



Evidence Summary

Google Scholar Out-Performs Many Subscription Databases when Keyword Searching

A Review of:

Walters, W. H. (2009). Google Scholar search performance: Comparative recall and precision. *portal: Libraries and the Academy*, 9(1), 5-24.

Reviewed by:

Giovanna Badia

Librarian, Royal Victoria Hospital Medical Library, McGill University Health Centre
Montreal, Quebec, Canada

Email: giovanna.badia@mail.mcgill.ca

Received: 2 June 2010

Accepted: 19 July 2010

© 2010 Badia. This is an Open Access article distributed under the terms of the Creative Commons-Attribution-Noncommercial-Share Alike License 2.5 Canada (<http://creativecommons.org/licenses/by-nc-sa/2.5/ca/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly attributed, not used for commercial purposes, and, if transformed, the resulting work is redistributed under the same or similar license to this one.

Abstract

Objective – To compare the search performance (i.e., recall and precision) of Google Scholar with that of 11 other bibliographic databases when using a keyword search to find references on later-life migration.

Design – Comparative database evaluation.

Setting – Not stated in the article. It appears from the author's affiliation that this research took place in an academic institution of higher learning.

Subjects – Twelve databases were compared: Google Scholar, Academic Search Elite, AgeLine, ArticleFirst, EconLit, Geobase, Medline, PAIS International, Popline, Social Sciences Abstracts, Social Sciences Citation

Index, and SocIndex.

Methods – The relevant literature on later-life migration was pre-identified as a set of 155 journal articles published from 1990 to 2000. The author selected these articles from database searches, citation tracking, journal scans, and consultations with social sciences colleagues. Each database was evaluated with regards to its performance in finding references to these 155 papers.

Elderly and *migration* were the keywords used to conduct the searches in each of the 12 databases, since these were the words that were the most frequently used in the titles of the 155 relevant articles. The search was performed in the most basic search interface of each database that allowed limiting results by the needed publication dates (1990-2000).

Search results were sorted by relevance when possible (for 9 out of the 12 databases), and by date when the relevance sorting option was not available. Recall and precision statistics were then calculated from the search results. Recall is the number of relevant results obtained in the database for a search topic, divided by all the potential results which can be obtained on that topic (in this case, 155 references). Precision is the number of relevant results obtained in the database for a search topic, divided by the total number of results that were obtained in the database on that topic.

Main Results – Google Scholar and AgeLine obtained the largest number of results (20,400 and 311 hits respectively) for the keyword search, *elderly* and *migration*. Database performance was evaluated with regards to the recall and precision of its search results. Google Scholar and AgeLine also obtained the largest total number of relevant search results out of all the potential results that could be obtained on later-life migration (41/155 and 35/155 respectively). No individual database produced the highest recall for every set of search results listed, i.e., for the first 10 hits, the first 20 hits, etc. However, Google Scholar was always in the top four databases regardless of the number of search results displayed. Its recall rate was consistently higher than all the other databases when over 56 search results were examined, while Medline out-performed the others within the first set of 50 results.

To exclude the effects of database coverage, the author calculated the number of relevant references obtained as a percentage of all the relevant references included in each database, rather than as a percentage of all 155 relevant references from 1990-2000 that exist on the topic. Google Scholar ranked fourth place, with 44% of the relevant references found. AgeLine and Medline tied for first place with 74%.

For precision, Google Scholar ranked eighth among the 12 databases when the complete set of search results was examined, but ranked

third within the first 20 search results listed. Within the first 20, 55% of the search results were relevant. This precision rate put Google Scholar in third place, after Medline (80%) and Academic Search Elite (70%).

Google Scholar's precision and recall statistics may have been positively affected by its search for a keyword in the full-text content of indexed articles, rather than just searching in the bibliographic records as is the case for the other 11 databases. The author re-calculated the recall and precision rates for a title search in Google Scholar using the same keywords, *elderly* and *migration*. Compared to the standard search on the same topic, there was almost no difference in recall or precision when a title search was performed and the first 50 results were viewed.

Conclusion – Database search performance differs significantly from one field to another so that a comparative study using a different search topic might produce different search results from those summarized above. Nevertheless, Google Scholar out-performs many subscription databases – in terms of recall and precision – when using keyword searches for some topics, as was the case for the multidisciplinary topic of later-life migration. Google Scholar's recall and precision rates were high within the first 10 to 100 search results examined. According to the author, "these findings suggest that a searcher who is unwilling to search multiple databases or to adopt a sophisticated search strategy is likely to achieve better than average recall and precision by using Google Scholar" (p. 16). The author concludes the paper by discussing the relevancy of search results obtained by undergraduate students. All of the 155 relevant journal articles on the topic of later-life migration were pre-selected based on an expert critique of the complete articles, rather than by looking at only the titles or abstracts of references as most searchers do. Instructors and librarians may wish to support the use of databases that increase students' contact with high-quality research documents (i.e., documents that are authoritative, well written, contain a strong analysis, or demonstrate

quality in other ways). The study's findings indicate that Google Scholar is an example of one such database, since it obtained a large number of references to the relevant papers on the topic searched.

Commentary

This study evaluated keyword searching in Google Scholar by calculating the recall and precision rates of the search results with regards to finding references to a pre-established set of 155 relevant papers on the topic. These relevant papers were selected by looking at several factors, such as the subject and importance of conclusions, in the content of the complete articles. According to the author, evaluating Google Scholar's search results with a definition of relevancy that is based on the content of the published literature that exists on the search topic, rather than looking at just the titles and abstracts of references, is what makes this particular study unique.

The study's findings suggest that Google Scholar will obtain above average recall and precision when using a keyword search to find references on a multidisciplinary topic. There are confounding variables in the study that may contradict these findings. The search terms used in each database were based on the two words that appeared most frequently in the titles of the 155 relevant articles. Rather than measuring the search performance of each database, the study may have actually assessed the author's search strategy. The search strategy was very precise and strongly favoured the relevant papers that were pre-selected on the topic. Furthermore, this does not match the author's intention, which was to reflect the actual behavior of inexperienced searchers. Instead, the search strategy does the opposite, since it is unnatural; searchers will not know the exact words that are used in the titles of the majority of documents on a specific topic.

A major concern for this reviewer is that this study does not assess Google Scholar's search performance in obtaining references to recently published documents, which is an

extremely important factor for examining the relevancy of search results on many topics, especially in the health sciences. Google Scholar's recall and precision rates were calculated based on finding references published between 1990 and 2000.

Despite its weaknesses, this study improves our understanding of recall and precision for keyword searching in each of the 12 databases examined. This will help reference librarians to recommend the best database to inexperienced searchers who wish to find a few relevant papers on a specific topic.