenance, available in a format in which it can be readily accessed and understood by designers developing the next generation of engines.

We have also collaborated with many other partners, including the Semantic Web research group of Hewlett Packard Laboratories, at Bristol and the Environment Agency (Bath). Applications of the adaptive Information Extraction technology developed at Sheffield have been used at the Lawrence Livermore National Lab and Boeing for supporting automated document annotation. And we have also been part of a US Defense Advanced Research Projects Agency (DARPA) team trying out novel technologies to support multinational forces in peace-keeping missions where diplomatic, economic and other non-military interventions might be more appropriate. Many others could be highlighted.

One more indication of AKT's industrial relevance is that there have been several AKT spinoff companies. I-C2 Systems ([www.i-c2.com](http://www.i-c2.com)) is evaluating the use of I-X and other task support technologies, while Level E Limited ([www.levelimited.com](http://www.levelimited.com)) has been formed with the aim of delivering to the investment management community state of the art, smart decision-support and decision-making technologies suitable for high-frequency trading. Garlik ([www.garlik.com](http://www.garlik.com)) is a startup with involvement from AKT staff to deliver services designed to aid trust and privacy. It specialises in consumer products and services to put people and their families in control of their own digital information. This has exploited a great deal of experience in AI,

and in particular exploits the possibilities for extracting information from large and heterogeneous sources using techniques pioneered in AKT. Garlik received first round Venture Capital funding of £3,000,000 and currently has 52,000 registered users. Garlik has used many of the open source AKT technologies (such as 3store) in a real-world, large-scale problem space, and has therefore been able to show that the technologies do stand as scalable and practical solutions to many of the problems of information management

AKT produced almost 400 publications during the six year period of funding, ranging from books and journal articles to international and national conference papers. We have published in all relevant major journals and conferences. Four collections of *Selected Papers* highlight over 100 refereed papers in premier journals and conferences in total. Our researchers, papers and systems have won several awards over the period of the project; many of us have been invited speakers at prominent events, and we have organised a number of high-profile conferences. Numerous AKT personnel sit on the editorial boards of noted journals, while AKT members have contributed to relevant standardisation bodies such as the W3C and OASIS.

As I noted earlier, during the course of AKT there has been a tectonic shift in the landscape of our understanding of knowledge and its management. The IRC created an integrated research environment which allowed resources to be reconfigured and repurposed, and for assumptions to be challenged, in order to adapt to this new environment. This was the foundation for a great deal of added value for the consortium, which was large enough to be an influence on industry, and to create a cohort of young researchers who were at home in this landscape. Brand recognition has amplified



each partner's effort and visibility, while the six-year timescale allowed a longer and more sustained attention span. We have been able to attract important researchers to our sites, and to recruit top postdoctoral researchers, while we have been able to provide a stimulating environment for our DTA students. AKT funding has enabled a critical mass to be reached having a dramatic impact on the worldwide research community.

The experience of the IRC concept was that the synergy between partners, the interdisciplinary ethos and the level of resources delivered much more value than 3 or 4 individual projects funded each year to separate Universities. Many ideas were circulated and created; many were followed and returned the investment (e.g. annotation for the semantic web, hybrid searching, architectures for language engineering, photo annotation, etc.). It is very important to stress that the financial stability gave the possibility to approach very challenging and exploratory research.

One perhaps unanticipated effect of AKT's critical mass is that it is much more feasible to allow small projects to fail, thereby promoting a greater culture of risk-taking, and the learning of lessons. Failure can be an important driver for science; of course some failure is normal and indeed inevitable in scientific research, but the lessons that failures can teach us are often lost because research communities break up and new research is not stimulated. The AKT IRC created a very close community in which failures could be analysed and the lessons applied to new areas of development.

There is no doubt in my mind that the IRC experiment has been a good one, and, talking to my fellow IRC Directors, it seems that that feeling is widespread. There were also significant collaborations between IRCs. Each IRC evolved its own distinct identity and

culture of work. It is perhaps timely for an event to enable Directors and researchers to share the insights and lessons, the products and results of research. As IRCs together we should present the benefits and outcomes to a wider audience nationally and internationally.

Indeed, it is worth noting that a recently completed international review of UK ICT commissioned by the EPSRC (<http://www.bcs.org/upload/pdf/ICT-International-Review-Final-Report.pdf>) found that the IRCs "have been very successful", and have "been excellent in fostering large collaborations and the creation of infrastructure". The panel, supported by the IET, Intellect and the BCS, was chaired by William Lorensen of GE Research and comprised 14 leading ICT international experts. They recognised the added value not just of individual IRCs but of their collective contribution to UK research and its international standing. They stated "The UK ICT community is doing exciting, high quality, world class research across the spectrum of CS and EEE. The Panel agreed that the UK is in the premier league in many research areas. The UK, through its Interdisciplinary Research Collaborations (IRCs) and e-Science programmes, has made great strides in interdisciplinary research and in some areas this is attracting new talent into the ICT community."

So my final act as Director is to thank the EPSRC for its creative focusing of funding, and the development of a style of collaboration that has enabled us to change our research world and potentially the wider world.

Professor Nigel Shadbolt