

Editorial: Evaluating Web search engines

Dirk Lewandowski

Dirk Lewandowski, (2011) "Evaluating web search engines", *Online Information Review*, Vol. 35 Iss: 6

Purpose – The purpose of this Guest Editorial is to introduce the theme of this special issue.

Design/methodology/approach – A brief summary is given of the research purposes in Web search engine evaluation.

Findings – While there are established evaluation methods in IR, Web search engines have to be treated separately. Evaluating them is of special importance, not only due to their wide usage, but also from a societal point of view. As search engines are used for a variety of purposes, studies with a wide focus are needed, and studies dealing with specialised information are required as well.

Originality/value – This editorial sets the theme for the special issue and establishes the context of the work presented in the papers.

Keywords – World Wide Web, search engines, information retrieval, evaluation

Paper type – Viewpoint

Every month, more than 130 billion queries worldwide are entered into the search boxes of general-purpose Web search engines (ComScore, 2010). This enormous number shows that Web searching is not only a large business, but also that many people rely on the search engines' results when researching information. Additionally, users largely rely on Google, which enjoys a 65 percent market share in the United States (ComScore, 2011) and has a greater than 90 percent market share in many European countries (see Lunapark, 2011). However, this focus on Google may obscure the fact that the total number of search queries executed on all major search engines is continuously growing, i.e., even search engines that fall behind in terms of market share grow in terms of the total number of queries.

Speaking of the importance of Web search engines in general, it is surprising to find that this importance is quite contrary to the users' approach towards them. In general users are competent neither in the use of search engines nor in evaluating their results (Machill, Neuberger, Schweiger, & Wirth, 2004).

An obvious goal of all search engine evaluation efforts is to generate better systems. This goal is of major importance to the search engine vendors who can directly apply evaluation results to develop better ranking algorithms. However, there are a multitude of other goals in search engine evaluation that are often forgotten. This failure has to do with the assumption that the developers of search systems themselves need to agree to have their systems evaluated and to let them be compared in a lab setting, such as in the TREC evaluation initiative. However, one can see that none of the major search engine vendors takes part in these evaluations. Also, it is questionable whether the results from these relatively small-scale evaluation efforts also hold true when scaled to Web-size.

Comparing Web search engines may answer an important question, i.e., Do users choose to use a search engine (e.g., Google) solely because of its lead in quality, or are other

reasons responsible for this choice? When thinking into which millions of Euros are invested in developing search technology in government-funded initiatives such as *Theseus* in Germany and *Quaero* in France, it would be interesting to know whether the proposed quality lead of such efforts would be useful in terms of users' choices or whether a simple shift in market shares between the existing major search engines would be enough to give the users a good choice between high-quality search engines.

Another question one could ask is why the quality of search engines, and especially the quality of their results, is of such importance. While it is evident that users would like to use a search engine that produces relevant results, the usefulness of direct comparisons might not be that evident at first. Or, as Péter Jacsó put it in his column in this journal, 'In the ideal world one perfect search engine would suffice' (Jacsó, 2008, p. 864). However, this statement overlooks that a choice among search engines would be desirable for the users, as search engines rank documents differently and, therefore, return different results for the same query (Spink, Jansen, Blakely, & Koshman, 2006). Accordingly, users might see the results from an additional search engine as a 'second opinion' on their query.

Regarding 'search neutrality', a topic that has generated greater interest recently, it is at least questionable that the major Web search engines always show the 'best' results for a query at the top of the Web page they return. Some researchers have found that the search engines favour their own offerings within their results lists (Höchstötter & Lewandowski, 2009; Edelman, 2010; Edelman & Lockwood, 2011). Thus, some monitoring of the search engines would be desirable from a societal point of view. Also, it is not desirable that just one search engine should determine what users get to see.

In this special issue, we have methodological, as well as empirical, papers. However, the major difference between the papers lies in their approaches to evaluating search engines in meeting 'general-purpose' or specialised information needs. As Web search engines are used for so many purposes, we need both types of studies. It is greatly important that we know how good search engines are in general, but we also need to know more about their ability to answer specialised queries.

Although evaluating the results of Web search engines is of great importance, one should bear in mind that, regarding the results, quality is just one factor in overall search engine quality measurements. According to Lewandowski & Höchstötter (2008), search engine quality measurement can be grouped into four major areas: index quality, quality of the results, quality of the search features, and search engine usability.

Finally, we want to stress that the results from the studies of Web search engines can, at least to some extent, be adapted to other information systems. We need to develop evaluation methods for the quality of the results of Web site searches, of intranet searches, and of many other search purposes. The results from the studies presented in this special issue can help show that a variety of possibilities exists and that there is not only one general approach to evaluating search results.

Acknowledgments

The guest editor would like to thank all the reviewers for their good work and timely responses.

References

- ComScore. (2010). comScore Reports Global Search Market Growth of 46 Percent in 2009. Retrieved from http://comscore.com/Press_Events/Press_Releases/2010/1/Global_Search_Market_Grows_46_Percent_in_2009.
- comScore Releases May 2011 U.S. Search Engine Rankings. (2011). . Retrieved June 20, 2011, from http://www.comscore.com/Press_Events/Press_Releases/2011/6/comScore_Releases_May_2011_U.S._Search_Engine_Rankings.
- Edelman, B. (n.d.). Hard-Coding Bias in Google "Algorithmic" Search Results. Retrieved April 11, 2011, from <http://www.benedelman.org/hardcoding/>.
- Edelman, B., & Lockwood, B. (2011). Measuring bias in "organic" web search. Retrieved January 20, 2011, from <http://www.benedelman.org/searchbias/>.
- Höchstötter, N., & Lewandowski, D. (2009). What users see – Structures in search engine results pages. *Information Sciences*, 179(12), 1796-1812. doi: 10.1016/j.ins.2009.01.028.
- Jacsó, P. (2008). How many web-wide search engines do we need? *Online Information Review*, 32(6), 860-865. doi: 10.1108/14684520810923971.
- Lewandowski, D., & Höchstötter, N. (2008). Web Searching: A Quality Measurement Perspective. In A. Spink & M. Zimmer (Eds.), *Web Search: Multidisciplinary Perspectives* (pp. 309-340). Berlin, Heidelberg: Springer.
- Lunapark (2011). Suchmaschinen-Marktanteile. Retrieved from <http://www.luna-park.de/home/internet-fakten/suchmaschinen-marktanteile.html>.
- Machill, M., Neuberger, C., Schweiger, W., & Wirth, W. (2004). Navigating the internet: a study of German-language search engines. *European Journal of Communication*, 19(3), 321-347.
- Spink, A., Jansen, B. J., Blakely, C., & Koshman, S. (2006). A study of results overlap and uniqueness among major web search engines. *Information Processing & Management*, 42(5), 1379–1391. Elsevier.