The evolving internet

Anni Rowland-Campbell – March 2008

The internet has profoundly changed the way we live and work, much like the printing press once did. As with all eras of radical change and innovation it usually takes approximately thirty years for a new technology to become socialised and far longer for it to become imbued within the human psyche, but the internet, and the World Wide Web which sits on top of it, has now become not only part of the modern lexicon it is even beginning to pervade the air we breathe.

When Tim Berners-Lee launched "The Web" to the world he did so because of the opportunities afforded by the internet. He created a medium and environment within which humans could publish, share and search for information at almost instantaneous speeds and with increasing usability in terms of navigation. Web 1.0 was just the first step and, as Web 2.0 evolves the real potential of the internet is beginning to emerge. However, for the machines themselves the unstructured data that pervades the web is both confusing and unclear. Machines and artificial intelligences need clear instructions as to what to do and are unable to "infer" as we humans do. Thus their ability to contextualise and attach meaning to different bits of information as decreed by the whims of human need is limited. Enter the Semantic Web.

Berners-Lee envisioned that the web would evolve to become for data what documents are for humans by a series of iterative tagging and guided by agreed ontologies. If you and I and a friend spoke three languages we would need eight different dictionaries to translate a conversation, and even then the nuances of colloquial language would probably be missing. For humans and machines it is the same - we need agreed terms and definitions about what we are discussing, and these need to evolve with time.

How real is the need?

For many organisations the reality of their information and data systems is a cacophony of legacy systems usually disconnected, unable to communicate with each other, and their design and management often driven by politics and organisational cultural environments rather than bigger picture strategic goals. Whilst the amount of data and information we have and produce is increasing exponentially our knowledge, let alone wisdom, remains as always - largely tacit and in peoples' heads. The generations of knowledge management systems that have been developed have only added to the confusion and have sucked resources from IT budgets in the hope of finding the “magic bullet”.

For many organisations, and especially for government, the twin challenges of records management and an ageing workforce (which means knowledge retention and management) are amongst the most pressing problems and, with a population increasingly expecting access to information, the need for privacy and the "right to know", this poses a real quandary. eGovernment - in its many forms - dominates much of the political agenda and open systems and social networking are becoming influential in campaigns and governance.

The dream solution would be to know what is known, to be able to ask the right questions and finally to get the right answers at the right time, in the right place and about the right things, now. This is the promise of the Semantic Web, which is slowly and steadily becoming a reality.

Reuters have just announced their plans for "Calais" (http://opencalais.mashery.com/), a semantic web application that will facilitate search as envisioned by Berners-Lee within the domain of Reuters' knowledge base. Although not fully functional as yet Calais has put a stake in the ground for all media and publishing organisations, and pushed the "content" agenda a step forward. It matters not how one consumes the information - be it print, iPod, PDA or visually imprinted retinal projection. The value is in aggregating the data and information to produce true knowledge of value to a particular person at a particular time. The Calais system once perfected could do the same for any publisher with content regardless of what that content is. With the consumer being able to access it when they want it how they want it the entire publishing supply chain changes, and with it the digital revolution turns another notch.

There are two approaches to developing semantic technologies and the true solution lies at the mid-point between the two. For many within the IT world there is the bottom up approach of tagging the data and meta-data then building the ontologies with which to describe them on the top. Once these ontologies exist then "semantic interoperability" can be achieved leading to new services and applications, and finally to new organisational business models.

The flipside of this is the tip-down approach, of starting with the end user and using processes such as Topic Maps to identify agreed areas of interest and then interrogate the data according to these taxonomies. This is where the Semantic Web becomes a creature of philosophy, psychology, sociology and politics - the human world.

And, the gap in the middle? Herewith the true challenge. When asked most IT researchers are fairly confident of the machines being able to do their bit and build both intelligent agents and learning systems based around RDF and the other standards to interrogate the data. But they all hesitate when it comes to the human domain. One senior researcher who sits on the W3C Standards Committee with Berners-Lee believes that the Semantic Web could be a reality in 3 to 5 years dependent on three things.

Firstly, there needs to be a major education campaign undertaken of senior managers (both IT and other) within organisations of the potential value of semantic technologies so that adequate funding is allocated to develop them and political will is put behind the agreement to develop and adopt standards.

Secondly if pilots are undertaken they need to be properly scoped, thought through and given the right resources - time, people and budget - to have a realistic chance of working. Calais will be an interesting test case.

This is because of the third potential stumbling block, that of expectations. Many a technological innovation, and particularly those within ICT, are undertaken with unrealistic expectations of both the people and the technology. These technologies are evolving and changing very quickly, but, like any true work of genius, they almost always need time to be conceived and mistakes to be made so that true learning and development can occur.

It has taken us as a research team almost two years to get to the point where we can explain this coherently and cogently to others, and to begin to see the implications and opportunities. Yes, there will be semantic applications which you can buy out of the box, but the biggest challenge is and will be having your organisation ready to utilise them. This means starting now to get your data and information in order, developing ontologies, ensuring you are employing the right people and, in many cases, looking critically at your current organisational structures, business models and marketing strategies to leverage the opportunities the technologies will bring.

This is a longer term process and will take any organisation a number of years to complete, just in time for semantic applications to reach the market.

Our research is showing that semantic technologies are very real and they are coming. They are a logical extension of everything that has been developed in the information technologies field and whilst the “Turing machine” may be still a way away the semantic world is coming, now.

If you are interested in finding out more about this project please contact Anni by email at Anni.RowlandCampbell@aus.fujixerox.com

Anni Rowland-Campbell is Project Manager of the “Semantic Technologies Research Project” at Fuji Xerox Australia. She is currently a PhD candidate at the Open University Business School, UK, investigating innovation and knowledge within the graphic media industries.