

Living in a world of biased search engines

Dirk Lewandowski

Associate Editor – Europe

Online Information Review

This is a preprint of an article accepted for publication in *Online Information Review* Vol. 39 (2015)

About the author

Dr Dirk Lewandowski is Professor of Information Research and Information Retrieval in the Hamburg University of Applied Sciences, where he specialises in search engine information retrieval. He is also editor and a contributing author of *Web Search Engine Research* (Emerald, 2012) and Editor-in-Chief of the *ASLIB Journal of Information Management* (Emerald). His papers have appeared in the *Journal of the Association for Information Science and Technology*, *Journal of Information Science*, *Online Information Review* and many other scholarly journals. He is also Associate Editor – Europe for *Online Information Review*. Professor Lewandowski may be contacted at dirk.lewandowski@haw-hamburg.de.

When looking at internet usage, we see that searching is one of the dominant activities (Purcell *et al.*, 2012), with more than 18 billion queries entered into the general-purpose search engines every month on desktops in the US alone (ComScore, 2015). We all search every day, and we predominantly use Google for that purpose. In many European countries Google has a market share of well over 90 percent (ComScore, 2013), and even in “multi-search engine markets” such as the US there are only two real competitors, namely Google and Bing. While Yahoo is often regarded as a search engine, the company gave up their own search technology in 2009 and have been showing results provided by Bing since then.

On the surface there appears to be quite a variety of alternative search engines from which to choose. However, on closer inspection we find that, as with the Yahoo-Bing search alliance, search engines power each other: a search engine with its own index gives its results to other seem-to-be search engines. This so-called partner index model has served to thin out the competition in the search industry (Clay, 2011). Why should a company invest in search

technology (and this is a heavy investment indeed!) when plenty of income can be generated through providing someone else's results and advertisements?

Then there are so-called alternative search engines, which are being founded on a regular basis. The press often presents them as real alternatives to Google; but the reality is that they often rely on the partner index model, or, if they use their own databases, these are much too small to compete with Google or Bing.

As the web contains billions of documents and is very dynamic, it requires a large technical and financial effort to build an index that is both as complete and as current as possible. Even the best search engines are not able to cover the whole web, as well as keep their indices current (Lewandowski, 2008). Looking back over the last decade or so, we can see that no company, except Microsoft, is able and willing to invest in its own web index.

As a consequence, we rely on just two alternative views of the large amounts of web data when we use search engines – either Google's view of the web world or Bing's. Both search engines determine what we see when we type in our search queries.

The presentation of a certain set of results is what I call an algorithmic interpretation of the world, that is, the web data. However, people often assume that search can produce right and wrong results. They think that, if a search engine has found the "magic formula", it can provide its users with the best possible results. But there is no such thing as a "right" results ranking (as opposed to a "wrong" results ranking). At least for informational queries there are often hundreds if not thousands of relevant results. The goal of the search engines in these cases is not to provide a certain set of right/relevant results, but to list some of the potentially relevant results in the top few positions.

Our understanding of the quality of our search results is probably biased by the results we can actually judge. These are mainly the answers to our navigational queries, where we want to find a certain website about which we already know, or assume that it exists (cf. Broder, 2002). If we search for a website like Microsoft's, we can assess with certainty whether the search engine produces the right result in the first position. I assume that users extrapolate their positive experiences with this type of query to informational queries, where the right/wrong distinction does not apply.

That said, we can see that searches are always biased, and there is no such thing as an unbiased search engine. It would be impossible to construct such a search engine, because human beliefs and assumptions influence the design of algorithms, and they therefore prefer certain documents to others. It is even at the core of every idea of ranking that, based on certain technically mediated assumptions, certain items are preferred over others.

This would not be a problem if we had a variety of real search engines – meaning that they provide their own ranked results based on a large-enough database and are not just displaying the same results as one of the big players. However, keeping in mind the current market situation, something needs to be done about the one dominant interpretation of web data.

As Google is the company with the overwhelming market share, we can call this the “Google problem”. It is very welcome that, especially in Europe, a discussion on how to deal with this problem has not only started but also reached the wider public. This can, for instance, be seen from the large number of newspaper articles on the topic in the last year and also from the European Commission’s competition investigation of Google (whatever the result from that investigation may be).

However, whilst the problem has now been recognised, we still lack a solution. Some say that we will just have to wait for the market to create a real competitor to Google. I doubt that a market dominated by one single company for many years will give us a relevant new search engine, and one more search engine will not solve our problem. We need not only one or more new search engines, but also a multitude of players in the search engine markets, serving searchers’ needs from the general to the very specialised. This is also a strong argument against state funding of new search engines, as has been proposed.

The only fruitful solution I can see is building a publicly-funded infrastructure for querying and indexing web data, and having many companies build their services on this infrastructure, whether searching or other applications. Such an open web index would not only benefit competition in the search market but also foster plurality in search results and end the control that one company has over what we are allowed to see from the web (Lewandowski, 2014).

References

- Broder, A. (2002), “A taxonomy of web search”, *ACM SIGIR Forum*, Vol. 36 No. 2, pp. 3-10.
- Clay, B. (2011), “Search engine relationship chart histogram”, [online], available at: http://www.bruceclay.com/serc_histogram/histogram.htm (accessed 1 March 2015).
- ComScore (2013), *Future in Focus: Digitales Deutschland 2013*. Reston, VA: ComScore, Inc.
- ComScore (2015), “ComScore releases January 2015 US desktop search engine rankings”, [online], available at: <http://www.comscore.com/Insights/Market-Rankings/comScore->

Releases-January-2015-US-Desktop-Search-Engine-Rankings (accessed 1 March 2015).

Lewandowski, D. (2008), "A three-year study on the freshness of web search engine databases", *Journal of Information Science*, Vol. 34 No.6, pp. 817-31.

Lewandowski, D. (2014), "Why we need an independent index of the web", in König, R. and Rasch, M. (Ed.), *Society of the Query Reader: Reflections on Web Search*, Amsterdam: Institute of Network Culture.

Purcell, K., Brenner, J. and Raine, L. (2012), *Search engine use 2012*, Washington, DC: Pew Research Center, [online], available at: http://pewinternet.org/~media/Files/Reports/2012/PIP_Search_Engine_Use_2012.pdf (accessed 1 March 2015).