
Just what I wanted! The perfect Christmas gift – Google scholar - or is it?

Martin Myhill
Deputy University Librarian, University of Exeter
Tel: 01392 263870
Email: M.R.Myhill@ex.ac.uk

Google Scholar (<http://scholar.google.com>) was launched as a beta product during November 2004 and instantly hit the headlines. I suspect I was not alone in receiving a considerable number of requests from colleagues in academic schools asking the library to put up links to this new search engine or even circulate details of its arrival to the entire university. I suppose I should be grateful that the library was even viewed as the 'gatekeeper' when many of our users often turn to search engines ahead of our catalogue and various databases. While most information professionals rightly respect the provenance of Google, few would regard its search capabilities as the elusive 'Holy Grail' of literature searching in an academic context.

The nomenclature of Google Scholar implies more. The good points of it are:

- Easy access open to all (no membership or login, direct URL link) and it's free
- One step, one stop searching
- Simple 'free text' searching supported by an advanced search mode; a concept well understood by the vast majority of users of web search engines
- Proven development path in a short time span (e.g. advanced search options and 'preferences' added since the launch); supported by a massive organisation
- Academic in outlook – avoids the huge, and often irrelevant 'hits' of Google
- Links include respectable academic sources including Ingenta, Crossref and OCLC WorldCat

- Offers citation information (thereby extending to online references for print-only resources)
- Results are relevance ranked so you see the most appropriate resources first
- Potential – there's a developing world of institutional repositories which will need inter-linking plus new initiatives in digitisation (such as the non-copyright, nineteenth century collections at the Bodleian which Google is closely involved with in addition to other, similar projects)¹. The recently-announced Common Information Environment² has also made reference to the potential of Google Scholar in helping improve access to content for the maximum number of people.
- Add-on's. A number of third-party 'extras' have been developed such as the Firefox extensions developed by Peter Binkley³ and regionalised for the UK by Andy Powell⁴ at UKOLN. The two mentioned maximise the potential of passing OpenURL data to local resolution servers.

Unfortunately for everyone, there are also serious limitations:

- Apart from the current pilot of 'preference sites' which deal with site authentication issues, Scholar only crawls the open web. This is, of course, the iceberg effect as estimates put the 'hidden web' at seven times the size of the open web⁵ – and the hidden bits are often academic resources protected from unauthorised usage. It is possible to find a resource in Ingenta, be prompted to 'pay per view' and eventually find that you have free access via your own institution.
- It isn't possible to undertake a sophisticated, set-based search akin to what a researcher might use on a truly academic database such as PsycINFO.
- The limiting 'academic' criteria employed by Scholar are not widely known and the results are less than comprehensive – even for 'open web' resources. Users may find Google a more successful tool than Scholar - if taken in context. Scholar defines its coverage as 'including peer-reviewed papers, theses, books, preprints, abstracts and technical reports from all broad areas of research'.
- Searching doesn't always find what it should. A significant (and excellent) survey of Scholar's search capabilities was carried out by Jacso⁶. Admittedly, Scholar has developed further since that work was completed but

most of the findings still hold true. It is clear that the search results are often not as accurate (comprehensive and relevant) as they ought to be.

- Some references are 'empty'.
- It is clear that the offerings from CrossRef, and OCLC, in particular, are very limited versions of their full products. Scholar also has agreements with a number of publishers with a strong leaning to science disciplines.
- The OpenURL content of most data passed from Scholar is very limited making the third party extensions of far less value than they might be when many libraries are investing in OpenURL technology.

While 'limited' adequately describes Scholar's current capabilities, it does have potential. However, much of this will only be realised through specific negotiations at institutional level direct with Google to allow the profiling of local resources. At present, metasearch engines do a much better job for academia in this area although those carry a heavy financial cost and profiling load and also lack perfection in areas such as search syntax sophistication.

For anyone intending to use Scholar, I would offer the following advice:

- Read the Jacso review referred to above
- Read the help screen first⁷
- Familiarise yourself with the advanced options.
- Experiment with an area you know well and compare the results from Scholar with other search systems such as metasearching tools, online databases and even Google or Yahoo.
- There is much further discussion at the [Web4LIB] archive⁸

In recent months, there have been a number of search engine announcements. These include offerings from Microsoft⁹ and an academically-orientated development from Yahoo. The latter provides a drop-down 'library' option on the search screen to link to WorldCat records¹⁰. Whatever the drawbacks in an academic setting, search engines have a significant role to play in the seamless delivery of information and have the confidence of our users. At the current time, however, it would seem that a pair of socks would have been a more welcome Christmas present than Google Scholar. There is much room for improvement in this system but I'm sure that will come in time as already demonstrated in the development path and external interest since the beta launch.

References

- 1 Oxford-Google digitisation agreement (2005). <http://www.bodley.ox.ac.uk/news/news58.htm>
- 2 JISC (2005). Common Information Environment – What’s that? http://www.jisc.ac.uk/index.cfm?name=cie_interview_news_100305
- 3 Binkley, Peter (2004) Firefox extension for Google Scholar OpenURLs <http://www.ualberta.ca/~pbinkley/gso/>
- 4 Powell, Andy (2004) Firefox extension for Google Scholar OpenURLs <http://www.ukoln.ac.uk/distributed-systems/openurl/googlescholar/googlescholaropenurl.xpi>
- 5 Eric Digest (2002). Uncovering the hidden web, Part I: Finding what the search engines don’t. <http://www.ericdigests.org/2002-2/hidden.htm>
- 6 Jacso, Peter (2004). [Péter’s Digital Reference Shelf] Google Scholar <http://www.galegroup.com/servlet/HTMLFileServlet?imprint=9999®ion=7&fileName=reference/archive/200412/googlescholar.html>
- 7 Google Scholar help screen (2005). <http://www.scholar.google.com/scholar/about.html>
- 8 Web4Lib Archive (2004). <http://sunsite.berkeley.edu/Web4Lib/archive/0411/>
- 9 The Guardian (2005) ‘Search and you shall find: MSN search’ <http://www.guardian.co.uk/g2/story/0,3604,1403533,00.html>
- 10 Quint, Barbara (2005) OCLC and Yahoo! offer joint toolbar <http://www.infotoday.com/newsbreaks/nb041115-3.shtml>