



Will Web Search Engines Replace Bibliographic Databases in the Systematic Identification of Research?



Jessica Bates ^{a,*}, Paul Best ^b, Janice McQuilkin ^c, Brian Taylor ^d

^a School of Education, Ulster University, United Kingdom

^b School of Sociology, Social Policy and Social Work, Queen's University Belfast, United Kingdom

^c Library, Ulster University, United Kingdom

^d School of Sociology and Applied Social Studies, Ulster University, United Kingdom

ARTICLE INFO

Article history:

Received 29 July 2016

Received in revised form 22 November 2016

Accepted 23 November 2016

Available online 27 December 2016

KEYWORDS:

Academic librarianship

Bibliographic databases

Google Scholar

Health research

Literature searching

Social care research

Systematic searching

Web search engines

ABSTRACT

The availability of web search engines offers opportunities in addition to those provided by bibliographic databases for identifying academic literature, but their usefulness for retrieving research is uncertain. A rigorous literature search was undertaken to investigate whether web search engines might replace bibliographic databases, using empirical research in health and social care as a case study. Eight databases and five web search engines were searched between 20 July and 6 August 2015. Sixteen unique studies which compared at least one database with at least one web search engine were examined, as well as drawing lessons from the authors' own search process. Web search engines were limited in that the searcher cannot be certain that the principles of Boolean logic apply and they were more limited than bibliographic databases in their functions, such as exporting abstracts. Recommendations are made for improving the rigour and quality of reporting studies of academic literature searching.

© 2016 Elsevier Inc. All rights reserved.

INTRODUCTION

Academic librarians have the task of guiding higher education staff and students in their search for relevant materials. A particular demand in higher education is the retrieval of empirical research such as is required as part of research grant applications and research degree theses. As the volume of available information grows, the task of retrieving relevant research (and avoiding retrieving irrelevant research) becomes ever more demanding. The availability of the World Wide Web provides possibilities for efficient searching if it can be used effectively. The aim of the study reported here was to investigate whether web search engines have the potential to replace bibliographic databases for use in the systematic identification of research.

Although the lessons of the study were expected to be generalizable across higher education subject areas, to make a manageable project the focus was restricted to searching for empirical research and also by discipline. A search within the disciplines of health and social care was selected because this spanned both health disciplines where searching is

well-established and a social science discipline where there is emerging interest although the methodologies are not well developed.

Although there is a considerable literature base on systematic searching in some disciplines, examination of the implications of utilising web search engines, such as Google Scholar in the systematic identification of research, is a more recent development. Niu and Hemminger (2012) found that the academic scientists in their study were equally split in their preference for starting a search on the library web page or with Google. However, those using Google were less confident that their searches were returning all of the relevant information. The purpose of this study was to undertake a review of research which directly compared searching with web search engines against bibliographic database searching, so as to give a sound knowledge base for use by librarians in advising staff and students in their searching endeavours.

The project involved three distinct stages:

- a systematic search using eight bibliographic databases and five web search engines (including Google Scholar) to retrieve articles that either evaluated or used at least one bibliographic database and one search engine as part of a review or empirical study;
- appraisal of the retrieved relevant documents using a quality appraisal framework; and
- synthesis of findings and conclusions.

* Corresponding author at: School of Education, Ulster University, Coleraine campus, Northern Ireland BT52 1SA, United Kingdom.

E-mail address: j.bates@ulster.ac.uk (J. Bates).

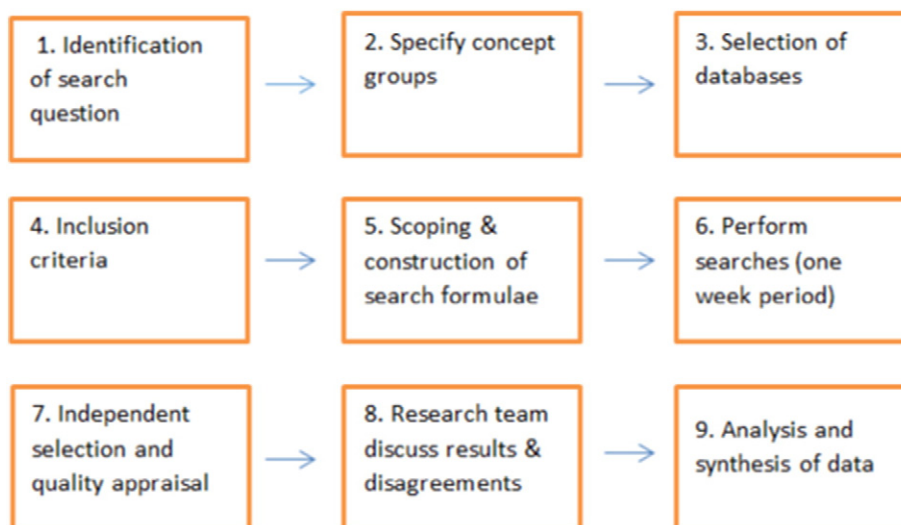


Fig. 1. Nine-step process.
(Adapted from Best et al., 2014).

METHODOLOGY

The systematic review followed the PRISMA guidelines (Moher et al., 2009) for the identification, screening and inclusion of relevant papers, which culminated in the following nine step process (Best, Taylor, Manktelow, & McQuilkin, 2014). See Fig. 1 below.

The search question was defined as *'what is published in peer reviewed articles comparing the performance of bibliographic databases and web search engines for topics relating to health and social care?'* To achieve this, three main concept groups were identified (1) literature searching; (2) bibliographic databases and (3) search engines. These groupings formed the basic structure of searches across all databases (see Appendix A). Health and social care was tested as a potential fourth concept group during piloting but was removed to increase sensitivity or recall and simplicity of search structure.

SELECTION OF DATABASES

The selection of appropriate databases is a key factor in the literature searching process (Lawrence, 2008). Eight bibliographic databases (Fig. 2) were chosen ranging from inter-disciplinary databases, such as Scopus, to those focusing on health (Medline), social care (Social Services Abstracts) and information retrieval/library and information science (Library and Information Science Abstracts). The final selection of bibliographic databases was taken following a number of consultations with a subject librarian at the University and based on approaches established within previous research (Best et al., 2014; McFadden, Taylor, Campbell, & McQuilkin, 2012; Taylor, Wylie, Dempster, & Donnelly, 2007). Other databases considered, but removed following scoping exercises, were Communication Abstracts, PubMed and Social Care Online. Reasons for their omission included limited search sophistication, utility for topic and retrieval of articles predominantly available through other databases.

Search engines were defined as websites which search public areas of the web in response to the terms that the user types in.¹ Five web search engines (Fig. 3) were selected based on usage (traffic) as assessed using the following tools: Alexa, Reliablesoft and eBiz.² Each

of these sites identified the same top four search engines: Google, Yahoo, Bing and Ask. Google Scholar was also included given its availability and use within academia.

INCLUSION AND EXCLUSION CRITERIA

Inclusion criteria were defined as articles published in peer-reviewed journals, available in full text, in English language that evaluated or used bibliographic databases or web search engines to retrieve research. To be included, articles must have utilised and compared at least one or more bibliographic database against one or more web search engine for their literature search and focused primarily on a topic relevant to health and social care. Whether or not an article related to health and social care was assessed independently by two qualified health and social care professionals and was a major factor in bibliographic database selection. Given that Google Scholar was launched in November 2004, it was decided that articles published pre-January 2005 would not be included.

SEARCH STRATEGY

Where a database contained a subject index or thesaurus, it was agreed that the indexed term (if found) would be used in preference to user-generated 'free text' keywords. Where no subject index was available, user generated keywords were to be entered. The searches also employed the relevant truncation and wildcard characters for the selected databases, increasing the number of available search terms.

Most databases offered a 'sort by relevance' function. However, database suppliers rarely describe how 'relevance' is determined. Measures may include how frequently the term is present in the results, whether the term appears in a subject index, abstract, article title or author-supplied keywords or a number of other algorithms.

Despite these databases using automatic relevance sorting as standard, it was evident that some of the results appearing high up in the relevance sort, were there based on partial keywords from within a multi-word concept instead of as precise phrases. To overcome this, it was decided to include verbatim 'phrase-searching' on key terms, where necessary. In addition, the search strategy used more advanced proximity connectors to more precisely control the relationship between the key terms.

Medline and PsycInfo did not offer an option to sort results by relevance. The OVID sensitivity and precision filters were not used, being primarily for clinical queries.

¹ A useful comparison of library databases and web search engines is provided by the University Library, University of Illinois at Urbana-Champaign (<http://www.library.illinois.edu/ugl/howdoi/compare1.html>, last accessed 31 October 2016).

² Alexa (<http://www.alexa.com/topsites>), Reliablesoft (<https://www.reliablesoft.net/top-10-search-engines-in-the-world/>), eBiz (<http://www.ebizmba.com/articles/search-engines>).

Download English Version:

<https://daneshyari.com/en/article/4939015>

Download Persian Version:

<https://daneshyari.com/article/4939015>

[Daneshyari.com](https://daneshyari.com)