

# SEARCHER

The Magazine for Database Professionals



## THE **BETTER** MOUSETRAP

by **Laura M. Felter**

Principal Information Consultant  
CInC, Inc.

## Google Scholar, Scirus, and the Scholarly Search Revolution

Most busy professionals like to know up front what the bottom line is, so they can decide whether a piece deserves a chunk of their reading time. So let's have the bottom line of this Better Mousetrap column appear first. Read on!

As of mid-November 2004, a beta of Google Scholar had arrived, forcing information industry professionals to face its challenge. Amidst all the debates and insightful analyses of its searching capabilities and content sources, the bottom line was a choice between worrying or enrolling. This Better Mousetrap column advises all information professionals, whether in the information industry or libraries, to not only embrace Google Scholar, but to expect it to become one of the tools of choice for many of our researchers. The true bottom line? We need to help our researcher clients to integrate their tools of choice, whether Google Scholar and/or Elsevier's Scirus, with our own offerings.

Google Scholar is in its infancy, still in beta-testing, and projected to remain so for some time; enhancements to its search capabilities and added features are anticipated to be on the horizon later in 2005. So, I have chosen to compare Scholar to Scirus at fairly high-level, though, I urge you to stay glued to the listservs and blogs where experienced searchers are really putting Google Scholar through its paces, providing excellent search tips and limitations.

Let's move on to Google Scholar's present position at press time. By the way, I hope this column still correctly refers to Google's new service by the time you read it. As *Searcher* went to press, the American Chemical Society had sued Google for using the word "Scholar" in the name of its new product, accusing Google of creating confusion with the STN SciFinder Scholar service.

### Putting Google Scholar in Its Place

The launch of Google Scholar gives searchers a tremendous opportunity to do the following:

- Compare the free Google Scholar service with other tools that our researchers can access.
- Communicate to our clients about the supplemental tools available.
- Anticipate how output from other tools or research processes could integrate with Google Scholar searching and its results.

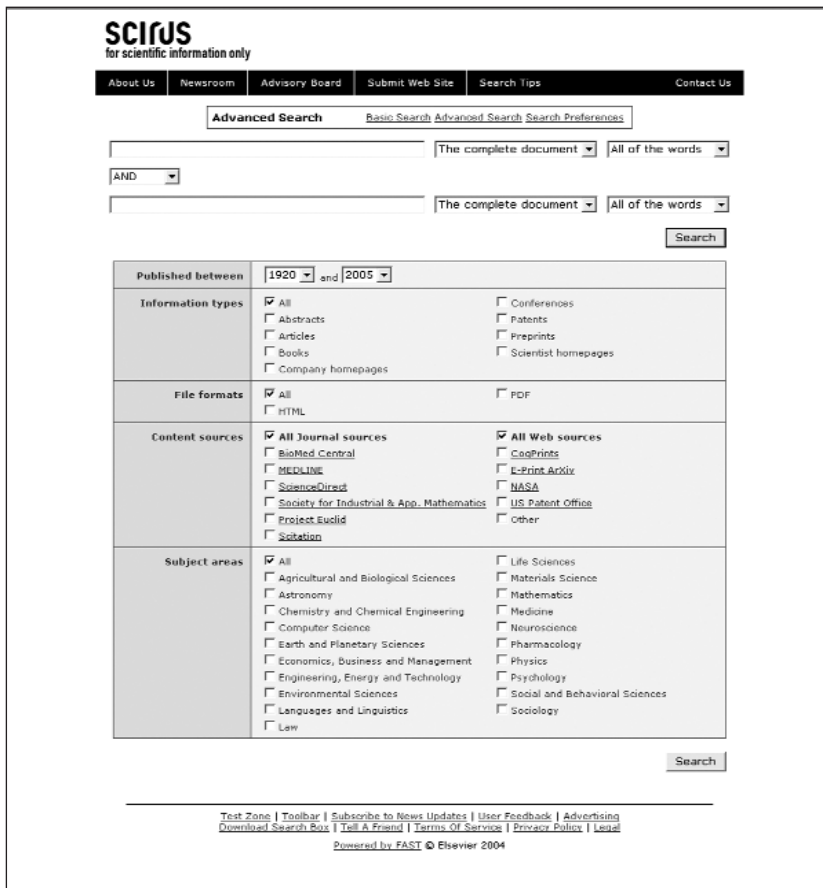
For the moment, I will exclude other sources of scientific and technical content such as institutional and/or fee-based databases. Why? Because most people search the simplest and cheapest sources first, leaving the more complex and more expensive sources for the next step in their online research. Remember, it's all about accessibility, and, quite frankly, without federated search engines, many library



Google Scholar search screen, December 2004



Scirus Basic Search Screen



Scirus advanced search screen

offerings are viewed by end users as a maze of unconnected databases with differing interfaces and search techniques. *And* that's assuming these end users even know which databases to seek out. Let's not even get into the information gaps of rich versus poor, database-wise.

Information professionals usually refer clients pursuing academic and scientific research away from Google Web to more scholarly search engines, such as Scirus, Elsevier's free product. Efforts to move clients away from Google often have mixed results, as some researchers shy away from unfamiliar searching venues. ("Google works just fine for me, thank you.") In a virtual environment, it's even more difficult to convince users to try something new. I've been in situations on the phone when I just *know* clients are not even writing down reminders of my carefully considered recommendations. (Hint: If they don't ask you how to spell it....)

Regardless of the comprehensiveness of the search, Google Inc. will most likely succeed with Scholar simply by shunting serious researchers *already in the Google domain* toward a subset of authoritative content. Even when libraries or institutions offer a federated search engine approach to licensed and unlicensed content, researchers may still bypass it with a Google-to-Google Scholar search technique. Therefore, we must assess whether they are headed in the right direction with Google Scholar and what, if any, additional tools they may require. We also must evaluate whether the subscription-based tools in our arsenal supply us with enough value-added features to make them worth the cost.

## Basic Search Interface as of December 2004

Google Scholar's basic and only search screen as of December 2004 is simplistic by design and mercifully familiar. By comparison, Scirus' basic search screen offers advanced features such as preference settings and sorting opportunities for separating journal and Web content, foreshadowing the offerings available in the Advanced Search interface.

## Advanced Search Interface

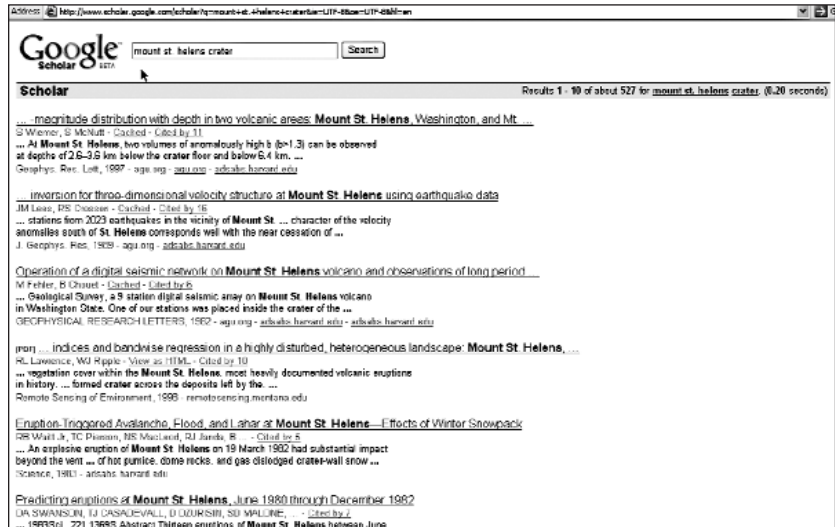
Compared to the Google Scholar search interface, Scirus' advanced search options are a welcome sight to the eyes of the information professional. However, aren't these features the very ones that confound our users and prompt comments such as, "Why can't it be more like Google?" As researchers work with Google Scholar and reach limitations of searching capabilities and options, they may become more receptive to other products.

Strike while the iron's hot and show them Scirus. Limiting results by document type? Check. Subject area? Check. Publication date ranges? Check. Boolean searching by field? Check. Clear identification of content sources? Absolutely. The Content Sources field alludes to what the About Us pages provide in more detail. After a general overview of Scirus, you may even be able to segue your client researcher into looking at other on-site products such as Ei Engineering Village 2 or ProQuest offerings.

To counter the features available now from Scirus, Google Scholar will need to add a clear layout of its sources and a collection of pointers to search by field, from specific sources, and other such advanced searching techniques. The Better Mousetrap anticipates an advanced search page offering for Google Scholar given the type of searches researchers will attempt and the audience Google Scholar is targeting with its content.

## Sample Results Display

In the display of research results and refining search options, you can really see some key winners. Your first sample searches on Scirus will reveal its capability for Citation Searching. This is also a key feature in Google Scholar, and, given the limited availability of such searching in even fee-based databases, Google Scholar's citation searching is worthy of historical note. As a vital research tool for any researcher, citation searching aids in the identification of key documents and authors, which can lead to more refined searches as well as to expanded search strategies.



Google Scholar result display, December 2004



Scirus result display

As of this writing, Google Scholar lacks the clear identification of full-text availability that Scirus so wonderfully illustrates. It takes quite a bit of clicking (or an experienced user) to know which of the multiple sources listed by Scholar could lead to a full-text source. Google Scholar also lacks search refining features such as Scirus' "Refine your search by these keywords," which offers sub-subject areas to refine results. And, addressing my personal gripe about most systems, how about exporting features such as "Save Result Set" and "E-mail Result Set"?

The Scirus capability of refining search sets by proffered keywords and

phrases is an important and useful searching feature. But for those who want to expand their search from the subsets of authoritative content to the World Wide Web, Google Scholar can conduct a Google Web search from Scholar result hits, simplifying the anticipated Google-to-Google search strategy that, inevitably, we all seem to conduct ad infinitum.

## Content and Search Capabilities

Scirus has entirely focused its Web content (.edu, .gov, .com, etc.) on "scientific, scholarly, technical, and

The screenshot shows a Google Scholar search for "calvin and hobbess". The results include "Homicidal Psycho Jungle Cat: A Calvin and Hobbes Collection" and "Calvin v Hobbes". Below the search results is the "Find in a Library" interface, which is powered by WorldCat. It displays metadata for "Calvin and Hobbes" and a table of libraries that hold the book.

Library Name	Distance (mi/km)	City	State/Province/County	Library Type	Library Type
Fort Vancouver Regional Library District	10	Vancouver	Washington	Public	Public
Concordia University	11	Portland	Oregon	Academic	Academic
Portland Community College	17	Portland	Oregon	Academic	Academic
Washington County Cooperative Library	28	Hillsboro	Oregon	Public	Public
Longview Public Library	40	Longview	Washington	Public	Public
Salem Public Library	55	Salem	Oregon	Public	Public

Google Scholar partners with OCLC Open WorldCat library locator service.

medical data,” sub-categories, which you can review in the Advanced Search screen shot. To supplement its all-Web content, Scirus taps material from the following specialty sources: MEDLINE citations, Elsevier ScienceDirect full-text articles, patents from the USPTO, BioMed Central full-text articles, Project Euclid full-text articles, Scitation, and SIAM. From the open archive initiative (OAI), Scirus reaches ArVix.org, NASA technical reports, and CogPrints. [For details on these sources, look at Scirus’ About Us page: <http://www.scirus.com/srsapp/aboutus/>.]

As I write this, Google Scholar content, as described by Barbara Quint in a NewsBreak (“Google Scholar Focuses on Research-Quality Content,” <http://www.infotoday.com/newsbreaks/nb041122-1.shtml>), “...includes a range of publishers and aggregators with whom Google already has standing arrangements, e.g., the Association of Computing Machinery, IEEE, OCLC’s Open WorldCat library locator service, etc.” However,

she also points out that “content in Google Scholar comes from a wide range of academic publishers, professional societies, pre-print repositories, universities, and scholarly articles across the Web.” Many of these sources, e.g., PubMed Central and MEDLINE, match Scirus’ content. Nonetheless, Google Scholar would be well advised to spell out some of the sources it accesses, for it currently sells itself short without such information.

Scholar’s partnership with OCLC provides simplified searching and locating of local library holdings. By using the “Library Search” link of your search hit, you go to OCLC’s WorldCat service, which identifies libraries’ holding items in or near whatever ZIP code the user inserts. Though not a unique partnership (Yahoo! Search offers a toolbar dedicated to facilitating this function: <http://www.oclc.org/toolbar/default.htm>), WorldCat access certainly enhances the use of Google Scholar. Google Scholar currently searches a limited 2-million-records

subset of the 57 million available in the WorldCat database, but reportedly is busy mining the full WorldCat database. Considering this database contains electronic books and journals, Web-based articles, and audio and video recordings, the resource potential here is high.

Google Scholar’s content covers broad areas of research but currently seems to focus on science and technology. Compared to Scirus, however, Google Scholar is certainly a better bet for research in the humanities and the arts, especially due to the availability of OCLC literary citations.

Scirus does a better job of consistently returning authoritative content, but how much of this is due to the provision of user filtering functions (e.g., by topic and field) on top of the relevance-ranking approach Scirus and Google both share? Regardless, be thankful for the options. Scirus is also an established product, having won awards such as the Search Engine Watch Award’s “Best Specialty Search Engine” in 2002 [<http://searchenginewatch.com/awards/article.php/2155921>]. Note that Scirus shared that designation with Google Groups and that Google products won eight out of the other 12 categories. What Google does, it does well, and one can bet Google Scholar will also succeed. Google Scholar is well on its way to becoming the tool of choice with both special and broad topic researchers.

## Integration and Communication — Google Scholar and Your In-House Offerings

Due to Google’s historical free-only stance, it is not anticipated that Google Scholar will partner with any one specific pay-for-reprint service, such as Infotrieve, ScienceDirect, or even the British Library. However, the existence of such a partnership in search engines such as Scirus and ScienceDirect should make information professionals out there consider how to communicate to their users which full-text publications and databases they already license. Google Scholar’s principal engineer, Anurag Acharya,

recognizes the problem. He commented to Quint in her NewsBreak, “So many people do not know that they have access through institutional subscriptions.”

Here are some basic questions for those within organizations to ponder:

- Will Google Scholar ever provide a fee-based service to help organizations integrate their tool with your in-house ones? Would that require your institution's IT department buying into the Google Search Appliance device? Imagine if Google Scholar could offer an organization the capability for users to get search results that highlight the sources accessible in-house. Until such a time, how do you teach researchers using Google Scholar what you have arranged for them through subscriptions?
- In the Google Scholar/OCLC WorldCat partnership, will Google Scholar capitalize on WorldCat's capability of highlighting items owned in your collection?
- If researchers can order a full-text reprint while in Google Scholar, say through an Ingenta hit, who's to say they won't do it despite in-house full-text access because it's simple, quick, efficient, and gosh-darn user-friendly?
- Are your department and other departments prepared to accept this as a cost of doing business?
- How will this user capability impact your current subscriptions to full-text content and document delivery services?
- Do you have plans for getting a product that easily navigates your electronic and print journals, specifically the full-text ones? The demand for this can be expected to rise.

In a nutshell, what does Google Scholar do, or what will it do, that we information professionals should prepare for? When Google Scholar misses some content, we should be prepared — aren't we always? — to have solutions ready for our researchers. Our clients will expect us, eight times out of 10, to help them obtain information they become aware of through Google Scholar or another Web search. For the other two times, we will have to make them aware of what Google Scholar and the Web missed completely.

By the way, Google Scholar's arrival has raised the bar of “value-added” for our subscription-based tools. If Google Scholar can include a feature that identifies which sources are citation, abstract-only, or full-text, then why can't some information aggregators hyperlink us to the file where the full-text version of the abstract we're looking at exists? Why can't some of our subscription-based tools list the citations of records external to their own systems if Google Scholar can do it so easily? What will we do if Google Scholar offers an information visualization tool similar to SciFinder's Panorama or Groxis' Grokker? Whether Google Scholar does these things or not, by offering its plenty for free, it leads us all to re-consider what

we're paying for and what we're demanding for our money (and time).

In the end, it's all about accessibility. The more we can integrate our content and our services with Google Scholar, whether technically or by working proactively with researchers on how to access the world Google Scholar presents to them, the more bang for our buck we will get out of our fee-based subscriptions. Google's Acharya quoted Vannevar Bush in describing the purpose behind Google Scholar:

Mendel's concept of the laws of genetics was lost to the world for a generation because his publication did not reach the few who were capable of grasping and extending it; and this sort of catastrophe is undoubtedly being re-

## Here We Go Again

Before this look at Google Scholar was even finished, Google announced a dramatic expansion of its Google Print program (“Google Print Expands Access to Books with Digitization Offer to All Publishers,” <http://www.infotoday.com/newsbreaks/nb041006-1.shtml>). Google has arranged to digitize all — or significant parts of — five of the world's largest research library collections. Millions and millions of out-of-print and in-print, out-of-copyright and in-copyright books will go digital. Search results from the Google Print program receive privileged positioning on the first page of Google search result displays.

At launch, Google had no specific plans to merge this product with Google Scholar, but Adam Smith, the Google Print library program product manager, did admit that it seemed a natural fit. [For a description of the new program, check the NewsBreaks on [<http://www.infotoday.com/newsbreaks> for December 20 and 27, 2004.]

...bq

P.S.

At press time, Elsevier announced the availability of a free, downloadable toolbar for Scirus. The toolbar searches Scirus directly as well as Scirus Web Sources and/or Journal Sources. A pop-up blocker removed online ads. A highlighter colors search terms as the terms appear on a Web page. It allows users to view a previous search. Searchers who find interesting sci-tech Web sites can submit them to the Scirus Team through the toolbar. Interested searchers can find the toolbar at the Scirus site [<http://www.scirus.com>].

peated all about us, as truly significant attainments become lost in the mass of the inconsequential.

—**Vannevar Bush**, “As We May Think,” *The Atlantic Monthly*, July 1945

## Final Words

I’m risking lecturing here, but we have no choice as information professionals but to embrace Google Scholar as well as Scirus and prepare for their widespread use. We info pros must avoid giving the appearance of being wary of any tool, especially their re-

searchers’ tool of choice. We must take advantage of Google Scholar as an opportunity to introduce our researchers to what they can access and to teach them search habits that best integrate Google Scholar and/or Scirus with institutional resources. If we can obtain products or streamline existing ones to help with this effort, all the better. But that’s not enough. We have to anticipate this need or else we risk appearing reactive versus proactive — a perception that is the death knell of any department or business in any industry out there.

As for me, I am ever so thankful that scholarly search engines such as Scirus and Google Scholar exist at the exact time I left the corporate world and its multitude of commercial database offerings. And guess what? For at least one of these products, we don’t even have to remember how to spell it. In fact, you can just “google it.” ♦

---

**Postscript:** For further readings and updates post-article deadline, please refer to the *CinC, Inc.* Web site at [www.cincinc.com](http://www.cincinc.com).