2008-07-08

How Scholarly Is Google Scholar? A Comparison of Google Scholar to Library Databases

Jared L. Howland
jared_howland@byu.edu

Thomas C. Wright
tom_wright@byu.edu

See next page for additional authors

Follow this and additional works at: https://scholarsarchive.byu.edu/facpub

Part of the Library and Information Science Commons

Original Publication Citation

BYU ScholarsArchive Citation
https://scholarsarchive.byu.edu/facpub/1263

This Presentation is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in All Faculty Publications by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.
How Scholarly is Google Scholar?
A Comparison of Google Scholar to Library Databases

Jared L. Howland | Thomas C. Wright | Rebecca A. Boughan | Brian C. Roberts
Brigham Young University
Methodology

Subject specialists

7 Subject specialists from 3 disciplines:
  3 from sciences
  2 from humanities
  2 from social sciences
Blind to purpose of study
Asked them to give us 3 things...
How does the acquisition and use of a second language in children affect their general cognitive development?
A structured query they would use to search a database:

(bilingual* OR L2) AND (child* OR toddler) AND “cognitive development”
The database they would use to search for that question
<table>
<thead>
<tr>
<th>Academic Discipline</th>
<th>Database Query</th>
<th>Library Database</th>
</tr>
</thead>
<tbody>
<tr>
<td>Science</td>
<td>(ACL or “anterior cruciate ligament”<em>) and injur</em> and (athlet* or sport or sports) and (therap* or treat* or rehab*)</td>
<td>SportDiscus</td>
</tr>
<tr>
<td>Science</td>
<td>lung cancer and (etiol* or caus*) and (cigarette* or smok* or nicotine*)</td>
<td>Medline</td>
</tr>
<tr>
<td>Science</td>
<td>“dark matter” and evidence</td>
<td>Applied Science and Technology Abstracts</td>
</tr>
<tr>
<td>Social Science</td>
<td>(“fast food” or mcdonald’s or wendy’s or “burger king” or restaurant) and franchis* and (knowledge n3 transfer or “knowledge management” or train*)</td>
<td>Business Source Premier</td>
</tr>
<tr>
<td>Social Science</td>
<td>(“standardized test”<em>” or “high stakes test”</em>”) and (“learning disabilit*” or Dyslexia or “learning problem”) and accommodat*</td>
<td>PsycINFO</td>
</tr>
<tr>
<td>Humanities</td>
<td>(bilingual* or L2) and (child* or toddler) and “cognitive development”</td>
<td>Linguistics and Language Behavior Abstracts</td>
</tr>
<tr>
<td>Humanities</td>
<td>(memor* or remembrance or memoir*) and (holocaust) and (Spiegelman or Maus)</td>
<td>JSTOR</td>
</tr>
</tbody>
</table>

This is what things looked like after we got all the information back from the librarians.
Then we took that information and used it in 2 ways.
The first was to actually run the search query in the suggested database.

We put the first 30 citations into a bibliographic citation manager and saved all of the actual full text.

We chose 30 because usability studies (Jakob Neilsen) tell us that less than 1% of all users ever go beyond the 3rd page of results and very few people ever change the defaults (ie, once they run a search they stick with it, success or failure).

Most of our DBs present 10 results per page so 30 results should represent a large enough sample to represent the actual set of results the majority of our users is ever going to see after performing a search.
We ran the same query in Google Scholar and saved the results again in a bibliographic Manager.

We used Zotero to quickly export all of the results.

We also saved the full text of each citation for later use in our study.
So, the first searches we ran using the native DBs and GS was for the query given to us by the librarian.

The second set of searches we ran was to see if the citations we found in the DB were available in GS and vice versa.
Is this citation available in Google Scholar?

Here is the same screenshot we saw just a minute ago.

We took the bibliographic information for each citation and searched for the citation within Google Scholar.
We then did the same thing in reverse.
We took the 30 results from GS and searched for each citation within the database
This allowed us to later calculate something we called “exclusivity”

We put the citations into 1 of 3 possible “exclusivity” categories

Shows proportion of citations within our study that overlap. As you can see, within our study we found that, on average, GS had a larger result set overall as well as more exclusively than the databases.
So now that we have the citations from the database and the citations from Google Scholar. We used the bibliographic manager to generate a list of references that we input into an Excel spreadsheet. Then, using a random number table, we completely randomized the order of the citations for each subject specialist.
Finally, to deliver the content to the librarians in a way in which it would be easiest for them to evaluate, we saved the full-text of each citation according to its randomly assigned citation number. Then we used Excel to create hyperlinks to the full-text of each citation and delivered this list along with the full-text on a CD to the subject librarians. We asked them to evaluate each citation using a rubric which we provided in hard copy form. As you can see, the subject librarians were only able to see the citation number and the bibliographic information. By clicking on the hyperlinked citation number, the full-text of that citation would appear and the subject librarians could easily rate the citation on the rubric.

Have full text appear on this page after click to simulate linking from provided document.
This screen shows the rubric that we used. It is based on a rubric that has popularly been used to evaluate print resources (Alexander, 1999)


We asked each subject librarian to assign a score of between 1 and 3 within 6 different categories to each of the citations (1 was below average, 2 was average and 3 was above average).

These six categories were:
- **Accuracy** – which looks at
  - reliability, fact checkers/editors, peer review
- **Authority** – specifically the
  - author’s qualifications, reputable publisher
- **Objectivity** – looking for
  - minimum bias, extent to which persuasion is the goal
- **Currency** – is the information up to date?
  - information up to date, date of publication indicated
- **Coverage** – how deep is the Coverage
- **Relevancy** – how well does the citation relate to the research question

This resulted in a total possible score of 18 for each citation – we called this a scholarliness score
Methodology

total scholarliness score = $\mu + E_i + L_j + EL_{ij} + \varepsilon_{ijkl}$

where

$\mu$ = Average total score
$E_i$ = Effect due to exclusivity ($i = 1, 2, 3$)
$L_j$ = Effect due to librarian ($j = 1, 2, \ldots 7$)
$EL_{ij}$ = Interaction between exclusivity and librarian
$\varepsilon_{ijkl}$ = Error term

We used this statistical model to evaluate the data. Essentially this formula says 2 important things about the way we used the data:

1. We controlled for the differences between the way librarians grade
2. We controlled for the differences in how exclusively the citation was available

This allowed us to pinpoint and measure any differences there may have been between disciplines in our data as well as any differences that can be attributed to the source of the citations
Results

Google Scholar was 17.6% more scholarly

Citations found only in GS had, on average, a 17.6% higher scholarliness score than citations found only in the DB
Results

Highest scholarliness score when found in both

Citations found in both GS and the DB were even higher than citations found only in GS
Results

No difference between disciplines

We found no statistically significant difference in the scholarliness scores between disciplines (ie, humanities citations in GS are just as scholarly as science citations found in GS)
AVERAGE 76% 77% 83% 100% 100% 97% 93% 82% 82% 82%
Physics 47% 47% 43% 97% 80% 28% 47% 35%
Linguistics 47% 47% 43% 97% 80% 28% 47% 35%
Biology 100% 100% 100% 100% 100% 100% 100% 100%
Medicine 97% 97% 97% 97% 97% 97% 97% 97%
Literature 80% 80% 80% 80% 80% 80% 80% 80%
Business 28% 28% 28% 28% 28% 28% 28% 28%
Education 47% 47% 47% 47% 47% 47% 47% 47%

- database citations in GS
- GS citations in database

Tuesday, July 8, 2008
This study can only be extrapolated statistically to the specific topics and subject specialists used in this study.

A more robust statistical methodology would need to be employed to make these results generally applicable.

We are encouraged by the results we received and feel that they would probably hold up but cannot say so until another study is done.
Future Studies

Improved rubric

If we had to do it over again, we would have increased the Likert scale on our rubric from 1–3 to 1–7 or 1–10.

This would have allowed for a more nuanced statistical analysis and made it easier to spot significant differences, if any, between GS and databases.
Our scholarliness calculation, ultimately, was based on the subjective opinions of librarians with subject expertise.

There are lots of ways to create a scholarliness score (citation counts, impact factors, etc). Which is best is still debatable.
Our study compared GS to individual library databases. A more appropriate comparison may be GS to federated search tools.
Questions?
jared_howland@byu.edu
tom_wright@byu.edu
rebecca_boughan@byu.edu

http://dspace.byu.edu/handle/1877/634