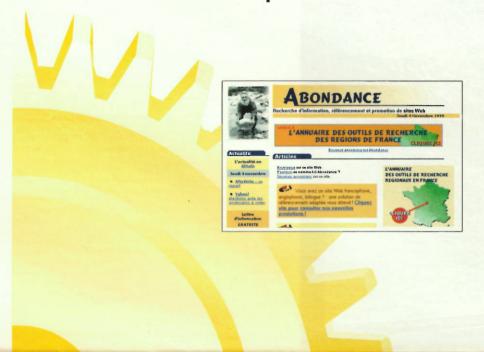# search insider

Wouldn't it be nice — no, damn useful — if you were able to use the same search commands across the major search engines? Well soon you could be able to do just that.

Launched by Danny Sullivan, creator of the highly influential SearchEngineWatch web site and monthly newsletter, the Search Engine Standards Project seeks to foster common standards in syntax and functionality across the major search providers. Other future developments also point towards some degree of search standardisation.

Despairing disparity

Sullivan describes how the initiative was born: "A growing user frustration, particularly among professional researchers, was that they have to use different commands to conduct the same searches on different engines. For example, if users want to restrict a search to a particular site on AltaVista they have to use the 'host:' command. Using Go the command is 'site:' while on Inktomi-powered search sites, it is 'domain:'. The main reason for these different syntaxes is that the search engines just don't talk to each other any more because things have become so competitive."

Sullivan believes that the major search engines could, and should, be talking on non-competitive areas. Partially as a consequence of the effects of portalisation, search syntax is no longer (if it ever was) a unique selling point. Rather, search sites now differentiate themselves by the various 'sticky' features that they offer to encourage users to stay on their sites for longer, such as news alerts, free emails and web hosting.

Talking heads

It is for these reasons that Sullivan created the Search Engine Standards Project. Members include representatives from the major search providers and several prominent academic and industry observers (including Jakob Nielsen). True to the democratic spirit that the initiative seeks to encourage, the project also includes members who produce overlapping web sites that cover search engine developments. Chief among these are Greg Notess and Avi Rappoport, the respective authors of the Search Engine Showdown and SearchTools.com web sites. While membership of the group is by invitation only, all membership suggestions are open to peer review on the project's email discussion list. This is to ensure that potential members are suitably qualified and that the project remains focused. There is, however, a public feedback facility.
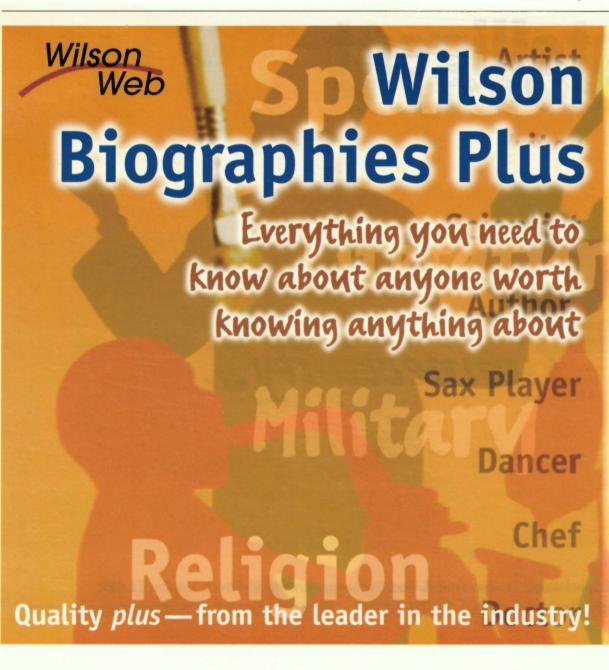
Old school rules, new school uniforms

Two proposals that the project team members are currently working on are:
- establishing a common availability of, and uniform syntax for, site-specific searches; and
- for all search engines to provide the ability to locate an exact URL within their indexes using a uniform URL: command.

There's nothing new here. Many of the major search engines have provided commands for restricting searches to specific web sites for years, but, as illustrated above, the commands vary wildly. Fortunately, for URL searching most of the search engines use the same command – 'URL:' – so a uniform standard could be established here relatively quickly.

The first proposal would be of great benefit to professional searchers who know what they are looking for and from which publisher. The second proposal would be of great benefit to web masters in determining exactly what has been listed from their sites. Anyone responsible for monitoring web sites of competitor organisations would also find this functionality extremely useful.

But are such proposals really useful, or are they merely 'nice to haves'? They are definitely needed. As covered in my last Search Insider feature (*IWR* 151), although the size of search engine indexes continues to grow, the percentage of the web each covers continues to decrease. Professional researchers are interrogating several indexes to ensure comprehensive coverage. Obviously it is an impediment to using a search engine if every time you use a different one you have to learn a new search syntax. Actually, this disparity in search engine syntax is one of the reasons for the burgeoning popularity

of search utilities: the user enters a search command and the utility 'translates' the command according to the syntax of each search engine index it interrogates (see *IWR* 147).
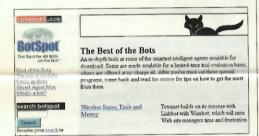
Getting better, if not bigger

In the meantime, search engines have been working hard to improve the relevancy of their search results. According to Sullivan: "There's been a lot of work to improve searching that doesn't get noticed. At the beginning of the year AltaVista introduced a 'related searches' feature and since then at least three other services —InfoSeek (Go), Direct Hit and AOL Search — have introduced this kind of feature. Search engines are becoming much smarter. Now they're examining the kinds of web pages that users view, they're refining their ranking algorithms and making use of links analysis — all of which make it harder for a web master to manipulate ranks. This, in turn, improves relevancy of results for search engine users." A virtuous virtual circle.

Sticky fingers?

So who is going to pay for these improvements and how will search engines make their profits? Project member Jakob Nielsen predicts that: "In the future we will have micro-payments for search. Realistically, to provide quality information over the long term requires serious effort. Companies have to be compensated for providing that."* Sullivan disagrees with micro-payments. He believes that, in order to realise the ecommerce opportunities that a mass market audience affords, search sites will continue to invest in providing improved free search services – at least for now.

Here come the X files

While acknowledging its data transfer benefits, Sullivan is sceptical about the ability of XML to revolutionise web searching. He believes that some searchers are "hoping in vain". Why? Because "searchers need a trusted source of

meta-data rather than user-provided meta-data."

True – but I disagree that users are "hoping in vain". There are widespread voluntary efforts to develop standardised XML tag sets for specific industries and interest groups. To fully realise the benefits of XML, each industry will need to agree standards for the tags used to describe the information that is specific to its discipline. Mathematicians, genealogists and chemists have already agreed standards. So, in other words, future XML meta-data will, to a certain extent, have originated from a trusted source – the peer-reviewed groups that established the standard XML tags for their own industries and topics.

Currently, text-based search engines will return a list of documents that will contain some information relating to the user's request. XML-enabled query searching, like any other query language, will return the relevant data that has been extracted from a document, rather than the entire document. Such query-based searching can also be used to perform computational analysis and manipulation of presentation on retrieved data items. Search engines will need to learn the standard tag structure that has been

agreed by each industry or interest group. They will also need to change their search interfaces to offer users the choice between text searching and field or tag searching. Currently there are no proposals by the standardisation project members to examine such issues.

www.online-information.co.uk

The future flux

"Search engines are here to stay": regardless of advances such as XML, Sullivan predicts there will be a continuing, and growing, need for on-demand information, and that web searching based on location and frequency of terms will continue to predominate. Search engines are also making much greater use of humans for categorisation of information on the web. The past year has witnessed a major shift in this direction, with four major services adopting human searching and categorisation rather than computer

web crawling – Lycos, MSN Search, AOL Search and Netscape.

Recognising the need for change, members of the standardisation project have suggested some further project proposals, including:

▶ Updating the robots.txt command to take into account advances in web publishing since the command was first developed.

▶ A new syntax for indexing. In this respect, the project could possibly become something like the W3C (the body responsible for developing technical standards for the web).

▶ A shared blacklist of web sites that spam in order to bolster their rankings. If a web site were caught spamming by one search provider then the other search engines would be alerted so that action could be taken.

The Search Engine Standards Project is a commendable voluntary initiative that could help to deliver real standardisation benefits, primarily for users, but also for search engines themselves. A real win–win idea that deserves everyone's support.

*Janah, Monua. 'Web directories profit motive complicates searches by consumers', *San Jose Mercury News*, 16 August 1999.