

search insider

David Green, this month's Search Insider, examines the latest generation of 'intelligent' search engines, and finds himself wondering just how much their autonomy needs regulating.

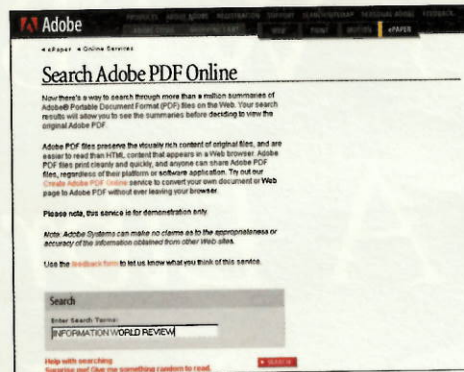
As the internet search genus continues to evolve, it has been speciating into new forms; meta-searching, popularity-based searching, link-analysis searching, natural language searching, search utilities and so on. This diversity and continuing specialisation is an indicator of what we could expect in the near future.

The shape of things to come

Search engines will continue to move away from interpreting statements to identifying the concepts that underlie the user's requests – such as natural language searching points towards the day when voice-based searching will become commonplace.

In the meantime, classification schemes will be increasingly used to refine users' keyword searches, suggesting alternative words or phrases to the user when presented with ambiguous terms. Specialisation, either through technological innovation or by type of content indexed, will continue apace. For example, most search engines cannot index PDF files, so Adobe stepped into the breach to develop a search engine that does (searchpdf.adobe.com)

No longer a simple issue of which search engine is the best or biggest, effective web searching now depends on choosing the appropriate content sources and search tool according to the requirements of each query – not unlike 'traditional' online business and scientific searching. For example, search utilities enable automation of many aspects of gathering market and business intelligence; running saved search



statements on a regular, allocated basis, providing alerts of changes to particular sites that you are monitoring and so on.

Insider searcher

Better algorithms and agents are delivering many of the current and forecast improvements in web searching and information delivery. While I covered intelligent agents in-depth in May's IWR (issue 158), search agents can learn from our behaviours and independently suggest new information sources to us, delivering truly 'proactive personalisation'. They deliver much of the functionality behind natural language searching and can determine not just the concept behind a sentence/phrase, but also identify other related concepts and links between those. This, combined with their ability to interrogate far more data than a single human mind could,

potentially opens the door to an explosive growth in human learning. New relationships between different concepts can be unveiled through the examination of mind-boggling amounts of information and the correlation of any links within.

Double Agents

However this Utopian *weltanschauung* is not without complication. There is a growing need for a public debate on the autonomy versus the accountability of intelligent agents. Just how far should their autonomy extend? In a near-future web crawling with intelligent agents, many of which will 'chat' to one another and share information, what will be the boundaries to regulate how much information about our behaviours they digress to other systems, databases and agents without our knowledge?



in association with



Information World review

Keeping you informed with the latest news, reviews and opinions in the information industry.

Get your own copy of *Information World Review* every month ...
... and don't miss the next issue:

- find up-to-date news and reviews on what's happening in your industry
- keep up-to-date with the latest internet developments, trends and technologies
- gain insight into the current industry issues with opinions from top information professionals
- find the latest jobs in the industry
- read our special feature focusing on company online information & cyber abuse

Visit our web site:

www.iwr.co.uk

or complete and return this form

Subscribe Today!

Yes! I would like to subscribe to *Information World Review*

£41 UK £51 Elsewhere

Special student rates available: contact rhart@learned.co.uk

☐ I enclose a cheque made payable to Learned Information Europe Ltd for £

☐ Please charge my credit card:

☐ Visa ☐ Access/Mastercard ☐ American Express

name on card

card number

expiry date

signature

☐ Please invoice me order number:

EC customers please quote your VAT number:

title first name surname

job title department

organisation

address

post code

country

tel

fax

e-mail

☐ From time to time we may allow other organisations to send you details of related products and services. Tick here if you do **NOT** wish to receive such information.

Simply complete and fax this form to: +44 (0)1795 414555
or post to:

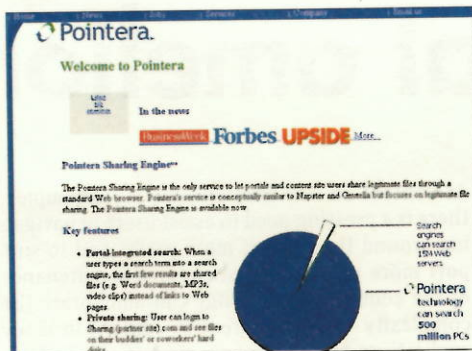
Information World Review Subscriptions
Learned Information Europe Ltd
PO Box 359, Sittingbourne, Kent ME9 8DL
T +44 (0)1795 414964 F +44 (0)1795 414555
E-mail iwr@galleon.co.uk

IWR 12/00

Agentware developers would be wrong to dismiss such privacy concerns, preferring instead to will the ascendancy of agents on the web through technological prowess alone.

Besides, on this latter point, agents have an Achilles' heel; like all computer-based systems, they are critically dependent on clearly defined parameters of operation. Once removed from this defined context their effectiveness collapses. Such neatly defined context can't mimic real-world complexity and agents can never, therefore, truly attain the personal levels of discrimination that we exercise for ourselves.

After all, humans can be wonderfully whimsical and circumstances are never truly predictable. Mike Lynch of, ahem, Autonomy (www.autonomy.com), was wrong when he flatly



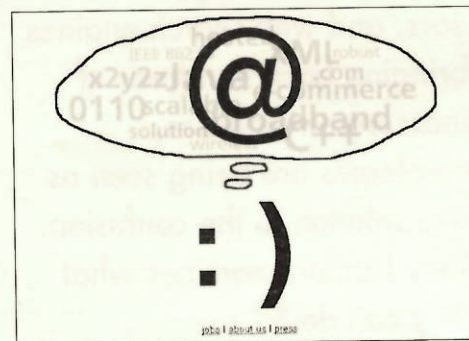
information they require on other user's computers, which provides numerous advantages:

- Extends the range of searching beyond web-servers to the millions of individual computers connected to the internet at any one time.
- The search is run in a 'real-time environment' so only information that actually exists on other computers will be listed in the results - so no dead links.

■ A greater range of file types can be searched - including dynamic data in databases.

■ Suitable for locating volatile or temporary information.

The downside is that, as the search is running in a 'real-time environment' across many computers, rather than a single search engine database, users will wait much longer for results to appear in their browser - often over a minute. This

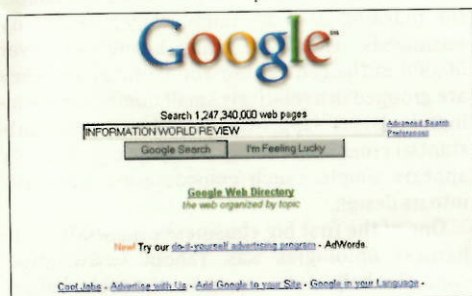


timescale is unlikely to reduce until broadband access becomes much more widely available. The programmes that enable such peer-to-peer searching are a new category of search utilities that are referred to as distributed search applications. Examples include Pointera (www.pointera.com) and InfraSearch (www.gonesilent.com).

Whilst big brand names such as Yahoo! and Amazon will continue to pull in new users on the basis of the trust their brand name brings, web searching is still a wide-open field. Relative newcomers such as Google, then FAST and later WebTop have all demonstrated that it is possible to rise to prominence very quickly.

I have outlined some of the possible future developments in this article - the rest may yet prove to be anyone's guess...

David Green is a council member of the Institute of Information Scientists. He writes here in a personal capacity. His website is www.clickmedia.freemove.co.uk

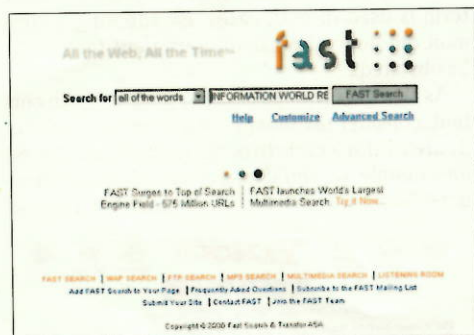


commented "Search is dead". Sergey Brin of Google (www.google.com) is perhaps more astute when he commented that "people will search more and more, because the diversity of content requires it".

Remember - the search genus is evolving, providing a variety of options to the searcher. Privacy is already an issue and in an online world where much crime will migrate from the theft of artefacts to theft of identity details, it will become vitally important. Agentware developers would do well to incorporate robust, user-determined levels of personal security settings that dictate what information the agents we use can digress, and to whom. Naturally we would want different security/privacy levels according to which sources we were using. Otherwise concerned users will simply opt for other search tools - such as the newly emerging distributed search applications.

Peering through a window to the future

Just as the nascent explosion of the web was enabled by open internet protocol standards, its ability to self organise with agents will be dependent on the presence of open standards and com-



mon protocols. While concerted corporate alliances such as Symbian seek to determine (and thus control) such future open standards, it is likely that the open source movement that brought us Linux, Gnutella and Napster will have a greater impact. The internet was originally peer-to-peer but it became 'centralised' when it became commercialised. However, peer-to-peer computing, information distribution and indeed infrastructure could be the future.

Almost all of today's search engines are based on a centralised client-server model in which many users interrogate a single mainframe server (or proxy copies). Even though FAST (www.alltheweb.com) uses parallel processing, it nonetheless falls under the 'centralised' category, as there is no decentralised user-to-user interaction. There are several major drawbacks to this prevailing approach:

- Spiders cannot index databases that are integrated to websites.
- They index varying, and often small, percentages of web.
- Search engine indexes often contain out-of-date pointers to content, i.e. dead links.

Also referred to as distributed file searching, peer-to-peer searching allows users to find the

Paper 1.0 - Information Management

Sheet ref. wm/1

Time - 2 minutes

Only write in the white areas

Answer all questions by circling the correct answer in the white panels on the left of the sheet.

Name:

Student No:

Question 1

The most popular feature of search technologies today is:

- A Their ability to return hundreds of thousands of results
- B Their ability to hide the relevant results amongst irrelevant ones
- C Their ability to misunderstand the users' intended meaning
- D None of the above

Question 2

What is "a controlled vocabulary of terms or phrases"?

- A A Tax Return
- B A Taximeter
- C A Taxidermist
- D A Taxonomy

Question 3

"Two great nations divided by a common language", Churchill would have benefited from which technology to interact more successfully across the Atlantic?

- A Wordmap® Taxonomy Technology
- B Spellchecker
- C GPS
- D Telepathy

Question 4

By providing building and maintenance services for taxonomies, which company is to search and navigation as the blood stream is to the body?

- A Blood Transfusion Service
- B Bob the Builder
- C Wordmap Ltd
- D None of the above

Hand your completed papers in at
Stand 170 at Online Information 2000
or e-mail to answers@wordmap.com

FOR ALL YOUR TAXONOMY NEEDS

wordmap®

For official use only

www.wordmap.com