Google Scholar: experiences ideas and plans

Gordon C. Tibbits
Thoughts presented by
Gordon Tibbitts
President
Blackwell Publishing
September 14, 2005
Google Scholar’s Success! or?

• This Presentation
  – Where are we?
  – Environmental analysis (this publishers perspective)
  – Product scope (Google’s scope)
  – A SWOT
  – Building a bright future
• A convergence of needs has resulted in:
  – broader support for federated searching
  – the launch CrossRefSearch
  – content-providers getting indexed by Google
  – And Google Scholar which serves these and some other real and perceived needs

• Let me take you through a few slides of feed-back from the scholarly community
On Google Scholar

8.16.2005

UC Library Staff Use of Google Scholar

The CDL put together [this fine report (pdf)] based on a survey of University of California Libraries.

On June 22, 2005, the CDL requested information from the campuses about librarian and library staff use of Google Scholar in their own work and at public service desks. Eight of ten campuses responded with a wealth of information about the creative ways in which the libraries use Google Scholar, as well as with their objections to its use. Immediately below is an overall summary of responses, followed by a document containing all the detailed responses received. At the end of the second section is a report from UCLA detailing how the UCLA library integrates and positions Google Scholar along with the rest of their electronic resources.

posted by tjs ondermann at 11:24 AM | 1 comments

8.02.2005

Uses

Uses for Google Scholar:

- For citation verification
- To locate chapters within multi-authored books and conference proceedings
- For interdisciplinary scholarship
- For cross-disciplinary topics not found via searches in traditional bibliographic databases
- As a quick starting point to decide which subject resource is appropriate
- For background info
- Good start up database/search engine
- To learn background info on a topic
- Use its SFX capabilities
- Only after exhausting other resources
- Picks up web-based materials not in databases
- To try to replicate user’s search
- It’s faster for locating a full text source
- A fast way to search for both free and library-only stuff at the same time
- When need something from a source where it is not accurately covered in our subscribed A&I databases
http://schoogle.blogspot.com/

**Uses**

**Reasons for usefulness:**
- Its speed
- Allows libraries to brand with SFX
- It’s fast and simple
- It’s freely available
- It’s open access
- May have government documents eventually
- It is a diverse database that covers book catalogs, journal publishers' sites, article databases such as PubMed, etc.
- It is another entry point for references and papers not found through normal channels
- The technology allows me to be as vague or as specific as I need/want.
- It's a great mechanism to find the pivotal works in a particular subject (the works that appear on a variety of sites are good candidates).
- Simplicity of its search interface
- Quick and simple
- It's easy and quick to search and linking to full text is available
- It searches things Google doesn’t search
Uses

- **I don’t use it because:**
- Have not found it to be useful
- I do use it, but not as a substitute for scholarly indexes and databases.
- Students usually want some verification from me or their professor that what they’ve found is really peer reviewed and from a reliable source.
- We have access to better tools
- We have other excellent resources; no need to use it
- Use “real” databases instead
- can do a much better search in a subject specific database
- Advanced Search lines in Google Scholar don't offer me as much searching power as, say, the Advanced Search lines in the CSA databases
- I'm waiting for Google Scholar to evolve some more--it's in its early stages. I'm mostly in a wait-and-see mode.
- I’m not sure what it includes
- It’s hit or miss
Needs and Wants:

- **Sorts** (ex. Publication, period, author, time cited, ability to re-sort current result set)
- **Selects** (ex. by: subject, descriptor [ontology], topic, and/or by material)
- ? “cited-by related items at a ‘older’ date”
- **Discipline** specific searching (ex. Database interfaces, taxonomies, demographics,...)
Critique and Universal Challenges

• “Unfortunately, none are even close to comprehensive. Each tool covers one segment exclusively or in very different ways.”

• “At this point, each covers a certain segment of scholarly material, but plenty of problems remain. Other search tools continue to serve the scholarly community better.”
Critique and Universal Challenges

• **Coverage**: how much is available to the spiders vs. what’s seems to be shown. What is not available (not searched) could be increase and is a different concern. * Missing year of pubmed content

• **Search stability**: search 15 articles separately all are there. Search using a collective (AND) term and not all 15 show.
Critique and Universal Challenges

- **Search stability**: how does consolidation of like articles across multi-sources (doi’s) work? Will it ever work?

- **Ranking**: “They may guess that records that matched the query term in the title field are ranked ahead of the ones with higher citedness score, but this does not seem to hold true”

- **Bug-Fixing**: links, Citedness
Opportunities

- **exploit** the highly structured and tagged Web pages with rich metadata readily available [from] scholarly publishers
- **create** field-specific indexes for distinct data elements
- offer … pull-down menus for limiting the search by publisher, journal, document type, publication year, etc.
- **consolidate** cited references through the ever increasing DOI registry
- **develop** utilities that enable libraries to launch a known-item federated search in the full-text aggregators' databases licensed by the library in order to check if any have the document from a journal that is not licensed digitally from the publisher
Opportunities

• **Reporting**: Show # articles, # authors [with abstracts, TOC’s, reviews, letters to the E., removed/hidden]

• **Query Improvement**: Allow variations of authors names semi-disambiguated for the authors to fully-disambiguate

• **Un-authorized copy detection** and automated removal

• Note: Universities are starting to put the link on their web pages (ex. U of Texas)
Environment

Primary Scholarly Journals
Books e-Content

Additional Scholarly IR's Other Pre-Prints

Users & Uses Student Public Research

Products & Bus. Models Find Sell Discover

Google Scholar BETA
More than one choice

Google Scholar or PsycINFO?
- I have done my search. Now how do I get the material?
- Return to Introduction

Google Scholar
- About
- FAQ
- Reading fulltext in Google Scholar

PsycINFO
- About
- FAQ

G2O Access
- About
- FAQ

government documents

SciCus
- About
- FAQ

science, medical, technical data

SMEAL Search
- About
- FAQ

business, technology

SOSIG
- About
- FAQ

social sciences

Voice of the Shuttle
- About
- FAQ

humanities, music, arts
* In the eyes of the beholder
<table>
<thead>
<tr>
<th></th>
<th>Most used, Already beneficial for all!, G’s “Can Do!,” Capital</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Built for Dynamic data, partner? Dedicated G resources?, Freq. of updates</td>
</tr>
<tr>
<td></td>
<td>Work close with scholarly comm., tie into user workflow, content use reporting, plagiarism/ misuse reporting</td>
</tr>
<tr>
<td></td>
<td>Others “Do It” better than G, Alienating partners, Conflict with content-provider Rev/models, Good federated engines, G’s conglomeration</td>
</tr>
</tbody>
</table>
Building a Bright Future

• Content providers, Google and others can help make the future brighter:
  – **Work as a community** to improve meta-data standards and practice
  – **Facilitate integration** of content with library and research services: Fewer steps to content- Fewer steps to answers
  – Facilitate competitive forces while **keeping an eye on value stake-holders** bring and the value of sustaining business models
References


Reading:

http://www.infotoday.com/online/jul05/OnTheNet.shtml