



Giant leap for the search unknown

Beverley Head asks the experts how close the world's web-trawlers are to a "Google-killer" breakthrough.

YOU fancy Thai — but aren't in your neighbourhood. So you get out your mobile and speak your request into the always-on search engine, which responds with the name of a restaurant your friends raved about last week, displays directions and a review, and also confirms that your preferred vegetarian options are available.

Welcome to smarter search, perhaps just five years away.

Dr David Hawking, science leader in information retrieval at CSIRO and chief scientist of enterprise search company Funnelback, says the future of search is all about personalised responses, tailored to your location, preferences, and social network.

These search tools will explore everything: personal files, emails, images and videos; corporate data; freely available information on the web; and

subscription databases.

If that sounds like an invasion of privacy, never fear. The personal preferences would be built at the user end, without sending too much private information to a third-party search provider.

Dr Hawking is supervising a PhD student at the moment exploring this idea of "personal meta-search" and so far the results seem promising, although Dr Hawking says a product may be a way off.

Even without that tool, there are already more than 1700 search engines in the marketplace.

Some offer global search, such as industry giants Google, Yahoo! and Baidu (which dominates China's global search market), and emerging players such as US-based Powerset, due to launch this month, and Sydney-based Lexxe, which will launch its beta version natural language search engine in

October.

Meanwhile Wikipedia founder Jimmy Wales has established start-up company Wikia, which will use a combination of computer tools (one of which is Grub, a web-crawler that emerged from Looksmart) and humans (which have proved adept at creating content for Wikipedia so should be equally adept at interpreting content) to sort and search information. In what might appear tit for tat, Google is also now exploring how wikis could be harnessed to refine search results.

Besides general purpose search there are enterprise search engines such as Funnelback, which is used by the ABC, ASX, Westpac and the Australian Government Information Management Office (AGIMO), that can be tuned to the enterprise's needs so that search algorithms rank policy docu-



ments more highly than, say, a blog and employees are permitted access to information on an as-needs basis.

Both global and enterprise search are developing rapidly. Dr Hawking believes within five years search in general will be far more contextualised and personal and heterogeneous search engines in enterprises far more sophisticated than they are today.

Certainly they are being constantly fine-tuned for specialist applications. Interwoven for example is this month releasing a search tool for professional services firms which allows them to search document management systems, records management, email archives and time and billing systems.

According to Professor Robert Dale, convenor of the ARC Network in Human Communication Science and Director of the Centre for Language Technology at Macquarie University, there are "three things that will be really

Earlier this year, Google launched a free service in the US that lets people use their mobile phone . . . to search for different services.

significant in the next five years: integration of linguistic information in the search process; voice search where instead of typing in (the search data) you speak it into a mobile phone — and there are all sorts of reasons why that's a challenge — and the personalisation of search where results are dependent on either

your geographical location or your previous search history."

Professor Dale acknowledges there are privacy issues, but he believes people would give up a little personal privacy for the potential benefits.

His group is exploring how to combine research into information retrieval with natural language processing advances to improve search results.

"Ask Jeeves was a first attempt in this area but the technology wasn't up to it. Now we see all the big players work toward where you can type in a question and get back an answer," Professor Dale says.

"Powerset has attracted a lot of attention as it has acquired the rights to some natural language technology that Xerox has developed over time. I'm a little bit sceptical, but Powerset is only the most visible of a series of companies working on this.

Their agenda is to change the nature of search. Of course Google may buy them, although it has its own people working in this area."

Powerset, based in Silicon Valley, plans to release its search engine and partner-program later this month. Harnessing natural language, Powerset's goal is to "break the confines of keyword search", making search more natural and intuitive.

Its premise is that a conventional search engine facing the queries "who bought Google?" and "who did Google buy?" would see the same sets of keywords and search on those, rather than true meaning of the phrases. According to Powerset, a semantic search would spot the difference.

While Powerset may be attracting attention, Sydney based Lexxe is already offering natural language search. Results even from the early version of

the engine seem pretty impressive.

A new version of the search engine is slated to be available in October and will have its own index and expand its use of natural language to deliver what founder Dr Hong Liang Qiao describes as semantic search.

Google's director of research, Peter Norvig, says the continued competition is healthy for the entire sector. For him the holy grail of search is that the engine performs "like a librarian or research assistant who knows everything. We're not there yet".

To get there, technologies were needed to synthesise multiple search results and generate a single comprehensive response to a query and also to steer people through the search process.

Yahoo! is also working on this and is testing an assistance layer which helps the searcher formulate the query.



Meanwhile, Dr Norvig says the search platform itself is changing, with people wanting to search from mobile phones or hand-held devices as well as PCs.

This would require search engines to have voice interfaces, and also to be able to package up results so that they could be delivered as an audio file, or massaged into a format that would fit sensibly onto a small screen.

Earlier this year, Google launched a free service in the US that lets people use their mobile phone or hand-held device to search for different services. Not surprisingly it's been controversial among providers of paid directory services.

But for Google it's an invaluable voice field trial, and the opportunity to capture a huge database of voice prints that it can use to tune future voice interfaces.

While the voice-driven search is limited at present, Dr Norvig believes free-form voice-driven search could be available within five years.

He also believes that search in the future will cover a much

bigger content spectrum likely to include a combination of free and subscription content, images, video and books.

Dr Norvig admits that in some ways, search today was "pretty impoverished in that it uses on average three words to search 100 billion pages".

HE believes that natural language processing was important in refining that and making sense of a query. He adds that it is also important to understand "the fit of the document in the world" which would differentiate a news report from *The New York Times* from that on *The Onion*. At the same time, search engines would have to distinguish between scholarly requests for information and pre-schoolers' searches.

"The role of search is to give to access to the right infor-

mation," Dr Norvig says.

Like Google, Yahoo! is exploring how to present the most useful data — be that text, web page, image or video — in the most useful form. According to Tim Mayer, vice-president of search product management, it's a significant engineering challenge requiring search across multiple domains.

It also requires a level of search engine intelligence, so that when people searched for "sunsets", the search engine recognised that most people might like to see images of sunsets, rather than say, a database of the times at which the sun sets.

Understanding context and also finding the most useful form of information is the key, he says.

The fact that the big search engines are pouring money into this should, he believed, arrest any potential fragmentation of the search market in spite of those 1700 plus search engines.

It's not possible at present to search YouTube or MySpace from Google or Yahoo!, but universal search tools which can search a wide range of different content (web pages, text, images and video) coupled with the ability of search engines to drive traffic to sites probably pointed to search consolidation rather than fragmentation, Dr Ross Wilkinson, research director of the ICT centre at the CSIRO, says access to the right information rather than just a range of information is important so that people can "make evidence-based decisions rather than opinion-based decisions".

He believes the big search breakthroughs will come from personalised search, where individuals can apply filters to fine-tune searching.

► Continued **PAGE 10**

LINKS

- www.powerset.com
- www.lexxe.com
- funnelback.com
- www.hcsnet.edu.au/nextgensearch07/abstracts

Giant leap for search unknown

◀ From **PAGE 9**

Allaying privacy concerns, Dr Wilkinson says personal filter data could be held on the user's rather than the search engine computer. He adds that the Federal Government is interested in delivering citizen centric services and that the CSIRO is working with AGIMO on how to maintain a person's profile so it could be used to direct search.

Personal filters would also influence the ability of a search engine to "push" information to users before they even knew they wanted it.

Tim Mayer believes that he "should be able to walk down a street in Palo Alto and one of my friends linked to my search engine should tell me about something. It should be more pushing information out than user pulling it".

He believes we are at a turning point — "over the last seven or eight years search result pages have looked fairly similar — things are starting to change as people are trying to understand user intent.

"So for example when you type in a movie, it will understand whether you want a preview, the showtimes, to book a ticket or look at reviews.

"We've started to launch a movie module where we respond to a movie inquiry — say about the new movie *Ratatouille* and rather than providing lots of different web page listings, we have all those things available on the one search page."

Search is about to become more interesting.



NEXT

Thanks to Telstra's advertising campaigns, we all understand the value of search as it pertains to the Great Wall of China. But what about rainbows?

To gauge how different search engines work, try the classic primary schooler's question, "Why is the rainbow curved?" in Lexxe, Yahoo!, Google, Baidu, LookSmart and Live.com.

One question —



many answers. For my 10-year-old, Lexxe's early version natural language search was far and away the most

useful and contextually appropriate, followed by Baidu, Google and Yahoo!.

The top response on the Live.com search was a strange non sequitur about neuropsychology, and LookSmart pointed to an article explaining why the Canary Islands are the "hottest queer sun and fun destination", which is rather more information than he currently requires.

BEVERLEY HEAD



Dr David Hawking (left) and project assistant Paul Thomas toss the frisbee — and ideas — at the ANU in Canberra in their hunt for the search engine to end them all.

PICTURE: ANDREW TAYLOR