THE SEMANTIC WEB FARI F MARTIN AND SIMON KENT EXPLORE A NEARBY PARALLEL UNIVERSE WHERE INFORMATION EXCHANGE MAKES SENSE. **GRAPHIC DESIGN BY LIZ** TURNER

We should have robotic butlers by now. Hovering intelligent robot butlers, a global village, and computers that don't crash. Predicting the real world impact of technical trends is a dodgy business to be in. Wave after wave of buzzwords, acronyms and technology can numb us into suspecting that we'll never see the promised goods. No robot butlers or intelligent agents, no seamless chattering of machine to machine, filtering, pushing and pulling factoids for their human overlords.

Techno vision statements all boil down to this - at once vacuous and inspiring: everyone and everything connected. Machines that work with other machines without swearing, expense, fiddling. People working with other people online and off without their gadgets and software getting in the way. Despite the cliches, despite the dot-com crash, despite the hype, it is happening. Intelligent agents, push, smart filters... long promised, rarely encountered. So what's different this time? What's happening?

The looming future of the Web is one that echoes its grassroot origins, as an information system built for knowledge sharing. By using a new machine-friendly Web page format, Resource Description Framework (RDF), we get to play a clever trick. Whatever we want to say is atomised into a collection of independent 'triples', each telling a tiny part of the story: Naomi aged 28. Naomi nickname 'naomibee'. Naomi arranged a Party. Announcement describes Party.

This feature of RDF is more important than its outward appearance as XML. Even though XML is the more widespread of the buzzwords, the architecture of the Semantic Web depends more than anything on the principle that RDF documents can be interpreted as simply "saving things about the world". What do they say? Just the sum total of whatever all the triples say. Unlike HTML, RDF's design allows information to be merged more or less just by adding together collections of triples. Software tools can load these triples from a scattered set of documents, and reach conclusions based on the merged collection.

RDF, along with other widely-agreed formal Web standards, promises to transform the Web into a *Semantic* Web; a Web whose basic data formats and protocols can be cheaply shaped to echo human concerns, relationships, priorities.

Communities who want to use the Web to describe things just need to decide on the kinds of things, and kinds of inter-relationships they care about describing, and then write a special kind of Web document (a schema or 'ontology') that tells machines about their concerns. We can't expect machines to deal well with the elaborate narratives we share in HTML documents. Machines want something simpler. RDF triples inhabit a simple cartoon world. Ken loves Barbie.



The Semantic Web of these interconnections grows as the set of RDF documents grows. Each new RDF-based data format (for news, for geographic information, for personal data, for webs of trust) creates opportunities for unexpected overlap and re-use. Each new page containing such information is potentially talking about things described in more detail elsewhere in the Semantic Web, increasing the density of interconnection. The more information available in RDF, the richer the search space becomes. By providing a vendor-neutral platform for sharing simple cartoon descriptions of the world, the W3C XML and RDF languages give smart machines a common basis for talking to each other on the Web. This creates an environment in which the remaining problems are largely social rather than technical: agreeing on the kinds of things, and kinds of inter-relationships that these machines should be designed around.

Hoodwinked by futurists, distracted by gadgets, we forget the obvious. The network was built by and for people, and it gains meaning only by making a difference in the web of our lives...

0700 NEWS

Naomi Bishop wakes and washes. Having done so, she makes herself breakfast. As she eats, she wakes her laptop from its power-saving mode and reads her mail and customised news feeds. The feeds consist of headlines, published by thousands of independent organisations around the world, and link to more specific articles. Naomi's chosen feeds are collated automatically by a small application on her laptop. She selects a few items that seem interesting, and skims them briefly. Naomi is a musician by profession, living in London. In her spare time, she releases music to the network for free - under a pseudonym, to prevent her record label finding out, and prosecuting her under the recent and stringent laws on copyright. In her mail this morning are plaudits for her latest piece from as far away as South Africa and Brazil. She likes to reply to her fans, but she can't reveal who she really is: so, she logs into a cryptographic proxy and replies from her alternate "face". Naomi likes to think of it as a *nom de* querre.

RSS 1.0 is the best known application of RDF. It allows syndication of news feeds such as the BBC's current headlines, available in XML. The original RSS 0.9 was an RDF application. There is also RSS 0.91, a simplification or bastardisation depending on who you talk to

Streaming news feeds can easily be browsed with applications like Amphetadesk or Netnewswire, Semantic bot interfaces to RSS feeds can pop up when they have new content filtered to your timespace and location, friendship networks, or personal interests. A lot of RSS development is being led by the blog community as they look for alternative channels to distribute their

content, and connect to each others' blogs.

http://purl.org/rss/

http://www.disobey.com/amphetadesk/ http://ranchero.com/software/netnewswire/

Public key cryptography allows secure communication and authentication, based around individuals' "key pairs" - a oublic and a private key.

The two keys are simply very long numbers, usually made up of 1024 or 2048 bits (or characters) of data, giving hundreds of millions of possible keys.

The keys have a mathematical relationship, such that if either key is used to encrypt a message, the other can be used to decrypt it. Currently, it is effectively impossible to "guess" the counterpart key or to derive it from encrypted data.

An individual's public key is made available to other users. The individual can then encrypt or add an encrypted signature to a message, using their private key. Any recipient of the message can decrypt or verify the message using the individual's public key, proving that the individual sent the message.

http://www.pgpi.org/ http://www.gnupg.org/

COFFEE

Feeling like having a walk somewhere in town. Naomi sticks her laptop into her bag, hops on a bus and jumps off somewhere at random a little while later. It's not a district she knows. Drawing out her laptop, she engages a sniffer app. As she walks, she keeps an eye on the onscreen meter, until it slides into the green zone: someone is broadcasting. She has bandwidth. Naomi sends a query to whatever node it is that she's connected to. It consists of two parts:

1) Who are you? 2) What can you tell me about your locale?

The node replies:

I am "oval". My coordinates are 534736E 183379N. I know...

then it rattles off a string of URLs. Naomi adds oval to her bandwidth bookmarks and then turns her attention to the URLs. Browsing one, she's shown a heap of information about where she is - local facilities, culture, shops, food venues... that's what she wants. She follows the links, crossreferencing with a map site, until she finds a cafe - indicated as having bandwidth - that's near her current location. Community reviews of the cafe on the web seem favourable. She moves.

Personal search engines, for individuals or small groups, become more viable and useful than massive global databases like Google. Such indexes gather content indiscriminately and rely on rule-based processing techniques, rather than human intent and human feedback, to make connections.

Something like a peer-to-peer network between small semantic nets, arbitrated by personal semantic bots, appears. If a local node doesn't come up with satisfying results, it can query other local sources, route information about *semantically*, based on what it knows about the context of the request. The need for central storage, central control of information is obviated.

In a world of public/private keypairs. correspondents can also encrypt a message with an individual's public key, meaning that only the individual can decrypt it using their private key. In this way, secure one-to-one communication can be achieved.

For extra security (effectively, to stop people creating keys in others' names), individuals can band together, and sign each others' keys to form a web of trust.

In effect, signing someone else's key with your own is stating that you know and trust the owner of that key to be who they say they are. The more known signatories a key has, the more another user can trust that key, and those of the other people inside the trust web.

http://rdfweb.org/foaf/

1030 NETWORK



As they talk, an idea emerges.

We should have a party.

When?

How about tonight?

Let's do it. I'll make the arrangements, you let people know.

Naomi types out an announcement: Party tonight. Stay tuned for further details. Spread the word. Then she kicks it out into her network of friends to go forth and multiply. First she flags the message for wide propagation, then connects to the crypto proxy; from there, she sends it to people she has listed in her friend-of-afriend - FOAF - information as "trusted".

What happens next varies from person to person. Some recipients will have her listed as trusted enough to immediately auto-redistribute it to their trusted friends. Other recipients, marginal contacts, will not. Some may choose to repeat the process and redistribute the announcement; others may not. Each time the message is transferred, only the immediately previous sender's name remains attached; reacting to threats of legal action, the network has adopted the cellular technique, familiar to generations of members of resistance movements. As time passes, news of the party spreads faster and faster, fanning outwards from every recipient.

FOAF, friend of a friend, is an RDF vocabulary for describing people and relations between them; their mailboxes and nicknames, images that depict them, projects they have worked on.

FOAF provides a grounding in the human network, where related media files, descriptions of events and reports about things happening in time and space can come together and be related back to real

FOAF describes a friendship network more than a trust network. in a free software kind of way, you expose yourself by describing yourself on the net. but also protect vourself and your network from semantic attack by staving open.

Foafbot, a bot used to interrogate the friendship network for relevant information, supports encryption and signing of files and masking of email addresses as a basic self-protection from spam attacks, etc.

The semantic web is not limited to descriptions of information as a 'resource' on the internet. One's own presence becomes a resource for others to use.

1130 BICYCLE

Naomi decides she's tired of public transport and resolves to get a new bicycle to replace her last one, which was stolen from her a few weeks earlier. Firing up a search bot, she tells it to dig through the web for people in London selling second-hand bicycles. In a short time it returns with a long list of names and web addresses. She drags the list into a separate data window, and instructs a filtering bot to remove all the names whose associated sites don't have FOAF data - it shrinks by about a half. Activating her FOAF browser, she drops the list in. Connect to the sites, she instructs it, and download these people's information. Then she tells it to try and find a connection between her and any of them She specifies a maximum of five steps. The browser then traverses the tree of each of the downloaded FOAF files, following the see-also links, downloading yet more people's files, trying to find anyone who knows someone who knows someone... who,

eventually, knows Naomi.

She lucks out. After a minute or so the browser chimes, and displays a window with a diagram of available sellers, generated on the fly and providing links to further information on each. One of the names on the seller list is three connections away from her. She doesn't have to go through all this, of course, just to buy a bicycle; however, by searching for someone with a connection to her she drastically reduces the chances of being ripped off by someone random.

Naomi stores the potential seller's information, and the information from the people connecting her to him, in a temporary cache; if he turns out to be reliable, she will add it to her personal database, her FOAF file: good connections are a thing to be cherished. She transfers his cellphone number into hers, and calls him; a moment later she's on the move, homing in to do the deal.

SVG, scalable vector graphics, is an open, XML-encapsulated alternative to Macromedia Flash. Vectors in an image are marked up in XML, and it becomes possible to apply metadata in a granular way to regions of an image, connecting parts of diagrams back to the outside world.

Adobe have also developed XMP, an RDF application providing a metadata format for documents, images. Acrobat PDFs and images in Photoshop 7+ already have XMP data embedded directly into the file.

1400 WAREHOUSE

As she's securing her bicycle - heavily - to the lamp-post outside her flat. Naomi's phone rings. It's her friend, the one who was supposed to be organising a location for the party, and he has good news. Some contacts of his have managed to procure access to an empty warehouse somewhere south of the river. (He's actually there now. They have electrification; amplification is on its way.) It's all coming together. Back inside, Naomi repeats her earlier actions and sends out another announcement. This time, she says that there's now a definite location, and tells people to watch out for a final announcement at about 6pm.

1700 MUSIC

Whilst she gets ready for the evening: Naomi decides to put on some music. She wants something new, something fresh, but she doesn't feel like putting any effort into looking. So, she lets her bot do the work for her. The bot has learnt her tastes by scanning the metadata of the music in her collection, and can extrapolate accordingly. She sets it to work searching FreeNet, and fairly shortly it finds something new and interesting. She cranks it up.

EVENT

The abandoned warehouse is abandoned no longer: it's packed to the rafters. Naomi's ripping it up, and the crowd are loving it. The live sound that they're hearing is piped to wireless network transmitters at the venue: these transmissions then reach out across the city streets, by means of "cantennas" - cheap, homebrew waveguide antennas fashioned out of reworked soup cans. These antennas propel the broadcast into range of more permanent nodes at other locations, from where the audio can be redirected to anywhere. Copies generally find their way onto FreeNet.

 $\bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet$

 $\bullet \bullet \bullet \bullet \bullet \bullet \bullet$

It's not only the organisers that are broadcasting. Members of the crowd at these events as often as not have their own laptops with them, with wireless cards enabled, walking nodes. As the crowd gathers, the nodes become aware of each other, and a temporary, autonomous network forms. Data streams through the collective from point to point, negotiating an outwards path: some members of the swarm have brought their own antennae, multiplying the bandwidth available.

> Wireless networking, currently encapsulated in the 802.11b standard, allows full network connections of up to 11 Mb/s. Public access "nodes", which allow you to connect to the internet on the move, are already springing up in many cities. Some are free, some require payment.

> A node on a wireless network - whether community or commercial - can offer information about itself, its environment, the people connected to it, as part of the semantic web.

Many in marketing dream of location-based, personalised agents' which will read your mobile phone or PDA data and announce their services as you pass by. Using semantic web technology and open, self-protecting personal description data, it will be possible to have filtering whitelists (implementations of which currently exist for email) to protect yourself from the intrusion of unwanted evangelistic sales agents.

Bots have their origins in chatterbots like 'Eliza' and 'Alice'. and later web bots, such as the 'spiders' and 'crawlers' used by search engines to index pages. They have begun to evolve into conversational interfaces to knowledgebases, more sophisticated versions of "Ask Jeeves", such as the IRC-based Infobot. Such knowledgebases also feed back into the web, via SOAP or REST services.

When bots meet the semantic web, everything gets a bit better. They can begin to provide a conversational, stateful interface to all sorts of web services, expressed in shared dialects which all bots can understand. Already most bots can understand and process RSS newsfeeds.

> http://www.ask.com/ http://www.infobot.org/

2400 POLICE

No-one is sure where the word comes from, but it spreads fast through the ad-hoc networks. The police have heard about the party and are on their way. Within five minutes, everyone in the place has heard, and escape routes leading away from the site are being co-ordinated and shared from machine to machine. When the first vans show up 20 minutes later, the warehouse contains nothing but a few carefully-gathered piles of litter.

