



Evidence Summary

Web-Scale Discovery Services Retrieve Relevant Results in Health Sciences Topics Including MEDLINE Content

A Review of:

Hanneke, R., & O'Brien, K. K. (2016). Comparison of three web-scale discovery services for health sciences research. *Journal of the Medical Library Association*, 104(2), 109-117.
<http://dx.doi.org/10.3163/1536-5050.104.2.004>

Reviewed by:

Elizabeth Stovold
Information Specialist, Cochrane Airways Group
St George's, University of London
Tooting, London, United Kingdom
Email: estovold@sgul.ac.uk

Received: 3 Mar. 2017

Accepted: 21 Apr. 2017

© 2017 Stovold. This is an Open Access article distributed under the terms of the Creative Commons-Attribution-Noncommercial-Share Alike License 4.0 International (<http://creativecommons.org/licenses/by-nc-sa/4.0/>), 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 results of health sciences search queries in three web-scale discovery (WSD) services for relevance, duplicate detection, and retrieval of MEDLINE content.

Design – Comparative evaluation and bibliometric study.

Setting – Six university libraries in the United States of America.

Subjects – Three commercial WSD services: Primo, Summon, and EBSCO Discovery Service (EDS).

Methods – The authors collected data at six universities, including their own. They tested each of the three WSDs at two data collection sites. However, since one of the sites was using a legacy version of Summon that was due to be upgraded, data collected for Summon at this site were considered obsolete and excluded from the analysis.

The authors generated three questions for each of six major health disciplines, then designed simple keyword searches to mimic typical student search behaviours. They captured the first 20 results from each query run at each test site, to represent the first “page” of results, giving a total of 2,086 total search results. These were independently assessed for relevance to the topic. Authors resolved disagreements by discussion, and calculated a

kappa inter-observer score. They retained duplicate records within the results so that the duplicate detection by the WSDs could be compared.

They assessed MEDLINE coverage by the WSDs in several ways. Using precise strategies to generate a relevant set of articles, they conducted one search from each of the six disciplines in PubMed so that they could compare retrieval of MEDLINE content. These results were cross-checked against the first 20 results from the corresponding query in the WSDs. To aid investigation of overall coverage of MEDLINE, they recorded the first 50 results from each of the 6 PubMed searches in a spreadsheet. During data collection at the WSD sites, they searched for these references to discover if the WSD tool at each site indexed these known items.

Authors adopted measures to control for any customisation of the product setup at each data collection site. In particular, they excluded local holdings from the results by limiting the searches to scholarly, peer-reviewed articles.

Main results – Authors reported results for 5 of the 6 sites. All of the WSD tools retrieved between 50-60% relevant results. EDS retrieved the highest number of relevant records (195/360 and 216/360), while Primo retrieved the lowest (167/328 and 169/325). There was good observer agreement ($k=0.725$) for the relevance assessment. The duplicate detection rate was similar in EDS and Summon (between 96-97% unique articles), while the Primo searches returned 82.9-84.9% unique articles.

All three tools retrieved relevant results that were not indexed in MEDLINE, and retrieved relevant material indexed in MEDLINE that was not retrieved in the PubMed searches. EDS and Summon retrieved more non-MEDLINE material than Primo. EDS performed best in the known-item searches, with 300/300 and 299/300 items retrieved, while Primo performed worst with 230/300 and 267/300 items retrieved.

The Summon platform features an “automated query expansion” search function, where user-entered keywords are matched to related search terms and these are automatically searched along with the original keyword. The authors observed that this function resulted in a wholly relevant first page of results for one of the search questions tested in Summon.

Conclusion – While EDS performed slightly better overall, the difference was not great enough in this small sample of test sites to recommend EDS over the other tools being tested. The automated query expansion found in Summon is a useful function that is worthy of further investigation by the WSD vendors. The ability of the WSDs to retrieve MEDLINE content through simple keyword searches demonstrates the potential value of using a WSD tool in health sciences research, particularly for inexpert searchers.

Commentary

Previous studies such as Ketterman and Inman (2014) have sought to compare WSDs directly with traditional bibliographic databases. However the authors of this study highlight research into typical library user behaviour that shows a preference for Google-style searching over traditional methods due to ease, efficiency, and relevance ranking. An assessment of WSD system performance using relevance of the results as an indicator is therefore warranted.

This study was evaluated using Perryman’s (2009) critical appraisal tool for bibliometric studies. The objectives are clearly stated and the methodology is described in detail for each aspect of the study. The chosen search questions are based on real life examples, and the retrieval methods are designed to reflect common user behaviours, and therefore both are appropriate for the stated aims of the study. All of the search strategies are included in the online appendices, and the processes for data collection and handling are well documented. Overall the methods section of this paper is strong and the authors provide an equally robust discussion of the limitations of their study, together with the controls they put

in place to help mitigate these, such as duplicate screening of the results when assessing for relevance.

Results from each strand of the study are clearly presented, however it would be helpful to see the tabulated results in percentages as well as absolute numbers so that the reader is able compare the performance of each WSD more easily. The authors collected a large amount of data and it would be interesting to see more reporting of this information, particularly the relevance assessments per search query, as the authors noted in their discussion section that relevance was often a function of the topic.

Although the authors were not able to recommend one WSD tool over the other, this study is a good starting point for library professionals considering promoting one of these tools to their library users or implementing one of these products in their library. There are many other issues to consider when evaluating a WSD, such as usability and compatibility with other library tools, and these are recognised by the authors. Deodato's (2015) comprehensive guide to conducting a full evaluation of WSDs is a useful resource.

The key finding of this study is the ability of WSD products to retrieve MEDLINE content with simple searches representative of typical student search behaviours. This has

implications for health sciences librarians who are involved in the training and education of library users and the selection of library resources. There are opportunities for further research to see if the findings of this study are consistent across other test sites and in different health science disciplines, and more studies designed to directly compare the performance of WSDs with MEDLINE are needed.

References

- Deodato, J. (2015). Evaluating web-scale discovery services: A step-by-step guide. *Information Technology and Libraries*, 34(2), 19-75.
<http://dx.doi.org/10.6017/ital.v34i2.5745>
- Ketterman, E., Inman, M. E. (2014). Discovery tool vs. PubMed: A health sciences literature comparison analysis. *Journal of Electronic Resources in Medical Libraries*, 11(3), 115-123.
<http://dx.doi.org/10.1080/15424065.2014.938999>
- Perryman, C. (2009). *Evaluation tool for bibliometric studies*. Retrieved from Carol Perryman website:
<https://www.dropbox.com/l/scl/AAAL7LUZpLE90FxFnBv5HcnOZ0CtLh6RQrs>