

Google Scholar: A source for clinicians?

The release of the beta version of Google Scholar (<http://scholar.google.com>) in November 2004 generated much media coverage and academic commentary.¹ With this service, Google hopes to make scholarly literature more accessible by indexing academic and professional sources, including peer-reviewed articles, preprints, theses and conference proceedings. This overview evaluates Scholar as an alternative for clinicians seeking information.

Most people, including physicians, who are using the Internet to search for information go first to Google (www.google.ca) because it usually generates useful links quickly. Google, however, emphasizes Web sites that are popular, as measured by the number of links from other sites, and does not weigh quality or date. Searchers are often frustrated by the large number of links Google will generate for common topics.

The ideal tool for finding clinical information would be a fast engine that provides the best hits from scholarly journal literature and clinical resources such as guidelines, perhaps emphasizing sites favoured by physicians in the way that Google emphasizes popular websites for general audiences. Busy clinicians would wish for succinct reviews and for the best evidence, with links to key papers that would be determined as such by the number of times they have been cited, thus balancing popularity with relevance and quality. Features enabling search refinement would be welcome, such as a tool to find related articles by subject or by using links or citations, including more recent articles that cite the retrieved items. Ideally, this engine would provide integrated, powerful access to many sources, including full-text journal literature and textbooks, evidence-based information, information for patients, and drug information, achieving for clinical sources what Google has for the entire public Internet.

The current version of Google Scholar focuses on Internet sites that contain information that is critically appraised, such as the peer-reviewed journal literature, or that are produced by reputable sources, such as universities. Through agreements with publishers, Scholar accesses the “invisible” or “deep” Web, that is, commercial Web sites the automated “spiders” used by search engines such as Google cannot access.² Using text analysis and the number of links from other sites, Scholar rapidly delivers a ranked listing, as Google does. Each item includes the number of links to it — in effect, a citation tracker, providing for free what interfaces such as Web of Science and Elsevier’s Scopus provide at much cost.

Scholar is collaborating with university libraries to develop a way to access full-text journals through institutional subscriptions, so that researchers and physicians affiliated with a university can go directly from a Scholar search to a full-text journal article if their university has a subscription to that journal. Also intriguing is the potential of future versions of Scholar to give free, efficient access to articles from commercial journals reproduced for open access on personal or institutional pages. When development is complete, Scholar may access the better quality sites now accessed by Google, supplemented by the electronic journal literature and additional reputable sources.

How well does Google Scholar do when it searches on medical topics? A search using “Vioxx” generated links to older research articles, with nothing on the drug’s withdrawal from the market in the first pages of citations (the searches for this article were conducted in April 2005). The first hit generated by a Scholar search on “C-reactive protein” was an important article pub-

lished in 2000 in the *New England Journal of Medicine*, but the next 100 hits showed only 9 articles from 2003 and 2004, and none from 2005. “Medicine” limited to 2005 publications resulted in 12 hits. However, a search on the “BODE index” for chronic obstructive pulmonary disease did lead to key 2004 citations. Searches on “cognitive behavioural therapy” and “cognitive behavioral therapy” produced quite different results.

There are other shortcomings. Because results emphasize pages that are cited more often, this creates a bias toward older literature. Many medical links found in both Google and Google Scholar are from PubMed (ncbi.nlm.nih.gov in search results); however, Scholar accesses only 1 million of the some 15 million records at PubMed.³ Although it enables citation searching, Scholar does not offer a “similar pages” feature as Google does to find pages on the same subject. Nor does it offer Google’s “Did you mean” feature, which addresses spelling mistakes and variations by providing alternates. The only major health database used is MEDLINE, which means that the “deep” Web stored in databases that index the journal literature, such as Excerpta Medica, CINAHL (Cumulative Index to Nursing and Allied Health Literature), and PsycInfo, remains difficult to access. These shortcomings make Scholar, for now, a supplementary tool for clinicians at best.

Competitors to Scholar will continue to provide better solutions for clinical information (Table 1). Access to the evidence-based literature continues to improve through tools such as the TRIP database (www.tripdatabase.com; 1 interface to several evidence-based medicine sites), the Cochrane Library (www.cochrane.org; an international collaboration to produce systematic reviews)

and a growing number of commercial products such as InfoRetriever (www.infopoems.com; succinct summaries of important clinical papers). UpToDate (www.uptodate.com; the large online-only textbook), although expensive, remains a key reference-based resource, and competitors such as Clinical Evidence and sets of textbooks such as MDConsult, Access Medicine (including Harrison's), and STAT!Ref are increasingly valuable.

MEDLINE remains the map of the medical literature. The full structure of the MEDLINE database is accessible through powerful interfaces such as Ovid (offered by the CMA's Osler

service) and PubMed, making them vital components of a thorough search of the medical literature. Although it might be hoped that Scholar could become the one-stop, all-encompassing interface integrating all sources for clinicians, the variety of needs and the specialized nature of the literature means that Scholar, even with needed improvements, will remain only one of a battery of information retrieval tools clinicians use.

Google's launch of Scholar indicates the growing sophistication of Internet searchers. It addresses concerns about the quality of information found on the Internet and integrates previously inaccessible, high-quality

commercial sites with more reliable sites available on the public Internet. Google Scholar may develop into a free, sophisticated tool, but, at least in the beta version, it is not a useful choice for clinicians.

Jim Henderson

Health Sciences Library
McGill University
Montréal, Que.

References

1. Google Press Centre. *Media coverage*. Available: www.google.ca/int/en/press/press.html (accessed 2005 Apr 29).
2. Lawrence S, Giles C. Accessibility of information on the Web. *Nature* 1999;400:107-9.
3. Jasco P. *Peter's digital reference shelf*. Available: www.galegroup.com/free_resources/reference/peter/index.htm (accessed 2005 Apr 29).

Table 1: Search engines to consider for clinical topics

Search tool	Audience	Full text	Advantages	Disadvantages
Ovid www.ovid.com or through www.cma.ca * PubMed pubmed.gov or through www.ncbi.nlm.nih.gov	Health professionals, researchers, patients	Ovid: links to much full text PubMed: links to much full text, indicating free and fee	Powerful interfaces to MEDLINE, the highly structured database that indexes clinical and research health journals	Map of the literature, not the literature itself; quantity of research and other health professional literature hides clinical information
TRIP (Turning Research Into Practice) www.tripdatabase.org	Practising physicians, health administrators	Guidelines and reviews in evidence-based medicine sources	Best evidence-based medicine search engine; offers various choices effectively	Complex searches difficult (except for clinical choices feature through PubMed)
HighWire highwire.stanford.edu	Scientific and technical researchers, scholars	Free and pay-per-view journal literature	A wide range of journals including many key clinical titles; good search interface; citation and subject search refinement tools	6 mo to 1 yr of some journal issues not accessible; limited number of journals
Scirus www.scirus.com	Scientific and technical scholars	Links to free and pay-per- view scholarly journal literature	Powerful search interface	Emphasizes research over clinical practice
OALster oalster.umdl.umich.edu Interface to free digital publications, including a growing number of repositories of articles and other digital publications submitted by authors	Researchers, scholars	Includes free full-text articles, preprints as well as other digital materials (e.g., images)	Good search interface; international; growing alternative to commercial sources	Emphasizes research over clinical practice; medicine is a smaller part of existing repositories
Google Scholar Beta test version scholar.google.com	Students, teachers, researchers	Links to full-text sources, many of which are not free	Rapid search of scholarly sites, some private or commercial	Still in development; covers all subjects; complex searches difficult
Google www.google.ca	All Internet users	Free; especially patient- oriented and promotional	Free, rapid search of publicly accessible sites	Quality measured by popularity; much unreviewed information; complex searches difficult
Yahoo, Microsoft search. yahoo.com search. msn.com	All Internet users	Free; especially patient- oriented and promotional	Developing capabilities and increasing number of sites indexed to overtake Google	Relevance ranking inferior to Google

*Free for CMA members when accessed through www.cma.ca.

†A subscription is required, but the Web site offers 5 free searches per week.