Quality in Primary Care 2005;13:171–5

Knowledgeshare

Web alert: advanced techniques for simple searches

Ben Skinner BSc (Hons) MA
Evidence-based/Knowledge Management Librarian, *KnowledgeShare*, The Library, Sussex Postgraduate Medical Centre, Brighton General Hospital, UK

Google: pros and cons

During a recent teaching session we asked a group of first year medical students: 'which resources do you use to find information for your work?'. The majority of groups placed the search engine Google near the top of their list, which came as a surprise to no one, although the vehemence of one respondent was particularly memorable. 'Google rules!' he wrote; 'The world at your fingertips!'. His list of drawbacks to using the search engine simply read: 'None'.

When approached uncritically, and without an understanding of their limitations, the use of internet search engines to source healthcare information is about as far from evidence-based practice as one can get. A search in the Cochrane Library for systematic reviews (www.thecochranelibrary.com) or the National electronic Library for Health for national guidelines (www.nelh.nhs.uk) will (if successful) provide you with the best available evidence. The NHS National Core Content provides free access via an Athens password to databases of research publications including the British Nursing Index, Medline, and Embase, and of course Medline can be accessed without the need for a password at PubMed (www.ncbi.nlm.nih.gov/entrez).

Nevertheless, the first port-of-call for many healthcare professionals continues to be Google (www.google.co.uk), or possibly Yahoo! (http://uk.yahoo.com), its biggest competitor in the online searching industry. Although the evidence-based resources and journal article databases listed above provide high-quality results and are sophisticated enough to allow precise information retrieval, this strength is also their greatest weakness. The very features that attract experienced searchers by allowing them to construct highly sensitive and specific searches are the ones that confound novice users and push them away.[1] In contrast, search engines are astonishingly simple to use, they almost always produce results, and they produce results that are, on the whole, easy to digest. This month's article will look at how to use search engines more effectively and the new developments that are making Google even more useful.

Search engines and misguided loyalty

Web search engines are programmes that trawl the internet for websites, following links from page to page and building up massive searchable collections of the resources that can be found online. This is done entirely by computers, and there is no quality control on what is found. When keywords are entered into a search engine a list of the web pages that include these words, often numbering in the millions, is presented to the user.

In the early days of the internet, competition among the different search engine providers was fierce. Every year brought new contenders, and reinventions of existing services, all hoping to become the dominant product on the market. Clear winners are starting to emerge, and most of us now have a 'search engine of choice' that we use to the exclusion of all others, but there is still a variety to choose from, each with their own advantages and disadvantages and methods of searching. Most importantly, each search engine holds a vastly different set of websites, and ranks those sites in different ways. This means that when we stick to one provider we run the risk of missing important resources, especially considering that most of us do not look past the first two pages of results.[2]

Thumbshots: http://ranking.thumbshots.com

A striking illustration of this problem can be found by visiting Thumbshots, a site which allows you to search for the same keywords in two different search engines and visually compare the results. Individual results from the two providers are represented as rows of dots, with lines connecting websites that have been retrieved by both searches. This allows the user to see the overall level of overlap, as well as the differences in ranking between the two engines. A search for *shipman inquiry* compared across Google and Yahoo! reveals only a 19% overlap in coverage. High-ranking results from Google do not appear in the first 100 results from Yahoo! and vice versa. This provides an eye-opening demonstration of why a careful searcher *must* use more than one search engine.

Search Engine Watch: www.searchenginewatch.com

If you are going to broaden the scope of your online searching in this way, you will need to know about which search engines are available and how to search them effectively. A good place to start for information on this topic is Search Engine Watch. This site contains introductions to the different search engines, with ratings and statistics to show how they compare in terms of popularity, size, coverage of the web, and advanced search features.

The yearly Search Engine Watch awards are highly regarded in the industry and can be found at http://searchenginewatch.com/awards. The fifth annual award winners have just been announced and Google has been beaten to the title of Most Outstanding Search Service (by Yahoo!) for the first time since the awards began.

Consider the alternatives

In addition to Google and Yahoo! some important names to be aware of are Dogpile and Scirus, both of which have won awards from Search Engine Watch.

Dogpile: www.dogpile.co.uk

Dogpile is an example of a metasearch engine. These are websites that send your search to a large number of search engines simultaneously and present you with all of the results at once. This can save considerable time, although there are some drawbacks. Only the top-ranked sites from each provider are reported, leading to considerably fewer results and a greater risk of missing something important. In addition, two searches on the same subject conducted at different times may produce wildly different sets of sites, so it is worth repeating your search a few times to make sure you have retrieved everything. Dogpile echoes the simplicity of design that has made search engines so popular. After searching, it allows you to refine your results by selecting a subcategory within your topic, and also shows a list of your recent searches.

Scirus: www.scirus.com

Scirus is another site that emulates the stripped-down, user-friendly design of Google. It is a search engine for scientific information only, and is thus ideally suited to healthcare professionals trying to avoid the 'white noise' that can be thrown up by other resources. The front-page allows the user to search within journal sources, within scientific websites, or both. The journal sources include a wide selection of freely available peer-reviewed full-text journals from collections such as BioMed Central and CogPrints, as well as journal abstracts from PubMed, Science Direct, and elsewhere.

Scirus clearly indicates the source of each search result. Just like some of the more advanced, subscription-based databases, the user is able to mark interesting results for saving or emailing at the end of a search. And like Google, Scirus has produced a 'toolbar' that can be downloaded to your desktop. As well as giving you direct access to the Scirus search facility, this toolbar blocks adverts and highlights your search terms as you browse through your results. Currently, Google is being forced to work hard to provide a search engine service that is as suited to healthcare professionals and academics as Scirus is (see references to Google Scholar below).

Quality control

Before going on to look at some tips for getting the best out of Google it is important to mention the quality issue. Because of the fact that anybody can publish to the web without their work being quality assessed, searchers must always be alert to the source of a piece of information. The Resource Discovery Network provides a virtual learning suite (www.vts.rdn.ac.uk) which covers the range of questions that should be asked about online sources, and provides pathways specifically for doctors, nurses or allied health professionals. Two of the main causes for concern are: commercial enterprises trying to promote their own products (either overtly or covertly); and pornography. The latter can be particularly relevant in

the healthcare context as many searches on subjects from *safe sex* to *breast care* may produce unwanted results. A recent study from the US showed that using filters to block pornographic content can inadvertently block large numbers of health-related sites.[3]

There are now a variety of accreditation processes running to ensure the quality of medical websites, in particular the European Union's eEurope: Health Online program.[4]

Getting the best from Google

'Google' is a variation on the word 'Googol' which means 10^{100} and represents the company's aim to provide access to the massive quantity of information available online. The search engine was created in 1997 and was one of the first to rank pages not just according to how often keywords appear, but also by how many other websites link to a page, in an attempt to factor in the site's popularity. Theirs is now the most used search engine in the world, and the phrase 'to Google' has entered the language as a verb. But while most of us use the search engine, we are not necessarily taking full advantage of its search facilities. 'While Google is transparently easy to use, it is not transparently easy to use *better*.'[5]

Be more specific

The most obvious problem with large search engines like Google is the overwhelming quantity of results that they produce, and the lack of relevance of many of those results. Luckily there are a number of techniques to increase the specificity of your search.

A search for *emergency admissions* will find all websites in the database that have the word 'emergency', and the word 'admissions' somewhere in the content, whether these words are next to each other or separated by many paragraphs. To ensure that the words are searched as one discreet phrase you should enclose them within speech marks. Searching for "*emergency admissions*" leads to a 95% reduction in the number of results. This is particularly important when dealing with so-called 'stop words'. These are short words (for example 'and', 'with', etc.) that appear so frequently that search engines ignore them. Just occasionally it is important that a stop word is included in your search, and you can use phrase searching to insist on this. If you search for *phase i trial* Google will ignore the 'i' completely; searching for "*phase I*" *trial* will produce better results.

It can also by very helpful to phrase your search as if you were answering your own question. Rather than searching for *arthritis treatment* you might type "*treatments for arthritis include*".

Another useful technique is to use the minus sign to remove terms from your search. This comes into play when there are associations related to your keywords that you are not interested in, for example the keyword 'depression' might bring up sites dealing with stock market crashes and failing economies. To make the results more specific you can search for:
depression – economic – financial, and all sites that include the words 'economic' or 'financial' will be removed from the list.

Field searching is another way to home in on the more useful results. When your keyword appears in the title of a page (known as the 'title field'), it usually means that the page is particularly relevant. You can restrict your search in Google to only search for pages that include 'thyroid disease' in the title by typing *intitle:"thyroid disease"*. There should be a colon after 'intitle' and no space before your search term. You can also restrict your search by language, file format (perhaps you only want to look for PowerPoint presentations), and date, by clicking on the Advanced Search link.

Develop your sensitivity

Although the most pressing problem with search results tends to be that you are presented with too many, it is important also to be aware that you might be getting too few. Before you begin any search you should spend a little time brainstorming all of the different terminology that might be used to describe the topic that you are looking for, so that you catch each of the important documents however the author has chosen to word them. To search for synonyms on Google, use brackets and the word 'OR' in upper-case, for example:
(hypertension OR "high blood pressure") elderly will search for everything with the word 'elderly' and with either 'hypertension' or 'high blood pressure'.

It is also possible to get Google to search for synonyms and variant spellings automatically by using the tilda (~) mark. ~"*complementary medicine*" will pick up sites that use the term 'alternative medicine' as well as those that use 'complementary medicine'.

Many of these search techniques are equally relevant in other search engines, like Yahoo! Always check the search tips and help pages to be clear about exactly what can be done by each resource.

Variations on a theme

There are a considerable number of additional resources available from Google, which go beyond the basic search facility.

Google Images: http://images.google.co.uk

One is Google Images, which allows the user to search specifically for pictures that appear on the web. At the time of writing there were over a billion images in the Google database, which makes this an extremely useful resource, particularly for healthcare professionals. If you need to confirm a diagnosis of dermatitis herpetiformis or some other uncommon skin disorder, simply plug it into Google Images.

Google Scholar: http://scholar.google.com

Until recently, articles from the Medline database were unlikely to be found via a search on Google, despite the fact that all of its journal references were included in the Google system. This is because of the Google ranking algorithm which places the most linked-to sites at the top of its list and relegates 'unpopular' sites to the bottom. Google Scholar, a version of Google that is specifically intended for academic use, has done away with this problem. Google Scholar only finds references to research publications, although access to the full text will not be available unless the publisher has made their content available electronically and the NHS library services have subscribed to it. It should be remembered of course that for very specific and sensitive searching you should always return to the advanced databases like PubMed that provide more flexibility.

One particularly useful facility that Google Scholar does provide, and most subscription-based databases still do not, is citation searching. When you find an article that looks relevant, Google will allow you to quickly pull up a list of all of the articles that cite it.

Google Print

A further development is Google's recently announced 10-year project to digitise millions of books and journals from the library collections at the universities of Oxford, Harvard, Stanford and Michigan and the New York Public Library. Users will not be able to view the full contents of material that is still under copyright, but they will be able to search the full text and see relevant excerpts and bibliographic details.[6]

Other Google services

Google News (http://news.google.co.uk) takes information from 4500 news sources, which you can search or browse by topic and locality. You can also set up personalised alerts that send you news according to your interests at a frequency to suit you. Google Groups (http://groups.google.co.uk) allows users to contribute to online discussions on a huge variety of topics. Newsgroups have been around for as long as the internet and groups like sci.med.diseases and uk.sci.med.nursing can be useful places to ask questions of fellow professionals and contribute to discussions on hot topics.

Conclusion

As we have seen there are many ways to take advantage of the rapidly developing search engine technology, and plenty of techniques to increase the efficacy of your search. However, it is important to keep in mind that web search engines are not always the most efficient tool for finding information. In a recent issue of *Evidence-Based Medicine*, Dr Heneghan, a general practitioner, writes about his search for information online. A search on Google produced 152 000 links, and after browsing through the first few results he concluded that none of them were useful. He writes: 'Internet search engines can be useful for answering background questions, but foreground questions (specific questions on the management of a problem) will often result in a time consuming search'.[7]

Sometimes the problem is with the results you find, but more worrying are the key sources that you might miss. A comprehensive search should take Google, Yahoo! and Scirus into account, but they should only be one stage of your search strategy. Although more advanced search facilities may seem more daunting: 'anyone who invests a short time in learning traditional search technology will save hours in the long run'.[8] For ideas about which sites *should be* your first stop for various different healthcare topics, look back over previous issues of this 'Web alert' column.

REFERENCES

1 Felter LM. Google Scholar, Scirus, and the scholarly search revolution. *Searcher* 2005;13:43–8.
2 Zhao L. Jump higher: analyzing web-site rank in Google. *Information Technology and Libraries* 2004;23:108–18.
3 Childs S. Editorial – internet filters can seriously damage your health. *He@lth Information on the Internet* 2003;31:1–2.
4 Spink A, Yang Y, Jansen J *et al.* A study of medical and health queries to web search engines. *Health Information and Libraries Journal* 2004;21:44–51.
5 Piper PS. Google spawn: the culture surrounding Google. *Searcher* 2004;12:26–34.
6 Butler D. Publishers irritated by Google's digital library. *Nature* 2005;433:446.
7 Heneghan C. The doctor's advice and sleepless nights: what can you find in 5 minutes? *Evidence-Based Medicine* 2005;10:36–7.

8 Kenney B. Googlizers vs. resistors: library leaders debate our relationship with search engines. *Library Journal* 2004;129:44–6.

ADDRESS FOR CORRESPONDENCE

Ben Skinner, Evidence-Based/Knowledge Management Librarian, *KnowledgeShare*, The Library, Sussex Postgraduate Medical Centre, Brighton General Hospital, Elm Grove, Brighton BN2 3EW, UK. Tel: +44 (0)1273 696011 ext. 3704; fax: +44 (0)1273 690032; email: ben.skinner@bsuh.nhs.uk; website: www.KnowledgeShare.nhs.uk