

Search engine user behaviour: How can users be guided to quality content?

Dirk Lewandowski *

Department of Information, Faculty of Design, Media and Information, Hamburg University of Applied Sciences, D-20249 Hamburg, Germany

Abstract. The typical behaviour of the Web search engine user is widely known: a user only types in one or a few keywords and expects the search engine to produce relevant results in an instant. Search engines not only adapt to this behaviour. On the contrary, they are often faced with criticism that they themselves created this kind of behaviour. As search engines are trendsetters for the whole information world, it is important to know how they cope with their users' behaviour. Recent developments show that search engines try to integrate results from different collections into their results lists and to guide their users to the right results. These results should not only be relevant in general, but also be pertinent in the sense of being relevant to the user in his current situation and in accordance to his background.

The article focuses on the problems of guiding the user from his initial query to these results. It shows how the general users are searching and how the intents behind their queries can be used to deliver the right results. It will be shown that search engines try to give some good results for everyone instead of focusing on complete result sets for a specific user type. If the user wishes, he can follow the paths laid out by the engines to narrow the results to a result set suitable to him.

Keywords: Search engines, user behaviour, web searching, search shortcuts

1. Introduction

This article is about some newer developments with web search engines. There have been some changes recently at Google, Yahoo and others that affect both the presentation of search results, and user guidance on the results pages.

In the first section of this article, the situation of search engines today and of searching in the scientific context will be described. Then, an overview on the user behaviour in search engines will be given. It is not only important to know the users' behaviour in order to improve our understanding of search engines, but also to understand what users want from search systems in general. Search engines are the trendsetters in user searching behaviour.

The main part of the article will be on quality content and how it can best be presented in search engines. We usually see search engines as providers of low-quality content from the entire web. However, search engines use quality factors in their rankings. Some of these factors will be described and it will be shown what could be regarded as quality content.

In the conclusion, we will suggest ways for libraries to adapt to the behaviour of the typical search engine user. We will show how the strategies of the search engines can be applied to library catalogues.

*Address for correspondence: D. Lewandowski, Department of Information, Faculty of Design, Media and Information, Hamburg University of Applied Sciences, Berliner Tor 5, D-20249 Hamburg, Germany. E-mail: dirk.lewandowski@haw-hamburg.de.

2. Current situation with search engines and user behaviour

On the web, search engines clearly dominate. Other approaches such as web directories, social book-marking or question answering services only play an underpart.

The major problem in web searching is still the same as it was 10 years ago: relevance. Surely, search engines have improved a great deal over the years, but their results are still far from perfect. In the scientific context, we find that both students and researchers use the same strategies in search engines and in the library catalogues or special databases. In the last few years, companies like Google or Microsoft created special search engines for scientific contents (such as Google Scholar and Windows Live Academic¹), where users can easily find the documents that they want without changing their searching behaviour. Even “classic” vendors of information services like Thomson started to add web search to their products.

This distinguishes scientific search engines from what libraries have to offer. Today, there are still some searching skills needed not only to get relevant results from the library catalogue, but also to find the resources and databases that are suitable for a certain scientific discipline, even if the library offers access to it [1].

When looking at the searching behaviour with web search engines, we find that the average query length is 1.7 words for German language queries [2]. For the English language [3], queries are a bit longer, but this has to do with the specifics of the German language, where there is heavy use of compound words. For example, “search engine user behaviour” in English is four words, while the German Suchmaschinennutzerverhalten is just one word. Therefore, approximately 50% of German queries consist of just one word [2], where with English queries, the percentage is a bit lower.

We all know from the general search engines that they return many results for almost all queries. However, users tend to consider only the results on the first page [3], which usually consists of only 10 results. Even more, users focus on just the first few hits, the ones that they can see without scrolling to the bottom of the page [4,5].

When evaluating the results, users view just a few results (usually not more than five) and the average session length is less than 15 min [6]. Adapting this to the library users’ behaviour means that the user wants to get to the relevant results within 15 min. Otherwise, he or she will terminate the search session. Research shows that with limited time, even for scientific information needs, search engines could be a better choice for the searcher than the library’s offerings [7].

Interestingly, search engine users are easily satisfied. When asked about the quality of the results they achieve, they usually express satisfaction with their searching strategies [8]. Therefore, they don’t see why they should apply sophisticated searching strategies to other search systems. If Google is able to return relevant results for a one-word query, then every other system should be able to do the same.

One problem that search engines have to deal with are the different types of queries posed to them. According to Andrei Broder, there are three query types: informational, navigational and transactional [9]. With informational queries, users want to find information on a certain topic. Such queries usually lead to a set of results rather than just one suitable document. Informational queries are similar to queries sent to traditional text-based information retrieval systems. According to Broder, such queries always target static Web pages. However, the term static here should not refer to the technical delivery of the pages (e.g., dynamically generated pages by server-side scripts such as php or asp) but rather to the fact that once the page is delivered, no further interaction is needed to get the desired information.

¹Microsoft put Windows Live Academic to an end in June 2008, see: <http://blogs.msdn.com/livesearch/archive/2008/05/23/book-search-winding-down.aspx>.

In opposition to that, navigational queries are used to find a certain Web page the user already knows about or at least assumes that such a Webpage exists. Typical queries in this category are searches for a homepage of a person or organization. Navigational queries are usually answered by just one result; the informational need is satisfied as soon as this one right result is found.

The results of transactional queries are Web sites where a further interaction is necessary. A transaction can be the download of a program or file, the purchase of a product, or a further search in a database.

Each query type accounts for a considerable amount of Web queries [10]. Determining the query type from the query itself is not a trivial task. However, it can help providing the right results.

3. Quality on the Web and quality results in search engines

“Right results” could mean many things. In this section, it will be explained how search engines consider the quality of web pages in ranking the results for a certain query.

First, it is important to emphasise that search engines do consider the quality of the web pages they show to the user. Ranking results is much more than just text matching. The most prominent quality rating for web pages are the link-based popularity measures such as Google’s PageRank [11,12]. The underlying assumption is that a page that has many other pages linking to it is more important than pages that do not.

However, link-based measures may be the most popular and best known, but there are also other factors. A page’s freshness is of great importance. For example, there are lots and lots of pages dealing with the TV show “I’m a celebrity – get me out of here!”, but only some deal with the current season. Older results are typically not relevant for a user who is typing in the name of the show.

Another factor is the user’s location. Results for the same query should be one thing for a user based in the UK and something else for a user based in the US.

But there is even more to consider. If a user types in a broad query, it is useful to show broad results. There could be many relevant pages where the information goes too much into detail for a user with a broad interest.

The problem with all these quality factors that search engines consider is that these can only be applied to the individual pages. It is not possible to control that every time a user types in the name of a film, a certain mix of informative pages would be provided.

On their results pages, search engines offer much more than just the results list. Figure 1 shows a results list from Microsoft’s Live.com search engine. This is a typical results page that consists of the same elements as those that are found in other search engines.

In the first red frame, we can see navigational elements. The links point to results from additional databases such as news and videos. In the second frame, an advertisement is shown.

The third frame is more interesting. Here, some news articles relevant to the query are shown. A user could get to the same results (and more) by clicking on the “news” link above the results list. However, click analysis shows that users only seldom click on these so-called “tabs”; in fact, they are so seldom clicked that there is a term for this: “tab blindness”.

Finally, in the last frame, the beginning of the regular results list is shown. It consists of the results from the web database. Note that in this case, the news results are positioned above the web results.

This sample results page shows just one additional set of results: that from the news database. However, there are many more specialised databases, some of which are video, blogs, books and scientific contents [13]. It is interesting to note that there are some databases that stand in competition to the

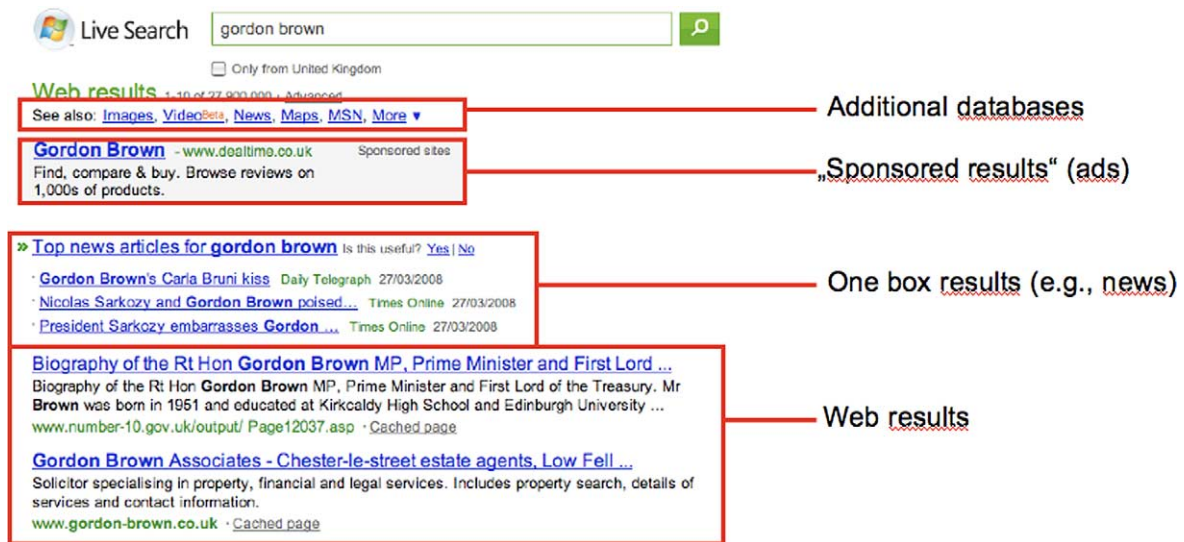


Fig. 1. Results page from Live.com.



Fig. 2. Google Scholar results above the regular web results list.

libraries' offerings, namely books and scientific content. When searching Google with "seem to be" scientific queries, this search engine offers a few results from Google Scholar above the web results (see Fig. 2). Users are guided to the specialised results set and with this, to the specialised search engine.

So, looking at quality content in web search engines a bit more systematically, we find three groups of quality content:

1. The first group are authoritative Web pages. These could be official sites that should be included in the results set. For example, for the query "George W. Bush", his official biography at white-house.com should be shown. There are also authoritative sources for certain topics. For example, in most cases the Internet Movie Database (IMDB) offers good results when a user types in the name of a film or an actor. Authoritative Web pages come from the general Web crawl of the search engine. Presenting these results prominently is just a question of ranking.
2. The second possibility is the inclusion of results from additional databases, as shown above. Such results are shown when a user's query produces results in such a database.

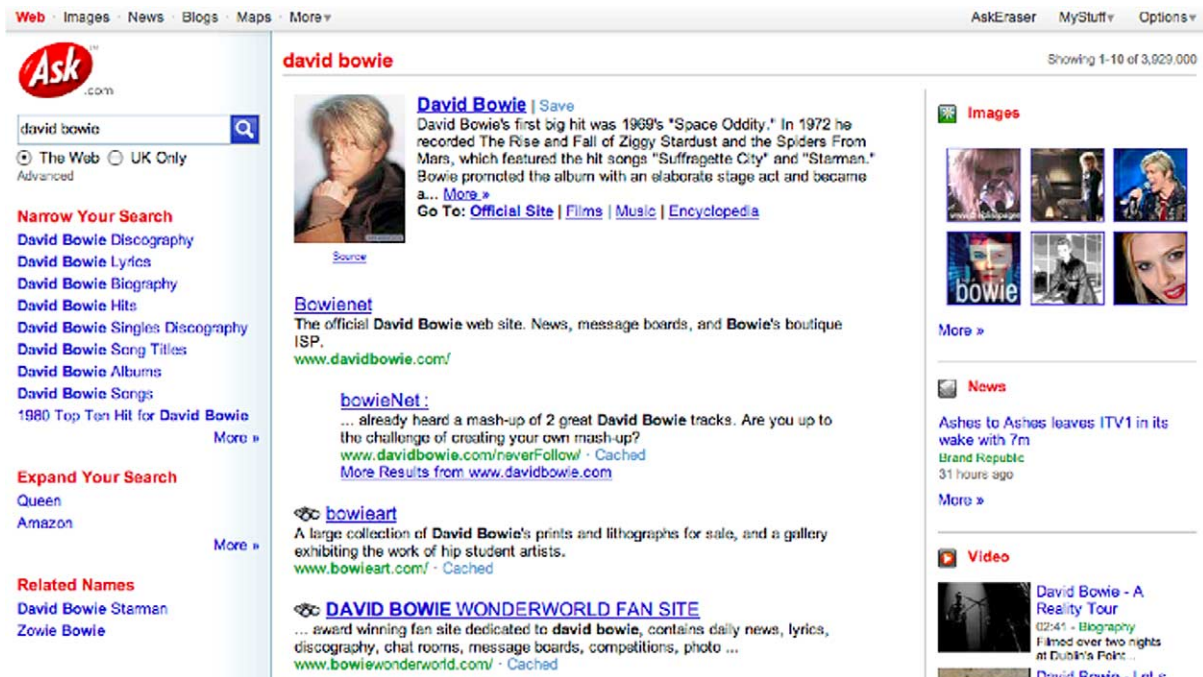


Fig. 3. Results for “David Bowie” on Ask.com.

3. The third possibility is to present pointers to external databases instead of directly presenting results from a certain database. This is shown in Fig. 3 for the query “David Bowie” on Ask.com. Above the web results, there is a box with some information on him, but also with links to other databases. These are not necessarily connected to the search engine. Instead, links to these external databases are shown each time a user searches for a celebrity. The links shown point to the Internet Movie Database, the All Music Guide and to Wikipedia.

In this case, the underlying databases don’t need to be queried when a user types a query into the search engine. The links are static and no processing power or time are wasted.

Figure 4 shows quite the opposite. This is a results list from Google’s “Universal Search” where results from several databases are mixed into the major results list. The example shows a part of the results list for the query “Nosferatu”. Apart from the usual Web pages the user gets a video of this famous silent film as the fourth result, and somewhere within the results list (here shown at the bottom), there is a selection of related searches.

It may not seem complicated to mix results from different sources. But it should be kept in mind that the results must be presented in less than a second. Every potentially relevant database has to be queried and results from all sources must be ranked into a single results list.

Yahoo offers similar results lists to Google’s “Universal Search”. These mixed results list show where user-centred search is heading. Users are not aware of all the databases the search engines have to offer. Therefore, search engines show results from all databases in one results list where the user can decide which results are the most suitable.

This leads back to the offerings of libraries. The problem there is very similar: libraries offer a wide range of specialised databases, but do the users know about them? If not, how can libraries publicise and promote these information sources?

Web [Images](#) [Maps](#) [News](#) [Shopping](#) [Mail](#) [more ▼](#)

Google [Advanced Search](#) [Preferences](#)

Web [Video](#) [Books](#) [Images](#)

Related searches: [nosferatu picture](#) [nosferatu pics](#) [nosferatu images](#) [nosferatu summary](#)

[Nosferatu - Wikipedia, the free encyclopedia](#)
 This article is about the 1922 silent film. For the 1979 remake, see **Nosferatu the Vampyre**.
 For other uses, see **Nosferatu** (disambiguation). ...
en.wikipedia.org/wiki/Nosferatu - 61k - [Cached](#) - [Similar pages](#)

[Nosferatu, eine Symphonie des Grauens \(1922\)](#)
 Directed by FW Murnau. With Max Schreck, Gustav von Wangenheim, Greta Schröder.
 Vampire Count Orlok expresses interest in a new residence and real estate ...
www.imdb.com/title/tt0013442/ - 57k - [Cached](#) - [Similar pages](#)

[Nosferatu: Phantom der Nacht \(1979\)](#)
 Directed by Werner Herzog. With Klaus Kinski, Isabelle Adjani, Bruno Ganz. Jonathan and Lucy live in Wismar and the Count wants a house there.
www.imdb.com/title/tt0079641/ - 55k - [Cached](#) - [Similar pages](#)
[More results from www.imdb.com »](#)

 **[Nosferatu](#)**
Nosferatu (1922) Originally released in 1922 as **Nosferatu** ...
 84 min - ★★★★★
video.google.com/videoplay?docid=-6185283610506001721

[Nosferatu - Directed by F.W. Murnau](#)
 Includes cast/crew information, review, synopsis, background, pictures and sounds.
nosferatufilm.com/ - 8k - [Cached](#) - [Similar pages](#)

[Dracula - Google Books Result](#)
 by Bram Stoker, Maurice Hindle, Christopher Frayling - 2003 - Fiction - 560 pages
 A completely revised edition of Bram Stoker's masterpiece of horror
books.google.com/books?isbn=014143984X...

[Official Nosferatu - Carpe Noctem Website](#)
 Let me again shout out a BIG thank you to all the fans and all my team members who are still supporting **Nosferatu** even though I had very little time for it ...
www.nosferatuthegame.com/ - 15k - [Cached](#) - [Similar pages](#)

[Internet Archive: Details: Nosferatu](#)
 Originally released in 1922 as **Nosferatu**, Eine Symphonie Des Grauens, director F.W. Murnau's chilling and eerie adaption of Stoker's Dracula is a silent ...
www.archive.org/details/nosferatu - 42k - [Cached](#) - [Similar pages](#)

 **[Nosferatu 1922](#)**
 B&W silent film with music. public domain. KEEP CLASSIC ...
 80 min - ★★★★★
www.youtube.com/watch?v=MEQsb6CRvNU

[Nosferatu Official Homepage Gothic Rock Band Nosferatu](#)
 The Official **Nosferatu** MY SPACE <http://www.myspace.com/officialnosferatu> ... The Official **Nosferatu** Forum www.nosferatuforum.forumup.org ...
www.nosferatu.fsworld.co.uk/ - 6k - [Cached](#) - [Similar pages](#)

Searches related to: **[nosferatu](#)**

[nosferatu picture](#) [nosferatu pics](#) [nosferatu images](#) [nosferatu summary](#)
[nosferatu dvd](#) [nosferatu clan](#) [vampires](#) [nostradamus](#)

Fig. 4. Results from Google's "Universal Search".

4. Conclusion

There are things that libraries can learn from the search engines. (In fact, there are many things that search engines could learn from libraries, but that is another issue . . .)

When we look at the new library catalogues, we see a lot of Web 2.0 features (see, for example, [14] or [15]). These are surely of great help, but one should not forget that they do not solve the search problem! Users still only search one collection when searching the online catalogue [1]. They are often not aware of the rich offerings in special databases licensed by their library.

Search engines can teach us that it's not only about offering these great collections, but also to guide users to them. Pointers to or even results from the databases should be shown within or around the results lists in the library catalogue. This will make the online catalogue more a scientific search engine than a classic catalogue. This is the way libraries should go. There is already competition from the search engine vendors that offer scientific search engines.

There was (and is) a great deal of discussion about Google Scholar [16–20]. Now it is time for libraries to offer their own academic search engines, with better content and a better presentation of their results.

However, this goal can only be reached if librarians, database vendors, publishers and OPAC vendors are working together on the next generation of the library catalogue (see [21]).

References

- [1] D. Lewandowski, Suchmaschinen als Konkurrenten der Bibliothekskataloge: Wie Bibliotheken ihre Angebote durch Suchmaschinentechnologie attraktiver und durch Öffnung für die allgemeinen Suchmaschinen populärer machen können, *Zeitschrift für Bibliothekswesen und Bibliographie* **53** (2006), 71–78.
- [2] N. Höchstötter and M. Koch, Standard parameters for searching behaviour in search engines and their empirical evaluation, *Journal of Information Science* **34** (2008), to appear.
- [3] B.J. Jansen and A. Spink, How are we searching the World Wide Web? A comparison of nine search engine transaction logs, *Information Processing & Management* **42** (2006), 248–263.
- [4] L.A. Granka, T. Joachims and G. Gay, Eye-tracking analysis of user behavior in WWW search, in: *Proceedings of Sheffield SIGIR – Twenty-Seventh Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Sheffield, 2004, pp. 478–479.
- [5] M.T. Keane, M. O'Brien and B. Smyth, Are people biased in their use of search engines?, *Communications of the ACM* **51** (2008), 49–52.
- [6] A. Spink and B.J. Jansen, *Web Search: Public Searching of the Web*, Vol. 6, Kluwer Academic Publishers, Dordrecht, Boston–London, 2004.
- [7] M. Machill, M. Beiler and U. Neumann, Leistungsfähigkeit von wissenschaftlichen Suchmaschinen. Ein Experiment am Beispiel von Google Scholar, in: *Die Macht der Suchmaschinen/The Power of Search Engines*, M. Machill and M. Beiler, eds, Herbert von Halem, Köln, 2007, pp. 327–347.
- [8] M. Machill, C. Neuberger, W. Schweiger and W. Wirth, Wegweiser im Netz: Qualität und Nutzung von Suchmaschinen, in: *Wegweiser im Netz*, M. Machill and C. Welp, eds, Bertelsmann Stiftung, Gütersloh, 2003, pp. 13–490.
- [9] A. Broder, A taxonomy of web search, *SIGIR Forum* **36** (2002), 3–10.
- [10] D. Lewandowski, Query types and search topics of German Web search engine users, *Information Services & Use* **26** (2006), 261–269.
- [11] A.N. Langville and C.D. Meyer, *Google's PageRank and Beyond: The Science of Search Engine Rankings*, Princeton University Press, Princeton, NJ, 2006.
- [12] L. Page, S. Brin, R. Motwani and T. Winograd, The PageRank citation ranking: Bringing order to the Web, 1998.
- [13] D. Lewandowski, Trefferpräsentation in Web-Suchmaschinen, in: *Information in Wissenschaft, Bildung und Wirtschaft; 29. Online-Tagung der DGI 2007*, M. Ockenfeld, ed., DGI, Frankfurt, 2007.
- [14] T. Sadeh, Time for a change: new approaches for a new generation of library users, *New Library World* **108** (2007), 307–316.
- [15] M. Mönnich and M. Spiering, Adding value to the library catalog by implementing a recommendation system, *D-Lib Magazine* **14** (2008), <http://www.dlib.org/dlib/may08/monnich/05monnich.html>.
- [16] P. Jacsó, Google Scholar revisited, *Online Information Review* **32** (2008), 102–114.

- [17] D. Lewandowski, Nachweis deutschsprachiger bibliotheks- und informationswissenschaftlicher Aufsätze in Google Scholar, *Information Wissenschaft und Praxis* **58** (2007), 165–168.
- [18] P. Mayr and A.-K. Walter, An exploratory study of Google Scholar, *Online Information Review* **31** (2007), 814–830.
- [19] J. Pomerantz, Google Scholar and 100 percent availability of information, *Information Technology and Libraries* **25** (2006), 52–56.
- [20] P. Jacsó, Savvy searching: Google Scholar: The pros and cons, *Online Information Review* **29** (2005), 208–214.
- [21] D. Lewandowski and P. Mayr, Exploring the academic invisible web, *Library Hi Tech* **24** (2006), 529–539.