

CHANGING DYNAMICS OF ACQUIRING KNOWLEDGE: A CRITICAL APPRAISAL OF WEB SEARCH ENGINE AND LIBRARY RESOURCES.

Disha Arti

Researcher, DLIS
University of Jammu, (U.T), India

Shivangi Gupta

Researcher, DLIS
University of Jammu, (U.T), India

Abstract

The present study deals with the comparison of an internet search engine with appropriate library databases and systems. A case study method considering various tasks on general subject queries, to be resolved through both internet and library databases are comparatively analyzed in order to assess the relative value, strengths and weaknesses of the two sorts of systems. Although they seem to be the parts of same coin but both demand a different approach to access knowledge related anything.

Introduction

The internet, and specifically the World Wide Web, has transformed the information era in the past decade, providing more rapid access to a greater volume of material than possible at any earlier time. Searching tools, though far from perfect, have played a major part in this transformation. One of these tools, the Google search engine, has become predominant, to the extent that “to Google” had become defacto a verb in the English language by mid-2003, despite the objections of the company (Quint, 2002; BBC, 2003).

Google is, therefore, representative of the variety of easy-to-use search engines, based on free-text searching of the content of public web pages. It is indeed their major representative, given the mission of the company “to make the entire world’s information available” (Library Journal News, 2003). The extension of the “basic” Google search function into Google scholar (providing access to non-copyright academic material.(Tenopir, 2005)), Google print (searching the digitized full text of printed books, from publishers, booksellers or libraries, and allowing the viewing of a small extract of copyright material) (Fialkoff, 2005), and other ventures, suggests that this may not be a wild ambition. While these engines have indisputably made much information searching quicker and more efficient, they have also led to the belief that all information is to be found there, and retrieved without undue effort: “library patrons expect to find it all in cyberspace ...for the purposes of academic research, such expectations are unrealistic and even dangerous” (Lawrence and Miller, 2000, p. 1). In turn, this leads to a dismissal of any other sources of information, specifically of libraries and the formal information sources which they provide: In less than a decade, Internet search engines have completely changed how people gather information. No longer must we run to a library to look up something; rather we can pull up documents with just a few clicks on a keyboard. Now ... “Googling” has become synonymous with doing research (Mostafa, 2005). Web search engines, and Google in particular, have created a generation of searchers who are choosing the simplicity of search engines on the open free web over the perceived complexity of library services. Libraries can no longer cater for “people who want fast, easy access to unlimited, full-text content using interfaces that require no critical thought or evaluation” (Bell, 2004). Fast and Campbell (2004) found that students “admired the organization of (an) OPAC, but preferred to use the web in spite of its disorganized state” It is, of course, inevitable that convenient access to information which, while it may not be comprehensive or of the highest quality, is good enough will be alluring. This is a natural human impulse, codified by Zipf into his principle of least effort(Sole and Cancho, 2003) and by Simon in his concept of “satisficing” (Tennant, 2001;Agusto, 2002), not to mention the complaint of some in the library/information area that we live in a society “fuelled by a culture of instant gratification” (Stoffle et al.,1996, p. 219).According to Bell (2004), “Google has become the symbol of competition to the academic library”. He uses the term “infobesity” to compare the way students now search for information with the modern consumption of fast food. Originally coined by James Morris, the Dean of the School of Computer Science at Carnegie Mellon University,

“infobesity” refers to the belief that searching Google for information provides a junk information diet. Bell believes that students often want to find something quickly and that they are generally not concerned about the quality. It is clear from a review of the numerous articles published on this subject, that there is a general belief in the library community that more “nutritious” information can be retrieved by using the specialized databases available in an academic library. There are debates as to the amount of information available through systems such as Google compared to the “hidden web” of library databases (Tenopir, 2004; Herring, 2001; Devine and Egger-Sider, 2004), as well as of concerns about the quality of material retrieved (perhaps uncritically) from a search engine (Herring, 2001; Tennant, 2001).

Purpose and competence of the study

The purpose of the present study is to compare the performance of Google to that of library database services in answering questions of the sort likely to be asked by students. While answering the question, it was hoped to gain an understanding of the optimal performance of both types of service, and of their relative advantages and disadvantages. By “library database services” I mean online catalogues and bibliographic databases, both general and specialized. Although there have been numerous comparative evaluations of databases, library systems, and web search engines (for reviews of these studies see Brophy, 2004; Xie, 2004) relatively few have attempted this sort of direct comparison of two kinds of service. One example is that of Xie (2004), who compared online database systems (Dialog and Factiva) with three different types of web search tool (search engine, directory and a meta-search engine). Students were asked to search the same two topics on each system and then required to give relevance scores, with precision of each system calculated on the total retrieved relevant documents (to a maximum of 20). Another is that of Fast and Campbell (2004), who examined the perceptions of students searching Google and a university library OPAC, using interviews, verbal reports and observations. A third is that of Griffiths and Brophy (2005), who report two studies of the use of Google and of various academic information resources, finding a predominant use of internet search engines. These examples illustrate the various tools which may be used to investigate a complex situation. The scope of this project is restricted on the one hand to the basic Google web search function. No attempt is made to deal with other internet search tools, or other Google services, nor the services provided by other library sectors. While this number of queries is not enough to allow a claim that the results are applicable to all situations, the depth of qualitative analysis should allow a good understanding of the issues.

Methodology

Only the main points of the study methodology are described here: full details may be found in Brophy (2004). A case study approach was adopted, with both quantitative and qualitative aspects. Only the main points of the method adopted are described here rather the full details may be found in Brophy (2004). Quantitative results allowed an evaluation of recall, precision and similar factors, while qualitative results allowed inclusion of ideas of quality and value of information (Bawden, 1990). Using a small number of cases, each with a test query searched by the investigator, allowed a detailed study of the documents retrieved, and of the reasons for their not being retrieved from each service. This led to a detailed understanding of the reasons for the relative performance of the services.

The queries used were open-ended “research based” questions (Bilal, 2001), rather than closed reference style queries, since these allowed the performance of the services to be tested to best effect. To avoid any kind of possible researcher biasness in the study the choice of topic are taken four general subject areas from different domains. For example we are taking here Environmental Science, business studies, law and music. It may be noted that the topics are as likely to be covered in web documents as well as in library database records or ~~print~~ vocabularies so that both the services are designed to perform well. The test queries are then searched on the Google, databases as well as resources suitable in a library. The readings of the searched queries are then analyzed to compare the quality, relevance and accessibility of the documents retrieved, and the coverage of these documents in the internet search engines and library resources.

Major Findings

The case study method is adopted to critically analyze the changing pattern of acquiring knowledge and comparison between the queries resolved based on the internet search engines and library resources. The results from the four different subjects are evaluated to ensure the quality, currency, relevance, accessibility and uniqueness of the retrieved information. A summary of results only is given here. A fuller treatment is given in Brophy (2004). A total number of 723 documents were retrieved over all the searches for the four test queries. Of these, 237 from Google and 163 from Library resources were evaluated on the basis of provided framework.

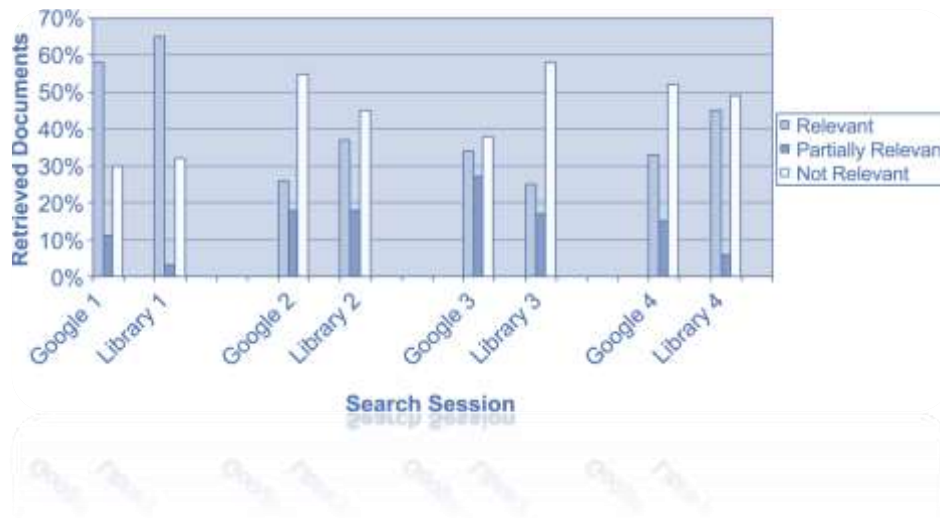


Figure 1 represents the percentage of retrieved documents within each of three relevance categories.

The below graphical figure shows a summary of the results of the Quality assessments conducted on the results of all the four tasks. Overall, 52 percent of the Google results are found to be of good quality while library resources came up with 84 percent of quality of the information retrieved (Figure 3). This seems to indicate that good quality results are only to be guaranteed from library systems. However, only a very small percentage of top-ranked Google results (4 per cent) were poor quality. Only two broken links were found when evaluating the four Google document result sets, indicating the system is updated frequently in order to remove any bad links. Google should not, therefore, be dismissed as providing consistently poor quality results.

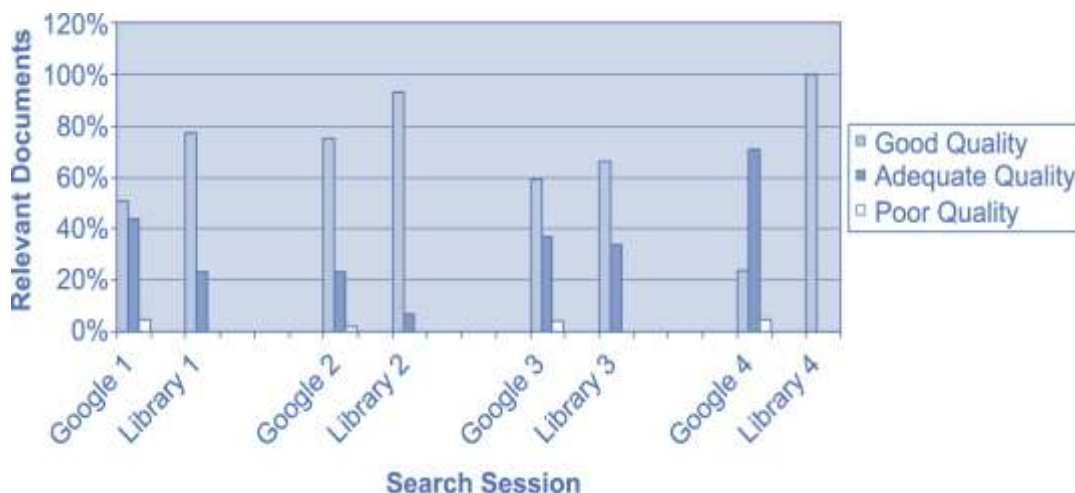


Figure 2 depicts relevant documents by quality.

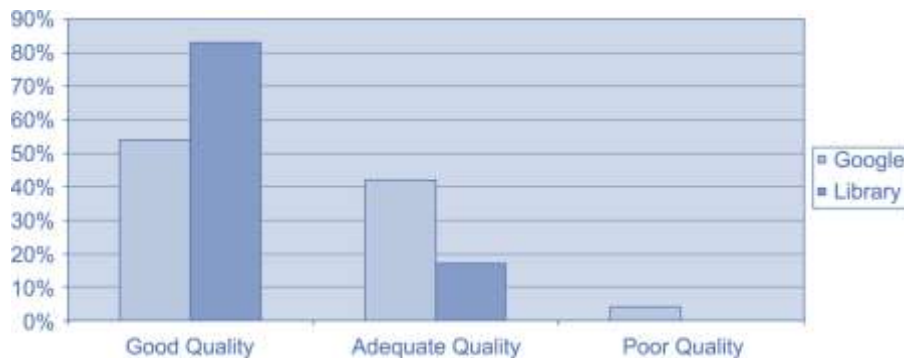
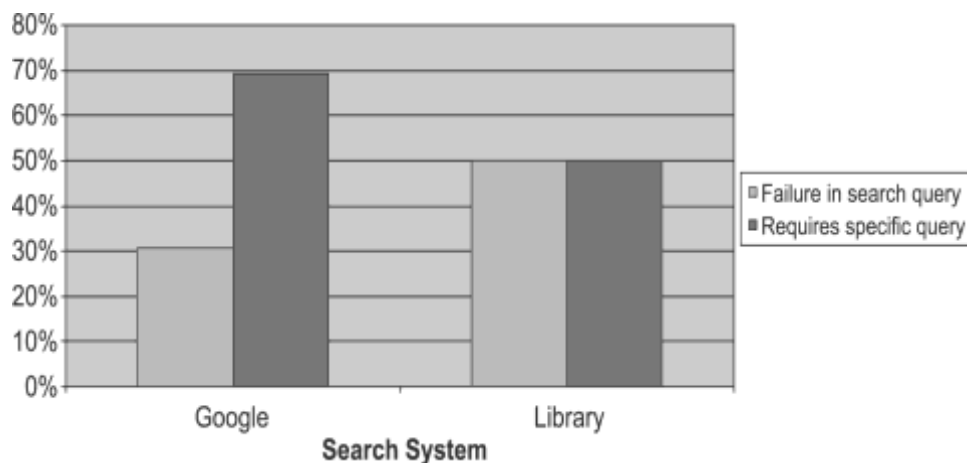


Figure 3 represents the quality comparison between results of google and library resources.

The presumed ability of search engines like Google to provide immediate access to full-text material is confirmed in the results for accessibility, which consistently shows Google providing immediate access to over 90 per cent (over 97 per cent in three out of four tasks) of its relevant results.



The results for the library systems were more varied than for Google, with 65 per cent of the total results being found to be immediately accessible. In search sessions 2-4, a significant number of documents had to be sourced via the British. This would undoubtedly be off-putting to many students due to the time and costs involved in accessing the information. Whilst the numbers are low (less than 10 per cent in three out of four cases), relevant good quality documents were ultimately found which were unavailable even in the British Library collections. Overall, almost 35 per cent of the relevant library documents were found to require either an Inter-Library Loan or be inaccessible altogether. Virtually all Google documents were immediately accessible or not accessible at all.

Conclusion

The study concludes that the two kinds of knowledge acquiring criterion i.e. web search engine and library databases seem to be complementary, as evaluated by their performance over this small set of test queries.

Quality Level	Google Mean (%)	Library Mean (%)
Good Quality	52	84
Adequate Quality	44	16
Poor Quality	4	0

Accessibility level	Google (mean) (%)	Library (mean) (%)
Immediately accessible	96	65
Accessible with difficulty	2	28
Not Accessible	2	7

Google appears clearly superior in both coverage and accessibility. Google also gives more unique items, though the difference is small, and both types of resource are needed if good coverage is required.

The library databases are superior for quality of results, though Google’s performance here is not unacceptable, when only the top ten results are considered.

Precision is similar for both systems, again with the caveat that only the top ten Google results are considered. The precision is not particularly impressive in either case, being less than 60 per cent. Intriguingly, improving the skills of the searcher is likely to give better results from the library systems, but not from Google. This has implications for user awareness and training programmes. It may be seen as a worrying factor, given the tendency identified above, and emphasized in the findings of Griffiths and Brophy (2005) for ease of use, and by implication lack of need for training, to be the major factor in choice of source. If systems like Google are indeed usable to best effect without training to improve competence, then it is highly likely that they will be preferred.

In comparing the systems, in terms of advantages and disadvantages, the conclusions can be summarized as follows:

Google

- a high proportion of relevant documents retrieved;
- an ability to retrieve a fairly precise set of documents;
- a high proportion of adequate or good quality results;
- a high proportion of unique documents; and
- no problems with accessibility.

Library databases

- a moderate proportion of relevant documents retrieved;
- an ability to retrieve a fairly precise set of documents;
- a high proportion of good quality results;
- a high proportion of unique documents; and
- some problems with accessibility.

In the shorter term, there may be a role for information specialists, in their capacity as facilitators of information literacy, in helping their users to appreciate the limitations of all available systems, and suggest strategies to overcome them. This could also include some understanding of the nature of academic information and its structuring. While it is perhaps unreasonable to expect a typical student to gain Borgman’s conceptual framework in its entirety, some realistic understanding of the world of information may be a realistic aim.

REFERENCES

1. Agosto, D.E. (2002), “Bounded rationality and satisficing in young people’s web-based decision-making”, *Journal of the American Society for Information Science and Technology*, Vol. 53 No. 1, pp. 16-27.
2. Barry, C.L. and Schamber, L. (1998), “Users’ criteria for relevance evaluation: a cross-situational comparison”, *Information Processing and Management*, Vol. 34 Nos 2/3, pp. 219-36.
3. Bawden, D. (1990), *User-Oriented Evaluation of Information Systems and Services*, Gower, Aldershot.
4. Brophy, J. (2004), “Is Google good enough? A study comparing a major search engine with academic library research”, unpublished MSc dissertation, Department of Information Science, City University London, London, available at: www.city.ac.uk/library
5. Chu, H. and Rosenthal, M. (1996), “Search engines for the world wide web: a comparative study and evaluation methodology”, *ASIS 1996 Annual Conference Proceedings*, pp. 19-24.
6. Fialkoff, F. (2005), “Access by Google”, *Library Journal*, 15 January, available at: www.libraryjournal.com/article/CA491137
7. Griffiths, J.R. and Brophy, P. (2005), “Student searching behaviour and the web: use of academic resources and Google”, *Library Trends*, Vol. 53 No. 4, pp. 539-54.
8. Mostafa, J. (2005), “Seeking better web searches”, *Scientific American*, Vol. 292 No. 2, pp. 51-7.
9. Schwartz, C. (1998), “Web search engines”, *Journal of the American Society for Information Science*, Vol. 49 No. 11, pp. 973-82.
10. Tenopir, C. (2004), “Is Google the competition”, *The Library Journal*, 1 April, available at: www.libraryjournal.com/article/CA405423.
11. Tenopir, C. (2005), “Google in the academic library”, *The Library Journal*, 1 February, available at: www.libraryjournal.com/article/CA498868