

The crossroads of academic electronic availability: how well does Google Scholar measure up against a university-based metadata system in 2014?

Niklas Karlsson

Electronic availability of information resources has increasingly become an important part of everyday vocation of academic libraries. This puts impetus on the libraries to know more about the way in which electronic information is being dispersed and handled. The present article aims to comparatively evaluate Uppsala University library's own metadata system Summon with the free, publicly available equivalent Google Scholar (GS). The evaluation is based on Péter Jacsó's theories on database evaluation which puts focus on Summon and GS via the use and application of ten different criteria. The uses of precision and relevance criteria were also implemented as additional evaluation tools. The results indicate that at present GS has to be seen as a necessary complement in retrieving electronic information due to the fact that Summon is not yet fully functioning on all levels and that GS has a wider intake of information sources. The use of web-based academic search tools is now vital. Will the open access movement evolve with Google as the main actor and take over the scene leaving costly databases and search tools behind? This article deals with the economic implications of comparing the practical functions of a costly in-house information system with a public equivalent. It reveals the complex situation that a world-class university is in as regards to information resources and the digitization and economic issues that follow.

Keywords: Academic libraries, databases, electronic information, Google Scholar, metadata system, web search engine.

In 2013, the Swedish Uppsala University library acquired electronic information resources at the cost of 33.7 million SEK representing 73% of the total acquisition of information resources (Uppsala University e-resources team, 2 February 2014). The corresponding percentage in 2002 stood at 32 and it is clear that electronic resources are about to take over transcending into a digital future. Uppsala University's acquisition policy mentions 'the quality, uniformity and breadth of the collections shall be guaranteed with a long-term perspective bearing in mind possible future areas of research and education¹.' This is where my research begins; the quality of the acquired and available material, and more precisely, the quality of the availability of the acquired material. The more material that is made available electronically, all the more important it becomes to refine and focus on the channels that are used to grant access to the material. Access to electronic resources is not dependent on competent librarians, but on well-functioning web-based discovery tools.

The author is in the Uppsala University, Carolina Rediviva, Dag Hammarskjölds väg 1 Postadress: Box 510, 751 Uppsala, Sweden. e-mail: 018365287@telia.com

In 2000, Allison *et al.*¹ discussed the problem faced by libraries with the altered media landscape with changing routines for information access. How should libraries respond to changing media resources and what strategies are appropriate to balance the issues of cost, access and local situation of a library? It is clear that there is a need for assessment tools in order to provide conditions for selection in an increasingly complex and growing range of information resources. There are many roads to information in today's technology and it is important to be able to identify them, both as a librarian and as a student. It is high time we put availability itself in focus, exposing it to an 'objective assessment'.

Since its launch in 2004, library and information science has emerged as a distinct research field that critically discusses, compares and evaluates the academic, freely available web-based search tool Google Scholar (GS). Uppsala University library uses GS in several capacities as a complementary tool via its library's website and GS has established itself as a well-used academic tool. A study in 2008 showed that GS has had good impact in the academic world, in that majority of research libraries in

the United States had links to it on their websites³. Another study in 2011 indicated that majority of the surveyed students in an American university used and had a positive impression about GS and the way it makes available documents perceived as useful and easy. The study notes how students increasingly are turning to GS and it is high time for university libraries to adapt to that reality³. Another study^{4,5} examined the usage statistics for GS in an American university which indicated that as early as 2006, GS was a well-used academic tool whose use had increased tenfold in 2011. In fact, GS was used more than the University's own meta-search engine⁵.

Howland *et al.*⁶ noted in their study that the ultimate evaluative comparison of GS would be with a completely locally indexed information system available within an organization, but produced by a third party. Uppsala University library today makes available such an information system, Summon, and it would be interesting to compare its abilities with those of GS.

The aim of this article is to examine and evaluate the quality of availability of information resources and the retrieval efficiency of GS compared to Summon. Quality of availability is a measurement on how well GS and Summon make available their information resources. Efficiency in this context is defined by precision and relevance in terms of both documents and citations. The aim is thus simply to evaluate and determine how good a free alternative like GS is today. The questions posed are: (1) How well does GS retrieve information resources compared to Summon? (2) How efficient are the search tools in retrieving relevant information resources? (3) What are their strengths and weaknesses respectively?

Related work

There have been several evaluations of GS since its conception in 2004 and I have them in chronological order for displaying the development of GS. Jacsó⁷ completed a comparative study of three multidisciplinary bibliographic databases in 2005; GS, Web of Science (WoS) and Scopus. He pointed out the shortcomings in terms of accuracy and citations and identified GS as 'often an extremely distant third' in comparison with WoS and Scopus. Andersson and Pilbrant⁸ comparatively evaluated GS and a similar search tool Scirus (Elsevier). Based on the assessment tools which measured relevance and academic content they found that GS retrieved a higher number of academic papers, while Scirus retrieved material with greater relevance to the topic. Overall, the differences were small and the authors felt that both the search services complement each other. Walters⁹ compared GS with seven subscription-based databases in 2006 and conducted searches on 155 articles on the topic 'migration in the latter part of the life' published between 1990 and 2000. The results indicated that GS retrieved the highest percentage of the searched material at 93,

which was 27% better than the best subscription-based database. However, the author pointed out that GS was not a serious alternative to the specified subscription databases because of its unsophisticated retrieval and document handling functions. Mayr and Walter¹⁰ evaluated GS retrieval precision in 2006 by selecting titles from five different journal lists which formed a wide academic spectrum with a total of 9500 searchable titles. The authors identified GS as an interesting alternative with its citation function and its freely available results, but that it was still a far worse alternative to specialty databases and library catalogues. Inconsistency, irrelevant material and that GS limits each search result to 100 documents were major disadvantages that were uncovered in the searches. Shultz¹¹ compared retrieving precision of GS and PubMed in 2007 by first implementing ten different searches on the subject, author, title, journal, as well as those in combination in the PubMed database via different search functions. The author then mirrored the searches in GS to simulate the user's search paths and search options. The results indicated that GS gave a higher number in the search results than PubMed in 8 out of 10 searches. Shultz reported that it is important for librarians to make it clear to users regarding the strengths and weaknesses of GS. Neuhaus *et al.*³ studied the presence of GS academic contexts by examining 948 US-based universities and their library websites. The results indicated that the universities in the highest academic category exhibited the largest degree with 73% presence of links to GS. Other categories indicated a declining trend with 33% attendance at the second highest academic level and a significantly lower share in the two lowest levels respectively. Howland *et al.*⁶ examined how GS compared qualitatively with traditional library resources. The authors asked seven subject librarians to design typical search queries from a student perspective across a wide range of subjects covering arts to science. The results indicated that GS had clearly higher academic quality and precision in the retrieval of citations over the entire subject spectrum than the databases that the libraries use in their everyday work. The authors see the outcome as a validation of the students' ever-faithful use of GS as a first source of information resource. GS showed better values than the subscription databases both in terms of accuracy and coverage. Also, GS strives to continue to develop features and capabilities. In 2005, Chen¹² followed up on the study of Neuhaus *et al.*³ on GS retrieval precision to examine how it has evolved and improved over the period of five years. With GS, the author sees potential for a shift of individual libraries from putting a large part of the budget on information resources to instead use free resources.

Research methods

As we can see above there are several ways of evaluating databases and search tools. My objective was not only to

measure efficiency via the use of precision and relevance criteria (constituting the ‘front-end’ of evaluation) as so many have done before me, but also utilize a wider ‘back-end’ evaluation basis. For this I employed the theories of one of the pillars in database evaluation, Péter Jacsó¹³, who outlines ten criteria for evaluating databases. This methodology helps identify and evaluate subject scope, range, accuracy, consistency and completeness (amongst other). The aim is to ultimately reach a result that describes the quality of the availability of the contents of the search tools that are assessed.

As for the ‘front-end’ the different relevance criteria are defined and used as a way to mimic the user, which puts basic functionality in focus when assessed¹⁴. Are the documents really full-text ones made available in the relevant language (the language of the search query)? Or are they merely reviews of documents, empty links or just old news bulletins? The precision measure quantifies these aspects and questions into hard data which in the end give a comparative result that represents the practical functionality of GS and Summon¹⁵.

The search queries were based on four types of selection groups constituting four different comparative evaluations.

(1) The first evaluation was based on search statistics representing the most frequent search queries typed into Summon during its first year in use at Uppsala University. This was done in order to measure how the databases answer up to real-life search queries.

(2) The second evaluation was based on Thomson Reuters ISI Web of Knowledge *Journal Citation Reports*. From ten different scientific categories, the top three journals were then searched in order to establish if and/or in what capacity they were present in both databases.

(3) The third evaluation was based on Thomson Reuters Research fronts 2013, which reports the most talked about and researched scientific questions in ten different scientific fields. This was done in order to measure how well the databases answer up to contemporary science queries; how up-to-date they are in reality.

(4) The fourth evaluation was designed to evaluate the bibliographic record content. For this, three different sources of top cited articles were used as well as Oxford Dictionary descriptions of the most commonly misspelled words.

To be able to carry out all the parts of the evaluation in a truly scientific manner, some serious methodological issues had to be dealt with and some ground rules established. The bias issue is always present, especially when undertaking a comparative evaluation using search queries based on different types of source. Different types of validity, completeness and strictness as well as methods for avoiding comparative bias have been employed and discussed. These include the so called ex-ante, post-ante and ex-post strategies that deal with comparative bias aspects in different stages of research process¹⁶.

(1) Standardization pertains to the definition of terms and proposals that are to constitute a common ground throughout the evaluation.

(2) Adaptation pertains to terms, search queries, selection groups, methods and how they are adapted to fit the specific evaluation and form a common framework for the whole research process.

(3) Correction pertains to the correcting and dismissal of non-relevant and non-usable parts in the final stages of the research process. This is a common method when scientists do not have enough opportunities to plan the scope and execution of the research process.

Of course, these are all methods in favour of being transparent and the common factor in employing these methods is simply to minimize bias and promote and maximize transparency. In this case a relevant prestudy would help to pinpoint the information needs of different groups within Uppsala University. Questions that would have been relevant to ask include whether there are different tendencies (larger/smaller) in different groups towards employing the services of a certain search engine. Groups would be recognized by their scientific field association as well as their position, i.e. students or researchers. Another obvious advantage in doing this kind of prestudy is to work out relevance criteria by approaching the said groups either with a quantitative approach, i.e. forms or a qualitative approach, i.e. interviews. This would enable the measure of information retrieval efficiency in both databases, aimed at specific information needs.

Notable results

The ‘front-end’ results indicate that GS is less efficient in terms of retrieving documents based on proven scientific search queries (Figure 1 *a* and *c*). Though while querying for the top journals in ten different fields (Figure 1 *b*), GS displays its strengths as it retrieves documents from all the selected journals compared to Summon which only retrieved documents from 14 out of 30 journals (Table 1). Another clear advantage for GS that becomes obvious during the test searches is that it gives working links to cited articles, while Summon does not. Also, 51% of the cited articles are available in full-text in GS. As for the ‘back-end’ evaluation, problems arise when the databases are queried on misspelled words. GS retrieves four times more documents than Summon when querying a misspelled word in relation to how many documents are being retrieved when querying a correctly spelled word. This means that the database does not correct either the spelling or the meaning of the misspelled word, which results in the fact that certain keywords to certain documents are only searchable when they are misspelled! As for Summon, the administrative team has not yet perfected the system in practice.

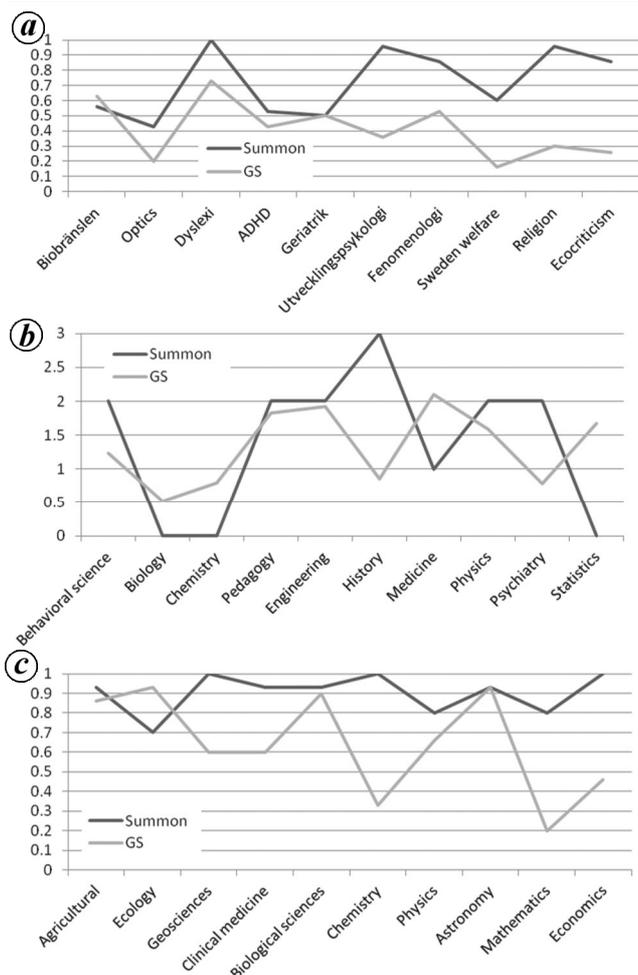


Figure 1. Precision Google Scholar (GS) and Summon test search 1 (a), test search 2 (b) and test search 3 (c).

Table 1. Precision of Google Scholar and Summon

	Precision percentage test		
	Search 1	Search 2	Search 3
Summon	73	47	90
Google Scholar	41	44	65

An example of this is a query for the misspelled word ‘accommodate’ and finds the article ‘Can phenomenology accomodate Marxism?’. However, when one queries for this article with the correct spelling ‘accommodate’, both databases do not retrieve it. There are similar problems with the way the databases index author names, as neither of them seems to employ a system that regulates author names according to a set standard. For example ‘Kobayashi, M’ could easily be a number of people (Masaki, Masato, Mahito). Why not always present the author as ‘Kobayashi, Masaki’, as would be the correct procedure in this example? These kinds of variation in presenting author names can be confusing and with faulty spelling there is further potential for confusion. When querying

for the correctly spelled ‘Péter Jacsó’ and the most common, according to Jacsó himself, misspelling of his name ‘Peter Jasco’, the misspelled version of the name retrieved 2719 documents in Summon and the correctly spelled name retrieved only 1085 documents (1634 documents less). GS was far less confusing, as the misspelled version retrieved only 22 documents and the correct one 128 documents.

Discussion

The significance of these results and this study as a whole is attained in reviewing both databases side by side. How do they perform against each other, and are there any practical reasons for using GS at all? Even though GS retrieves less full-text documents and thus functions with less efficiency, it is still a notable alternative in practice due to its wide intake of journals and documents as well as the simple fact that costly metadata systems do not yet function up to their full potential. The question is whether a system like Summon can attain a loyal following in competition with GS even if it optimizes its functions?

During the study Summon did not display working links to cited articles, which means that there is potential for a two-pronged slide effect when students and researchers need an alternative to find working links to cited articles as well as finding documents from journals not made available by the university. The benefits of using not only a ‘front-end’ evaluation, but also a theoretical ‘back-end’ evaluation are clear when assessing all the results. Structured ideas on how a database should function help evaluate new information that needs to be retrieved by students and researchers. Libraries/librarians need these tools today to handle the large amount of scientific information and evaluate the ways the acquired material is being made available. Alternatives like GS give credence to evaluation efforts in order to assess how to handle the increasingly expensive electronic resources. The information attained through evaluations like in the present study has the potential to creating strategic plans for the future. Electronic availability in the academic field stands is now at a crossroads, and there is an apparent need for a new solution that fits universities, publishers and researchers alike.

We need to qualitatively study relevant user groups and analyse their position to see for instance if there are any differences between different scientific topics. Another interesting study would be on the existence of bias in the working relationship between publishers of scientific material and metadata systems like Summon and GS. Are the metadata systems prone to retrieving certain sorts of documents more than others?

The future

The real significance of this study lies in how a university should spend its money. Should it continue to invest

millions in making available information resources when free alternatives are consolidating their strength? In November 2013, Thomson Reuters announced that they were entering into a cooperative effort with GS and maybe this could be a sign of things to come¹⁷? New technological solutions are being presented at a rapid pace and Google's dominance of the web is increasing. How can scientific publishing adapt and evolve in a way agreeable to all parties¹⁸. GS has a kind of on-demand function today already where free documents are instantly attainable and those which are not freely available are posted without a link to an actual document. This could help universities cut down their costs for information resources. The idea is to use GS for its free material and make a contractual agreement for direct access to documents not freely available. Google would be the main actor in this scenario, making available open-access documents as it does now and developing its intake of prized scientific material by collaborating with the main scientific information actors such as Thomson Reuters and Elsevier. The documents which come with a price tag could be retrieved via student portal type access, which means that the university only pays for the actual documents that are being downloaded.

1. <http://www.ub.uu.se/lana-och-ladda-ned/forvarvspolicy/?language-id=1>
2. Allison, DeeAnn, McNeil, B. and Swanson, S., Database selection: one size does not fit all. *Coll. Res. Lib.*, 2000, **61**(1), 56–63.
3. Neuhaus, C., Neuhaus, E. and Asher, A., Google Scholar goes to school: the presence of Google Scholar on college and university websites. *J. Acad. Lib.*, 2008, **34**(1), 39–51.
4. Cothran, T., Google Scholar acceptance and use among graduate students: a quantitative study. *Lib. Infor. Sci. Res.*, 2011, **33**(4), 293–301.
5. Wang, Y. and Howard, P., Google Scholar usage: an academic library's experience. *J. Web Lib.*, 2012, **6**(2), 94–108.
6. Howland, J. L., Wright, T. C., Boughan, R. A. and Roberts, B. C., How scholarly is Google Scholar? a comparison to library databases. *Coll. Res. Lib.*, 2009, **70**(3), 227–234.
7. Jacso, P., As we may search – comparison of major features of the Web of Science, Scopus, and Google Scholar citation-based and citation-enhanced databases. *Curr. Sci.*, 2005, **89**(9), 1537.
8. Andersson, C. and Pilbrant, M., Google Scholar eller Scirus för vetenskapligt material på webben? En utvärdering och jämförelse av återvinningseffektivitet. Master's thesis, Högskolan i Borås, 2005, p. 61.
9. Walters, W. H., 'Google Scholar coverage of a multidisciplinary field. *Inf. Proc. Manage.* 2007, **43**(4), 1121–1132.
10. Mayr, P. and Walter, A.-K., An exploratory study of Google Scholar. *Online Inf. Rev.*, 2007, **31**(6), 814–830.
11. Shultz, M., Comparing test searches in PubMed and Google Scholar. *J. Med. Lib. Assoc.: JMLA*, 2007, **95**(4), 442–445.
12. Chen, X., 'Google Scholar's dramatic coverage improvement five years after debut. *Ser. Rev.*, 2010, **36**(4), 221–226.
13. Jacsó, P., *Content Evaluation of Textual CD-ROM and Web Databases*. Database searching series, 99-0949100-1. Libraries Unlimited, Englewood, Colo, 2001.
14. Landoni, M. and Bell, S., Information retrieval techniques for evaluating search engines: a critical overview. *Aslib Proc.*, 2000, **52**(3), 124–129.
15. Carterette, B., Kanoulas, E. and Yilmaz, E., Chapter 5 evaluating web retrieval effectiveness. *Lib. Inf. Sci.*, 2012, **4**, 105–137.
16. Denk, T., *Komparativa analysmetoder*. 1. uppl. Lund: Studentlitteratur, 2012, pp. 22–65.
17. Day, J., Thomson Reuters pulling Web of Science from Discovery services. *Library Technology Launchpad*, 2013; <http://libtechlaunchpad.com/2013/11/21/thomson-reuters-pulling-web-of-science-from-discovery-services>
18. Thompson, A., Can Google be unseated from its search dominance throne?, 2013; <http://www.techwyse.com/blog/online-innovation/can-google-be-unseated-from-its-search-dominance-throne/>

ACKNOWLEDGEMENT. This article is an adaptation of a two years Master's thesis written without funding for Uppsala University ALM faculty.

Received 16 July 2014; accepted 21 July 2014