



Evidence Summary

Multiple Databases are Needed to Search the Journal Literature on Computer Science

A Review of:

Cavacini, A. (2015). What is the best database for computer science journal articles? *Scientometrics* 102(3): 2059-2071. <http://dx.doi.org/10.1007/s11192-014-1506-1>

Reviewed by:

Giovanna Badia
Liaison Librarian
Schulich Library of Science & Engineering
McGill University
Montreal, Quebec, Canada
Email: giovanna.badia@mcgill.ca

Received: 18 Sep. 2015

Accepted: 04 Nov. 2015

© 2015 Badia. 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.

Objective – To compare the coverage of computer science literature in four bibliographic databases by checking the indexing of a selection of journal articles. The purpose of this comparison was to identify the most comprehensive database in computer science and determine whether more than one database is needed to search for articles on computer science topics.

Design – Comparative database evaluation using citation analysis.

Setting – Computer science journal literature found within the INSPEC, Scopus, Web of Science, and DBLP databases.

Subjects – 1,135 computer science journal articles published by an Italian university's researchers from 1979 to 2014.

Methods – The University of Milan's institutional repository (AIR), containing publications authored by the university's researchers, was searched in October 2014 for journal articles that were assigned the subject heading “*informatica*” (the word for computer science in Italian). The author then searched the titles of these journal articles in each of the databases to check whether they were indexed. For articles indexed in all four databases, the author also examined the quality of the bibliographic records by looking for the presence of 20 elements (e.g., the “cited by” option, ranking of search results, precision of results, etc.) in each database's record. These overlapping articles were also searched in Google Scholar to help compare the quality of the records between the databases.

Main Results – Scopus indexed 75.86% of the journal articles found in AIR, Web of Science indexed 64.49%, DBLP indexed 61.15%, and INSPEC indexed 53.39%. Web of Science and INSPEC put together covered 74.80% of the articles, which is comparable to the amount indexed by Scopus. DBLP and Scopus contained the highest number of references to articles that were not found in the other databases, about 4% each. Out of the 1,135 journal articles, 391 (34.45%) were indexed by all four databases, with Web of Science scoring the highest for providing the best quality bibliographic records for these articles.

Conclusions – According to the author, the findings showed that INSPEC, Scopus, Web of Science, and DBLP “complemented each other, in a way that neither one could replace the other” (p. 2068) when searching the computer science literature. While there was overlap between databases, they each also contained unique articles.

Commentary

Based on the author’s literature review, there are many published studies that have compared two or more of the same databases being contrasted in this study, and examined one or more databases that index the computer science literature. However, it seems that none, thus far, have compared INSPEC, Scopus, Web of Science, and DBLP for their coverage of the computer science literature, or used publications from an institutional repository as the source of data for making database comparisons.

The EBL Critical Appraisal Checklist (Glynn, 2006) was employed to help objectively determine the strengths and limitations of this study. Its strengths lie with the data collection and study design. The author clearly outlines the procedure he utilized to conduct the research and describes the results concisely. Readers would be able to easily replicate the methodology.

The major limitation of this study concerns the study population, in this case the source of data used to compare the databases. As noted

by the author, only journal article titles that were assigned the computer science subject heading were extracted from the University of Milan’s institutional repository and used to compare the four databases. These journal articles accounted for only 29.64% of the total number of documents (1,135 out of 3,828 documents) in the repository about computer science. The author states that “in the computer science field, proceedings are usually a prime avenue for publications, and selected conference proceedings are as prestigious as journal articles” (p. 2069). Conference papers account for more of the published literature in computer science than journal articles. Therefore, the sample used as the data source in this study is not representative of the computer science literature, which means the results cannot be generalized to searching the entire corpus of published literature in computer science.

Additionally, another limitation pointed out by the author is that databases continuously update their list of indexed journals so that “identical searches might give different results if repeated over time” (p. 2069). The results in the study are a snapshot in time, requiring that the study be repeated to confirm the findings.

The author suggests that a future direction for research would involve searching the databases under investigation for the conference papers found in the institutional repository. This reviewer thinks that the future direction should have been done in this study in order to validate the findings. The author could have randomly selected documents from the repository to obtain a more representative sample or searched the titles of all 3,828 documents in the different databases.

Despite its limitations, this study will be of interest to librarians seeking to compare databases in a specific discipline for teaching, reference, or collection development purposes. The study successfully demonstrates that the source of data for a comparative database evaluation can also be taken from an institutional repository that provides references to all the scholarly output of its researchers. Readers should take a

representative sample of documents from the institutional repository to ensure the validity of their results.

Reference

Glynn, L. (2006). A critical appraisal tool for library and information research. *Library Hi Tech*, 24(3), 387-399.
<http://dx.doi.org/10.1108/07378830610692145>