search insider

The internet delivers content and connectivity. Both can be exploited for information research and collaborative knowledge sharing. The world's largest repository of information, however, grows at an estimated one million web pages *every day*. With so much available you must be focused. David Green offers a few pointers for those new to the searching game.

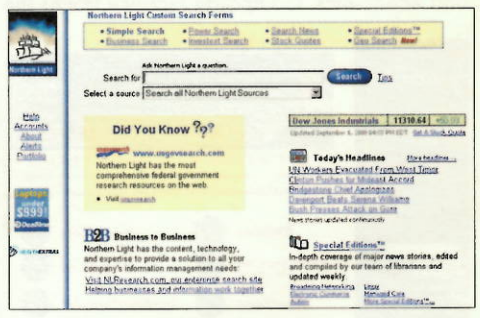
Stop searching. Start finding

There are a variety of tools available to the professional researcher. These range from human categorised web directories, through search engines, to search utilities that incorporate intelligent agents/bots.

Subject-specific directories are extremely good starting points because experts within that field usually compile them. Useful directories for news, company research and knowledge management are considered separately a bit later.

When using search engines it can sometimes feel as though your results have been generated by a magic randomiser, rather than a huge database that operates powerful retrieval and relevance-ranking algorithms.

Two stars do shine, though: Northern Light (www.northernlight.com) offers useful research features, categorisation of results and integration with pay-for business information sources. Google's (www.google. com) powerful link-analysis technology helps identify authoritative sources.



However, for targeted information research I would recommend using a search utility. Search utilities are programs that offer additional functionality such as automatic refinement of results, document summaries, active hyperlinking of query words, saved searches and language-independent searching.

They incorporate a range of intelligent agents and can therefore adapt to your changing preferences, delivering ever more relevant results. They are extremely useful for monitoring topics on a continual basis, providing alert reports according to the user's original customisation.

Some of the most popular utilities include LexiBot (formerly Mata Hari) (www.lexibot.com), BullsEye Pro (www.intelliseek.com) and Copernic (www. copernic.com).

Keep informed of the latest developments in the world of search engines and intelligent search agents at SearchEngineWatch (www.searchenginewatch.com), Search Engine Showdown (www.notess.com/search) and Botspot (www.botspot.com).

Reach the headlines

Before the widespread uptake of the web during the late 1990s, the dominant source of electronic news was pay-for online business information databases:

- **70s to mid-80s**: mainly premium business newspapers and some quality nationals available
- **mid-80s to early 90s**: plethora of business and professional publications appear online
- **mid- to late 90s**: most newspapers started to publish directly onto the web in order to reach a much wider audience than could possibly be offered by online hosts. Many news sources such as regionals and newspapers from emerging markets also became available electronically for the first time.

The vast majority of news web sites are free and many offer searchable archives. The changed economics and dynamics of news publishing has resulted in an explosion of news content, which in

turn has generated a new market to sort out this morass. Currently there are three main solutions:

- **News indexes**: directories of newspapers worldwide, usually sorted by country; for example, The Paperboy (www.thepaperboy.com), Yahoo!'s listing of newspapers by country (dir.yahoo.com/News_and_Media/Newspapers/By_Region/Countries) and Newscentral (www.all-links.com/newscentral). For magazines visit www.enews.com.

- **News aggregators**: hundreds of aggregated sources, with stories sorted by industry, category and so on. Try www.newsbot.com (part of the Lycos network of sites - archive only extends for one month, limited number of quality sites indexed) or www.newsindex.com (only aggregates current stories from various news sites, no archive).
- **News alerts**: set up profiles on your topic of interest - competitor news, key markets. Look at www.newsedge.com, www.newshound.com, www.newsnow.co.uk, www.newspage.com.

Several of the search portals offer free personalised news alert services as part of their portfolio of 'sticky' features. Of these, My Yahoo! is most directly relevant to UK users. The Financial Times (www.ft.com) has an extensive archive and offers email alerts.

Each of these three types of news sources has a role to play: news alerts help automate the process of monitoring particular markets or competitors, aggregated news sources are useful for more in-depth historic research, while news indexes can help you quickly identify relevant news sources in unfamiliar (geographical) markets.

Company information

Information on company financials and similar data is generally published at database-driven sites. Search engines have great difficulty indexing the contents of back-end databases - the very content you need - so it is necessary to visit individual web sites and interrogate their indexes. There are plenty of web sites to choose from - here are a few:

- **Corporate Information (www.corporateinformation.com)**: over 15,000 in-depth research reports and 20,000 profiles on publicly quoted companies from around the world. Famous for its extensive links to web sites that offer useful information for corporate research. A must visit.
- **Business Information sources on the Internet (www.dis.strath.ac.uk/business)**: maintained by Sheila Webber formerly at the University of Strathclyde, this site offers links to a variety of business information web sites, with comments and evaluation by this industry expert.
- **Hoover's UK (www.hoovers.co.uk)**: basic profiles of over 1400 UK and European companies, including contact details, sales, description of activities etc. Database can be searched by company, location or industry sector. Also contains a directory of relevant sites for company and industry research in the UK and mainland Europe.
- **CAROL (www.carol.co.uk)**: links the annual reports section of UK quoted companies' web sites. Now also includes sections for Asia, Europe and the US. Offers real-time alert service for merger, acquisition and take-over news.

Generals of knowledge management

- **Researcher's Interests (km.brint.com)**: has been praised as the best KM site on the web - and with good reason. This virtual KM library offers an extensive collection of articles, resources, analyses, calendar of events, news and much more (also check out www.brint.com/interest.html, a comprehensive listing of business-related web sites).
- **The Open Directory Project KM guide (www.dmoz.com/Reference/Knowledge_Management)**: compiled by volunteer KM experts from around the world, this excellent directory lists almost 600 sites. Categories include knowledge creation, discovery, flow and retrieval, KM glossary and processes.
- **SearchTools KM guide (www.searchtools.com/related/knowledge-mgmt.html)**: not as extensive but lots of useful articles and links to KM tools. Focused on the technology aspects of KM.

Can you digit?

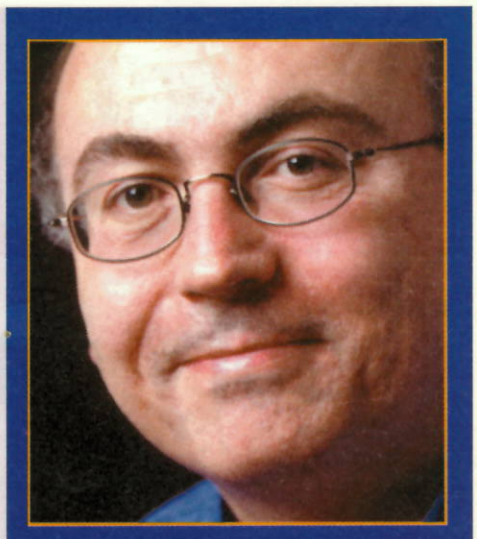
Keeping up to date with all things digital need not be a hamster-run round the wheel of technology - always feeling that bit behind no matter how much effort you apply. Bluetooth and third-generation mobile phones will be among the (seriously) next big things in the information industry. Want to know what they are? Visit What is? (www.whatis.com).

But don't struggle alone - the connectivity provided by the internet enables people to congregate in a virtual place to exchange information, opinions and advice. The support becomes synergistic.

Most of you will be familiar with Free Pint (www.freepint.co.uk), a fortnightly email newsletter written by UK information professionals. Identify other mailing lists and newsgroups at Reference (www.reference.com), DejaNews (www.dejanews.com) and Liszt (www.liszt.com).

Sites such as the Open Directory and Free Pint demonstrate that to keep up to date with the internet's content, it is necessary to plug into its connectivity. Such sites and mailing lists are only as useful as the contribution made to them by their subscribers - so you get so much more than you give. What a wonderful web!

David Green is a council member of the Institute of Information Scientists. He writes here in a personal capacity.



Flood on ... KM

You know, of course, that I thoroughly disapprove of the whole concept of 'Knowledge Management'.

That's mainly because of a nagging sense of guilt and doubt stemming from my period of work in the alleged 'real world' - that is, prior to being flung into a man o'war publishing company and ending up as one of Her Majesty's Press.

You see, before that I worked as an analyst/programmer for a large UK software house, in the AI department.

Remember AI? We're talking mid- to late 1980s here, when AI - artificial intelligence - was *the* trendy area of computing to be in. Magnificent verbiage was the order of the day: I helped build an assumption-based truth maintenance system for an AI toolkit; we talked of implicit and explicit knowledge, hard-wired and software mental skills. And on and so gloriously on.

Somewhere in there, by the way, we also had Knowledge Engineering: the self-proclaimed practice of extracting, codifying and maintaining domain expertise, typically in the form of *if–then* (production) rules, in (you guessed it) a knowledge base.

A Knowledge Engineer was a kind of deft psychiatrist/industrial psychologist/wizard computer programmer, able to sit down with any kind of expert - chemical engineer, proctologist, vintner, football manager - extract the heuristics (rules of thumb) that guided their thoughts, write it up in Lisp (a lovely computer language now as dead as Sumerian) and, hey presto, you had a genius-in-a-box (aka an expert system). Sounds tops.

It didn't work.

A bit like Trotskyism or medieval Catholicism, it was a beautiful theory brutally punctured by reality. You can't suck expertise out of experts like that. They either resist the enterprise altogether, or cannot articulate the fine-grained details of their thought processes in neat little capsules of logic that can be expressed in a line of code.

Oh, yes, there were expert systems that worked well - magnificently, in fact. But like great works of art, they were more the superlative expression of an individual's (or dedicated team's) effort and craft, not assembly line, repeatable and stampable products.

Expert systems saved firms money, found new mathematical formulae and diagnosed glaucoma, but these were merely marvellous individual computer programs, not a new template for working with computers.

In the same fashion, alas, I hear much the same kind of promises and wondrous charms in KM. I chaired a Knowledge Management conference two years ago. An intranet, an online search engine, a portal, an internal library *cum* document management system, and an internal London Transport retraining project were all offered up as examples of KM in anger.

Sorry - no. These were all excellent individual projects that gave their designers and clients value. But like AI or expert systems or

"A bit like Trotskyism or medieval Catholicism, it was a beautiful theory brutally punctured by reality."

Knowledge Engineering, what was not - and, in my view, cannot - be on offer, shrink-wrapped, on a CD-ROM, is a coherent, repeatable mechanism or method for capturing, controlling and exploiting implicit corporate knowledge.

Am I wrong? Is my comparison with AI ten years ago too harsh or specious? Why don't you tell me?

Or do you *like* sciences that aren't really scientific?

Gary Flood is deputy editor of one of *IWR*'s sister titles for IT professionals, *Computing*. To talk back, email: editor@iwr.co.uk.

TALKBACK

I am writing in response to Gary Flood's column 'The web's dark matter' last month (*IWR*, 161).

The fact that huge areas of the internet are untouched by search engines is quite well known - but the point that nobody seems to mention is that much of this material is better left un-indexed. American college students' home pages and sites about people's pet cats, to give just a couple of examples.

And if the internet really is 7500 terabytes, I'm pretty sure that the 'unknown' fraction is generally not worth knowing - people with important information to offer generally make sure that it *is* indexed by the search engines: I'm sure that the *IWR* site would have appeared pretty close to the top of Gary's list of 576,106 hits, if he conducted his search properly.

Incidentally, the first of the 'alternative' search engines that Gary mentions, Google, works by weighting sites according to their popularity (i.e. the number of hits that they record), so using Google would actually reinforce the 'black hole' effect, by producing sites that lots of other people have already visited, and ignoring all the sites that have had only a few 'hits'.

Jane Kerr, Associate Lecturer, OU Technology Department