The next wave of the web

Web gurus and geeks descended on Edinburgh, UK, this week for www2006. Chairing the panel 'The Next Wave of the Web' was Nigel Shadbolt, an artificial intelligence researcher at the University of Southampton, UK, and deputy president of the British Computer Society. Declan Butler asks him about the Web’s progress.

You've spent the week at www2006. What developments got you most excited?

The biggest theme is certainly what David Brown, chairman of Motorola in the UK, described as "the device formerly known as the mobile phone". The imminence of wireless broadband for mobiles means we are about to enter the phase of mobile and ubiquitous computing. It is also going to bring the Internet to the hundreds of millions of people who have no Internet access. We are talking now about creating mobile phones — devices — for $15 dollars apiece.

Another big theme was social software and the whole Web 2.0 movement. It’s this idea that by bringing lots of people's eyeballs onto tough problems you can generate large amounts of interesting activity. Companies like Amazon are now using automated software to farm out paying computing tasks. The Web allows you to build a social workforce.

Does that have an impact on science?
It wasn't so long ago that people were reluctant to put their data out there on the Web. But mashups [see 'Mashups mix data into global service'] and so on have shown that when they do, they unleash huge benefits and growth of activities. This is the whole open-data argument.

Another issue that came through strongly was e-science [the idea that research will increasingly be done via huge collaborations using shared datasets online]. Tony Hey, corporate vice-president for technical computing at Microsoft, was here talking about that.

Your background is in artificial intelligence. How is AI fitting into the Web?
I did my PhD here in Edinburgh in the late 1970s. We had interesting problems in trying to emulate human expertise and knowledge acquisition. But we couldn't get network effects going like what is happening on the Web. One of the problems of AI is that we've often been trying to do too good a job of emulating classic inductive reasoning; we've picked problems that are too hard. So AI hasn't really delivered on providing sentience in a box.
But, though most people don't realize it, the Web is already full of knowledge-intensive [AI] components. The Web is a brilliant place to get AI out there. Take Bayesian methods, a branch of statistics that allows a machine to make decisions, shifting probabilities based on its past knowledge and experience. Just because it's based on statistics, people think it must have come from statistics labs, but it came out of AI labs from our interest in reasoning. And Bayesian techniques are used all over the Web.

Like my wonderful open-source Bayesian spam filter, 'SpamBayes', which after a bit of teaching learns to filter my email almost perfectly. Absolutely.

The idea of a 'semantic web' — this notion of adding machine-readable tags to web pages so that a computer can read and 'understand' the text and data — has been around for years. But, like nuclear fusion, it always seems to be 'just around the corner'. Is it ever going to happen?

People at the meeting were asking the same thing; "It's been years now, what's really happening?" People joke that the semantic web is a refuge for tired old AI researchers. But I think we saw at this meeting that the semantic web is starting to happen. Perhaps we just need to think of it differently.

The semantic web has traditionally looked at very complex ways of expressing meaning, but for the web even adding very simple meaning can vastly improve things. Take RDF [Resource Description Framework: a simple ad-on that gives three properties about an online document or object (see 'From XML to RDF')]. Even simple RDF can basically characterize key classes, and relationships of interest, of components of data. Just put some RDF around your data and suddenly you have much more interoperability among machines.

You don't want scatter-gun search, you want rifle-shot results. You have to get inside the content. Google does a great job in a world of unstructured data, but if those data become more structured, it would do an even better job.

Will your average Jo Scientist or Joe Public ever bother adding RDF to their data and web pages?

This is interesting. The people in biology are really active because they have well-structured and described data and terms. They are now writing tools that will generate the RDF; with time these tools will get easier to use.

People using social networks, like FOAF [The Friend of a Friend Project], are already adding RDF about themselves almost without knowing it, when they enter details about themselves. But there are not enough tools to allow you to do it without thinking.

What do you think will be the hot topics at next year's meeting, or in five years?

Devices. We will be seeing real flexible displays. There are also lasers now coming out that project displays into mid-air. Computing power, hardware and Moore's law are the big drivers. Think about it — which has provided the bigger advances: the fact that I now have one million times more power on my desktop than when I was a student, or better algorithms?

Visit our newsblog to read and post comments about this story.