An Evaluation of the Usefulness of Multi-Search Engines in Scholarly Communication

Mirosław Górny, John Catlow and Rafał Lewandowski
Department of Information Systems, Institute of Linguistics, Adam Mickiewicz University, Poznań, Poland

Abstract: Academics often use publications from a wide range of sources. Multi-search engines enable simultaneous online searching of these distributed resources, which is claimed to make working easier for users interested in publications on a given topic. Because some multi-search engines are sold as commercial software, often at considerable cost, it may be asked whether the use of such software in an academic environment is beneficial and economically justified. Does the academic community in fact perceive a need to use these programs? Is their function not performed equally well by Google Scholar? Is it not more effective to make exclusive use of specialist databases? These questions are investigated by comparing the effectiveness of searching using Google Scholar and the Primo search engine.

1. Introduction

The sets of published resources used by scholars today may be described as "distributed". Publications that are thematically related may appear as a part of larger sets of literature representing quite diverse disciplines. This concerns particularly journals and conference reports, but many works are also published in book form, containing sets of articles on related topics.

In effect, the information user normally has to deal with some set of articles of interest, distributed among different publications and appearing at different times. If we add to this blogs and various types of repositories, we obtain some kind of simplified picture of the phenomenon.

Multi-search engines allow the user to search these distributed resources using a single query, identical for all of the resources. Naturally many queries may be

_____

submitted, but each of them is executed from one place against all of the resources being searched.

Because some multi-search engines are sold as commercial software (often at considerable cost), certain questions naturally arise, for example: Is it in fact profitable to use software of this type in an academic environment? Is the degree of distribution great enough to make the purchase of multi-search engines justified? Do academics in fact perceive a need for such programs? Is the role of multi-search engines not fulfilled equally well by Google Scholar? Is it not cheaper and more effective to make exclusive use of specialist literature databases?

Another argument advanced for the use of multi-search engines is that they are claimed to offer highly efficient searching. Since the factors important to the user are the way in which queries are formulated and the search results obtained, we shall not consider here the ways in which the search algorithms work. We shall limit ourselves to citing a short description of the Primomulti-search engine, as a typical representative of this category of programs:

"Primo analyzes a user's query and optimizes it so that the system can retrieve all relevant results, regardless of whether the user is searching for a known item or seeking items related to a certain topic. To optimize the query, Primo applies a number of techniques, such as the following: Identification of variations in terms and correction of spelling mistakes; Expansion of searches based on word stems; Expansion of search to full text if the original search yielded only few results; Recognition of citation formats (for example, APA, MLA, or Chicago) and year of publication; Recognition of compound words" (Exlibris, 2015a).

Moreover, "The Primo ScholarRank technology sorts the search results by relevance on the basis of several criteria: The degree to which an item matches the query; A value score representing an item's academic significance; An item's relevance to the type of search; The publication date (recentness) of an item" (Exlibris, 2015b).

Firms produce and sell such software, motivated primarily by profit. Academic libraries in Poland are generally willing to buy such programs, as they wish to prove that they are making information-related activity easier, thus creating the impression that they are meeting readers' needs and remain important to readers. There are therefore two players in the information product market that have an interest in the existence of multi-search engines. But is it similarly in the interests of readers? Do multi-search engines in fact provide such benefits to information users that it is worth incurring the substantial costs involved?

When users are asked about the usefulness of a multi-search engine available in a library, as a rule they answer in the affirmative. However, this is chiefly because it is not they who incur the cost of its purchase. It is therefore necessary

to apply a more objective method in evaluating the usefulness of such engines. The problem is important primarily in terms of costs. Polish academic libraries are currently spending hundreds of thousands of euro on the purchase of multi-search engines; but nobody is attempting to investigate whether this expenditure is justified.

Naturally, neither the manufacturers nor distributors of this software, nor the libraries themselves, have an interest in obtaining an objective answer to this question (unless the answer proves to be that the purchase is entirely justified). In turn, information users, who do not directly incur any of the costs, are generally indifferent to the problem.

Evaluating the usefulness of multi-search engines in these circumstances is a challenging task. Even establishing the criteria for evaluation gives rise to serious difficulties.

2. Formulation of the Problem

A multi-search engine is designed to indicate all sources of information on a given topic that are available within a given information infrastructure. It is expected that, instead of the need to send a separate query to each information source, it will be enough to create a single query to be executed against all of the sources. Manufacturers also advertise multi-search engines as capable of ordering search results based on the importance of the returned publications.

Wishing to consider this task more precisely, we may formulate the following hypotheses:

Hypothesis 1.A user will wish to use a multi-search engine chiefly when seeking publications on a given topic (different levels of detail may be specified – topic, publication date, language, author affiliation, etc.).

Hypothesis 2. It would appear that the overriding expectation of the information user will be the obtaining of knowledge about *all*existing publications on the topic in question, not merely a set of sources available locally. This second expectation arises later, when the user has made a preliminary assessment of the returned set of information. It is unlikely that a researcher will attach more importance to information *only on publications available to him or her*than to information on *all or almost all publications*– unless the researcher merely requires *any publications at all*–to produce a sufficiently extensive literature review or bibliography for a publication under preparation. It is true that the number of publications on a topic may turn out to be so large that many of them are not considered in subsequent work. However, at the stage of choosing publications of importance, the researcher will always expect a relatively complete set. At this stage, the availability of the publications is a secondary matter. If it transpires that there exists a publication which is of high importance

for the topic, but which is not available locally, then the reader will certainly take action to obtain it.

Hypothesis 3. Thematic searches can be assumed to be performed relatively rarely – usually only a few times a year. This means that it is difficult to speak of significant time savings. The time spent on performing such searches is a secondary factor compared with the value of the results.

Hypothesis 4. Thematic searching would appear to be an exceptionally important matter in academic work. It is the results of this searching that determine the information user's subsequent direction of work. It may therefore be assumed that the user is willing to spend a significant amount of time on achieving a valuable result. Then he or she will decide to use specialist tools, having greater confidence in them than in tools for general use – particularly since the latter are merely aggregators, and their completeness depends to a large extent on the efficiency of the process of aggregating information. We nonetheless know from experience that in at least 70% of cases users still begin with Google and Google Scholar. Factors behind this include ease of operation; the relative completeness of the results, covering not merely local but global resources; and the fact that these search engines are available free and without restriction.

3. Methodology

It would appear that the only method enabling evaluation of the usefulness of multi-search engines is testing carried out in a specific information environment. It is assumed that the most commonly used multi-search engine among academics is Google Scholar. Hence the results obtained by the use of that tool were compared with those obtained with the Primo multi-search engine. In some cases, a comparison was also made with the results of searching in a library catalogue and in the catalogue of the Digital Library of Wielkopolska.

The tests used randomly selected queries formulated by undergraduate and post-graduate students. They included both simple and compound queries: on terms of local significance, such as names of towns, or "vocational schools in Poznań in the interwar years"; and on broader terms, such as "inheritance law" and "the problem of anaphor in translation".

A score was then assigned to the results. First, the person performing the test assessed the value of each publication, and a score of 1, 2 or 3points was assigned depending on whether the value was assessed as low, medium or high. A publication deemed irrelevant to the topic scored minus 2 points. A publication with digital access was awarded a bonus of +3 points. In the case of Google, a publication that also appeared in the local multi-search engine was awarded +2 points. A publication available in local resources received +1 point. Further, 1, 2 or 3 points were awarded for the in formativeness of the description.

Query	Scholar	Primo
*Bnin (*name of a locality in Wielkopolska)	18	32
Kościan (name of a town in Wielkopolska)	38	59
Chwaliszewo (name of a part of Poznań)	76	25
gwara poznańska(Poznań dialect)	93	3
historia Islandii(history of Iceland)	81	(-4)
religijność Polaków(religiousness of Poles)	82	92
chemia koloidów(colloid chemistry)	28	6
żubr w Polsce(bison in Poland)	50	3
antropologia kulturowa(cultural anthropology)	39	51
kobieta na wsi(woman in the countryside)	36	9
Total score	541	276

Table 1. Scores for search results returned by Google Scholar and Primo (simple search terms).

Query (in translation)	Scholar	Primo
vocational schools in Poznań in the interwar years	37	0
the Wielkopolska Uprising in Inowrocław county	(-2)	0
assemblies of nobles in Środa Wielkopolska	43	0
Wielkopolska trade in the 1930s	5	8
diseases of children of pre-school age	(-12)	7

French cuisine in the Middle Ages	(-4)	0
population of wild geese in Wielkopolska	4	0
foreign language teaching in high schools	(-1)	3
manor farming in Poland in the 18th century	28	5
affective bipolar disorder	156	88
the problem of anaphor in translation	21	0
Szubin in the Wielkopolska Uprising	26	10
Total score	301	121

Table 2. Scores for search results returned by Google Scholar and Primo (compound search terms).

4. Conclusions

Theoretically, a multi-search engine working with resources that are exceptionally well catalogued ought to perform its function successfully – that is, locate all items relevant to the topic in question that are available locally. In practice, this is generally impossible. This is a result of certain limitations linked to the creation of metadata.

Multi-search engines generally do not ensure a high level of completeness of search results on a given topic. Sometimes they do not even return a complete set of results among local resources. It is hard to establish what proportion of the resources they overlook, as this is dependent on many factors.

It may be assumed that the process of searching for information often conforms to a three-level model. Here, the first search operation using a multi-search engine should return descriptions of all of the sources of information of potential interest to the user. That set will then be analysed in detail to select the most valuable items. When these items have been selected, the user will attempt to determine whether they are available locally, on the basis of local catalogues.

Even the best multi-search engine relies on an indexing tool. (In local multi-search engines it is necessary for catalogue data to be supplied systematically by librarians, while global engines depend on the effectiveness of indexing robots and the availability of local sets of metadata.) Moreover, a multi-search engine uses only those metadata that are available on some level. Hence there can never be certainty that all potentially significant publications have been taken into consideration.

What will the information user do in such a situation, assuming that he or she is aware of the aforementioned limitations? The solution is to work with specialist tools. Naturally, there is always the temptation to check in a fast and easy

manner what has been published on a given topic. Serious researchers, however, tend not to give in to this temptation.

As a rule, researchers may find themselves in one of four information situations:1) They want to be up-to-date with the literature relating to their current topic of interest; 2) They are beginning to work in a new subject area and are carrying out a retrospective search; 3) They wish to find a specific item; 4) They are seeking inspiration and searching literature related more or less directly to the topic of interest.

In principle, only in the second and fourth cases is the use of a multi-search engine justified; and then preferably not a local engine, but a global one. This does not mean, of course, that the multi-search engine will be the researcher's only tool. It will perhaps make it possible to gain a preliminary overview of the scope and content of the existing literature on a topic. If the researcher decides to continue to seek information, then he or she will no doubt move towards specialist tools. The number of such tools is usually limited to a few databases or a few dozen journal titles. Thus the effort involved in browsing them is not especially great, particularly if done systematically.

The great majority of information users would no doubt wish to obtain a set of descriptions of sources of information on a topic by submitting a single query (we are not considering here those users who simply like to look through different information sources thoroughly). Nonetheless, the following conditions must be fulfilled:

1) the operation should be relatively easy to perform;
2) the operation should take very little time;
3) the results should be complete or offer a very high level of completeness;
4) the results should contain descriptions of the most important publications;
5) the results should allow the user to evaluate the value of the publication.

The results of comparative tests for the Primo multi-search engine and Google Scholar in the environment of a university library in Poznań did not give any indication that these conditions were satisfied. They also do not establish definitively which of the multi-search engines makes it possible to obtain better results, although the scores appeared clearly to favour Google Scholar. It seems, however, that the test sample was too small. The conditions of the test provided for the submission of a single query, and took no account of the gaining of experience by the user. It is true that the results of other studies also indicate the superiority of Google Scholar: for example, in a study of known-item searching, Primo scored just over 60 points, while Scholar scored over 90 (Singley, 2014).

In another study,Scholar and Primo achieved almost identical results, but with a slight advantage to Scholar (Zhang, 2013). In a very well-designed study, there was found to be no significant difference between Google Scholar and Summon or EDS in tests of known-item retrieval, but Scholar proved to be somewhat superior in topical search(Ciccone and Vickery, 2015).

Given this knowledge, it is recommended that readers should primarily make use of specialist narrow-topic databases. The use of multi-search engines may be considered only as a complementary activity.

On economic grounds, however, it would appear that the purchase of commercial multi-search engines by libraries is probably not a profitable investment. Before purchasing such software, each library should carry out tests involving a set of typical queries submitted by information users.

References

Ciccone, K. and Vickery, J. (2015). Summon, EBSCO Discovery Service, and Google Scholar: A Comparison of Search Performance Using User Queries. *Evidence Based Library and Information Practice*, 10(1), 34–49. https://doi.org/10.18438/B86G6Q

Exlibris Ltd. (2015a). Primo Discovery:Search, Ranking, andBeyond. March 2015, 1–2.

Exlibris Ltd. (2015b). Primo Discovery:Search, Ranking, andBeyond. March 2015, 3–4.

Singley, E. (2014). Discovery systems – testing known item searching. *Usable Libraries blog*, March 18, 2014. http://emilysingley.net/usablelibraries/discovery-systems-testing-known-item-searching/ (accessed 20 April 2018).

Zhang, T. (2013). User-centered evaluation of a discovery layer system with Google Scholar. In:A. Marcus (ed.), *Design, User Experience, and Usability. Web, Mobile, and Product   Design*   (vol.   8015,   313–322),   Springer,   Berlin. https://www.researchgate.net/publication/262155062_User-Centered_Evaluation_of_a_Discovery_Layer_System_with_Google_Scholar   (accessed 20 April 2018).