My favourite software


What it does
Google is far and away the fastest, most accurate and most reliable search engine on the Web. All one has to do is type the search terms into the box and Google will return a list of ranked results almost immediately.

How it does it
Google sends its Googlebot software off continuously to ‘crawl’ around the Web looking at all the Web sites it can find – currently about 2 billion pages. It then uses PageRank, a proprietary ranking system, to score the indexed pages using a number of factors: not only the number of pages that link to them, but also the ranking of those linked pages, as well as other criteria involving the text within the page. This complex citation ranking allows an automatic and objective ranking of pages, and is the reason for Google’s uncanny accuracy.

Benefits
Since its inception Google has used a laudably clear and intuitive interface, which it has managed to maintain even with the addition of a number of additional services. It is extremely fast, highly accurate, and utterly reliable. Many search engines accept payment from Web sites in return for higher rankings but Google refuses to allow this, ensuring that its results are scored on objective criteria alone.

Google stores copies of searched pages in a cache, so one can read them even if the original site is down, and it can translate Web pages into any one of a number of languages, including Bork, bork, bork – the language of the Swedish chef from the Muppets.¹

The search facilities are sophisticated but easy to use. For normal searches common words such as ‘where’ and ‘how’ are ignored, but they can be included by adding a ‘+’ in front of them. Adding a ‘-’ in front of a word allows one to exclude terms, thus increasing sensitivity. One can search within domains (i.e. www.fhm.org.uk) to narrow things down even more, and also restrict by language, date, or where on the page the search term is found; Google even automatically checks the spelling of your search terms and offers alternatives if appropriate.

For ease of use, and added features, the Google toolbar is well worth installing; simply download it over the Web to provide a permanent Google search box at the top of one’s browser.

Drawbacks
Google only crawls the ‘visible Web’ – there is also an ‘invisible Web’ that is inaccessible to conventional search engines.² Crawlers are really only any good at finding simple text-based pages, and complex database-driven sites are out of reach using currently available technology. Google is the only major search engine to index PDF files (such as many Department of Health documents) thus rendering them ‘visible’ in a way that its competitors do not, but even Google only picks up somewhere between about 25 and 50 per cent of the machine-readable Web, which is many hundreds of times smaller than the invisible Web.

Every search involves a trade-off between sensitivity and specificity. Most of the time PageRank works so well that a useful link will appear on the first page or two of results, but some searches are so non-specific that they inevitably generate large numbers of unwanted results; at such times it can be hard to separate the wheat from the chaff with Google.

What’s the competition
Directories, such as Yahoo (www.yahoo.com) rely on human interaction to code and assess Web sites. This not only makes them less responsive, it also opens them up to subjectivity and bias; nevertheless, this can be useful for browsing within a particular topic area. As a response to the information overload resulting from many searches, some engines, such as Vivisimo (www.vivisimo.com), categorize their results to make them easier to search within. One approach to resolving the drawbacks of individual approaches is to use a meta-search engine such as ixquick (ixquick.com) – this searches using several different search engines, and has the benefit of sensitivity at the cost of decreased specificity. All these approaches have merit in different situations, and I have tried them many times, but I always return to Google as the best available compromise.

Declaration of interest – none

References

Dr. Harry Rutter
Specialist Registrar in Public Health Medicine
Oxfordshire Health Authority,
Richards Building, Old Road,
Headington, Oxford OX3 7LG
E-mail: harry@oxon.org

© Faculty of Public Health Medicine 2002