Search engines

Uses and challenges

Brigitte Simonnot

Translator: Inist

Electronic version

URL: <http://journals.openedition.org/questionsdecommunication/8879>

DOI: 10.4000/questionsdecommunication.8879

ISSN: 2259-8901

Publisher

Presses universitaires de Lorraine

Printed version

Date of publication: 1 December 2008

ISBN: 978-2-86480-981-4

ISSN: 1633-5961

Electronic reference

Brigitte Simonnot, « Search engines », *Questions de communication* [Online], 14 | 2008, Online since 01 February 2014, connection on 02 May 2019. URL : <http://journals.openedition.org/questionsdecommunication/8879> ; DOI : 10.4000/questionsdecommunication.8879

This text was automatically generated on 2 May 2019.

Tous droits réservés

Search engines

Uses and challenges

Brigitte Simonnot

Translation : Inist

EDITOR'S NOTE

This English translation has not been published in printed form/Cette traduction anglaise n'a pas été publiée sous forme imprimée.

- 1 Web search engines are barely ten years old, but they have become familiar and sometimes indispensable tools. Their use has become commonplace in a whole host of everyday life situations, within both professional and private settings (Savolainen, 1995). Planning trips, keeping up with the news, looking for online health information, during key milestones in life, or simply for leisure purposes, individuals are increasingly turning to online resources. Commercial search engines capitalize on this audience: several studies have shown that information retrieval (IR) is one of the most popular uses of the Internet, almost on a par with communication tools such as messaging. According to Comscore Inc.¹, during the month of June 2008 alone, Americans made 11.5 billion requests to the five major sites that hold the lion's share of online information retrieval: Google (61.5 %), Yahoo (20.9 %), Microsoft (9.2 %), Ask (4.3 %) and AOL (4.1 %). In France, in May 2008², 2.9 billion requests were submitted by 26 million Internet users, which translates into an average figure of 3.6 searches per day, with Google holding 82% of the market share. Médiamétrie/Netratings provides similar figures for May 2008: the study³ indicates that in France, Google received 26.6 million unique visitors. These results reflect the fact that search engines have become a part of everyday life for Internet users, a trend which highlights the underlying advertising stakes involved. They also show that despite the considerably⁴ high number of online search engines, only very few of them have become mainstream.

- 2 Scientific research in IR was previously confined to the field of libraries and scientific and technical or strategic information, however, it has been boosted by the development of online information services. The extremely specific corpus to which search engines are dedicated – heterogeneous, unstable and fast-growing – as well as the nature of their subjected uses, has led their designers to extend their scope of activity to all sorts of information – data which may be of interest to everybody. At the technical level, a search engine is structured around four main components: (i) a spider or crawler (*robot de collecte*) which crawls the Web to find documents and store them in a database, (ii) an indexing module (*module d'indexation*) that characterizes documents particularly from their key words or any associated citations, (iii) a user interface (*interface utilisateur*) that allows the Internet user to enter a request and consult the results and (iv) a matching module (*module d'appariement*) which, for any given request, prepares a hierarchy or ranking of web pages (*classement des pages*) liable to provide a suitable answer⁵. Each of these modules contains, in turn, several components that refine the data-handling carried out at each stage. The general public are not fully aware of the complexity of these tools, the creators of which continuously seek to simplify their use.
- 3 In the field of Human and Social Sciences, French studies on the uses of search engines remain rare, whereas there is abundance of English literature on this subject, and numerous articles have been published in specialist scientific journals (in particular, *Information Processing & Management*, *Journal of the American Society for Information Science and Technology*, *Journal of Documentation* or *Information Research*). For this special issue of the *Questions de communication* review, two main types of contributions have been selected: first, works based on state-of-the-art knowledge in the domain of search engine use analysis and which offer a deeper understanding of the subject, and second, recent results of empirical research that analyzes search engines or observes their use.
- 4 At present, French terminology in this area of research seems to be relatively poor. For example, the expression “*recherche d'information*” is used interchangeably to designate both “information search” and “information retrieval”: the former is an English term that refers to the activity of searching, to users' information searches, and the latter is a term coined by the American researcher Calvin Mooers (Chiaramella, Mulhem, 2007: 14) that covers all methods and algorithms that enable data collection, indexation, matching and page-ranking according to query. As for the commonly used expression “search engine” (*moteur de recherche*), it remains rather limited for describing these services, especially when they are of a commercial nature⁶: the English term *engine* has both an underlying mechanical and engineering meaning, whereas there is no hint of the latter element in the French equivalent term of *moteurs de recherche*. In the French language, an engine is a machine that transforms all forms of energy into mechanical energy, a force that directs movement⁷. In the figurative sense, it also incites or leads to action. Do engines, through facilitating the search for information, incite Internet users to become more informed? In 1959, Calvin Mooers (in Austin, 2001: 607-608) denounced the belief that “an information retrieval system will tend not to be used whenever it is more painful and troublesome for a customer to have information than for him not to have it”: he put forward the hypothesis that there are a lot of individuals who do not want information and avoid using a system which, precisely, would give them this information. Having information at one's disposal may be painful and troublesome when it must be read, understood and processed – activities for which IR systems have long been of limited use. Search engines may not necessarily mean that individuals are better-informed, but they

do contribute to making information available to the general public, data that, were it not for these engines, would probably have remained confidential.

Analyzing search engine uses

- 5 Analyzing the uses of Web search engines has two major objectives: to gain a better understanding of human informational activity and to improve the functionality of tools and interface design. Madjid Ihadjadene's and Stéphane Chaudiron's contribution⁸ focuses on the methodologies implemented in these studies from a comprehensive state-of-the-art understanding of the issue. Their contribution is doubly interesting, not only because of the considerable amount of research carried out on these issues and their disparity, but also from the point of view of the conclusions they make. The authors distinguish between four principal research categories. The first two fall within the context of behavioural studies on search engine users, and seek to highlight the tactics and strategies implemented by these users and how they read and evaluate the results. These studies face two difficulties: for the results to be representative, they must cover a large amount of data. However, these data are not easily accessible to independent scientific researchers. Search engines' transaction logs provide information that may appear to be rather succinct: IP address from which requests are made, date and time, key words, the operators that Internet users select for their requests, and sometimes the chosen options. However, for scientific researchers with access to these data and for the search engine designers themselves, this information enables them to analyze the often intuitive ways in which Internet users formulate their requests and consult the results. A third research category that is mainly found in the field of cognitive psychology, seeks to differentiate users by extending use(r) typologies. In this category, the most common differentiation is undoubtedly the one drawn between experts and novices. Madjid Ihadjadene and Stéphane Chaudiron show that many of these studies suffer from shortcomings, notably as a result of unclear or inconsistent definitions of the concepts of 'expert' and 'novice'. The final research category focuses more generally on the context of information retrieval and falls within the mainstream study of individual information behaviour. From the current state of the scientific literature, the authors conclude that methodologies are still imperfect and fragmented. They recommend adopting a holistic approach that would enable the complex variables involved in Internet users' IR activities to be addressed.
- 6 Young people are the customary users of search engines, to such an extent that the term "Google Generation" is often used to refer to individuals born after 1993 (CIBER, 2008). In this context, Nicole Boubée analyzes the strategies used by young information seekers online. Her study – which is based on direct observations of middle and high school students seeking information and on explicitation interviews – calls into question the results of previous studies, especially those asserting young people's browsing preference to be "direct" query-type searches. Returning to the overly formulaic novice/expert distinction, she endeavours to show how inventive and creative these young people can be in their search engine habits when they try to devise search strategies that serve their purpose, or to understand the applications' functionality. She denounces a "dominant theoretical perspective" liable to become enclosed in a stigmatization of the knowledge gap amongst young people on informational activity, and calls for further studies that truly take into account the complexity of this activity; an activity which designers of

commercial search engines are continuously striving to simplify, or at least appear to do so.

The quest for the perfect search engine

- 7 The following two contributions focus on the functionalities that the perfect search engine should have. Judit Bar-Ilan's contribution concentrates on the very specific context of uses by researchers working in the field of Web analytics, otherwise known as webometrics or webology. It provides her with the ideal opportunity to revisit the list of ideal functions she has already suggested in a previous article (Bar-Ilan, 2005) and to analyze whether or not these functionalities are offered by the major commercial search engines, and if they are, in what form. Her study highlights that these services are not tools specific to the domain of webometrics or exclusive to scientific research on the Web. Researchers who use these applications to compile their corpus must be warned: data gathered empirically by commercial search engines can in no way be considered to be representative of the Web as a whole. The engine's results are also characterized by a strong temporal instability and disparities in their coverage of resources from different countries. The major commercial search engines have been designed to analyze English documents and it remains unclear as to how they have been adapted for research in other languages. She attributes some of these limitations to the fact that search engines are targeted at the general public – for whom these functions would be less important, whereas webometry researchers have quantitative requirements in relation to results (Thelwall, 2008) – and also to the commercial nature of these services. In fact, commercial search engines are often clouded in secrecy regarding the details of their algorithms, both for competition reasons⁹ and also to combat spammers. The lack of transparency and the applications' inconsistencies require the utmost caution when tapping into them for scientific purposes.
- 8 Dirk Lewandowski and Nadine Höchstötter broaden and extend these remarks by proposing a reference framework for “measuring the quality of Web search engines” and comparing them with one another. Based on both state-of-the-art knowledge and their own investigations, they show the difficulties encountered by researchers in this field. They also underline the methodological issues posed by some studies, thus preventing a generalization or a comparison of obtained results. The authors call for research activities in this sector to be intensified in Europe, in such a way that different countries have different tools adapted to their own informational resources. In the cultural heritage sector, Jean-Noël Jeanneney's plea (2005) has been heard and actioned with the setting up of the Europeana digital library project. However, very few European actors have managed to make their mark in the non-specialized IR domain, a field in which the economic stakes are significant nonetheless. The evaluation framework proposed takes into account both the technical performance of search engines and their adaptation to Internet users' needs, in accordance with Kalervo Järvelin and Peter Ingwersen's appeal (2004; Ingwersen, Järvelin, 2005). As both of these studies show, there is still room for improvement with search engines, not only at the technical level, but also in relation to the information that they provide on the way they function and the usability of their interfaces.

Behind engine interfaces

- 9 The Google engine interface is often quoted as being an example of ergonomic excellence. Brigitte Simonnot notes that it is recreational on several levels, a means perhaps of winning over users, of creating a rapport. She also analyzes some of the games created by and around the famous engine. They are often representative of both the economic stakes involved – online referencing, which is the other face of commercial engines –, but also the sector's political and ethical issues. The main objective of traditional IR systems was to enable users to find relevant information that met their needs, however, when it comes to commercial engines, their purpose is more ambiguous: creators' marketing strategies are continuously being adapted to the habits of Internet users who, in turn, take ownership of the applications and use them in new ways. By facilitating the activity of information searching, even by making it fun, engine service providers create confusion as to their real objectives and work on appeal, probably to raise users' confidence in their applications and to develop customer loyalty¹⁰.
- 10 Certainly, engines focus on facilitating online information searches and help people to find their way around the wealth of content available. Nevertheless, Michael Zimmer speculates about the "Faustian bargain" that may represent the drive for the perfect search engine, whose aim is to provide access to the world's knowledge – through the widest possible coverage of Web content – and to attain enough recall capacity to predict user intentions (Jansen *et al.*, 2008).
- 11 The author underlines how threats to privacy increased when robust audiovisual data processing applications were made available to the general public both through the development of Web 2.0 – the Web as universal processing platform (O'Reilly, 2005) for data and social networks – and the interoperability of these technologies with search engines. He suggests a group of measures that should be put in place to protect and respect individual privacy. In fact, by seeking to continuously customize the results, engine designers are driven to systematically gather all data liable to improve application performance – data that can be further exploited using sophisticated data mining algorithms. The latter not only describe what Internet users do, but they can also supply a host of details on their core interests and ideas, which poses a real threat for privacy protection. Search engines are involved in a form of "economy of visibility" (Heilmann, 2007: 310), under the pretence of preventing not only public security issues, but predicting the needs, wishes and desires of information researchers online. This year, the French Data Protection Commission (CNIL) and its German counterpart celebrate their third anniversary. CNIL's primary mission is to inform citizens on their rights and obligations regarding the protection of freedoms and privacy, to ensure that application designers respect legislation and that everybody has the right to access personal computerized data. There is also a pressing need to inform citizens about the handling of their personal data online that is often collected and used without their knowledge. There is no doubt that application designers must make more of an effort to provide better information about the data gathered or produced, and make it easier for people to exercise their right to access and amend personal data, whilst also making provisions for a "right to be forgotten".

Conclusion

- 12 At a time when search engine designers are working towards making all information available at the click of a mouse, harbouring suspicions about these applications that are otherwise extremely useful should be avoided. Rather, knowing the limitations of these tools and deconstructing the myth that informational activity should be simple, easy and free from any charge, is a first step towards regulating practices in this sector. Equipping the French language with a more precise vocabulary in this scientific field is certainly a work in progress that needs further development. However, the translations of three articles submitted in English – those by Judit Bar-Ilan, Dirk Lewandowski and Nadine Höchstötter, and Michael Zimmer – have sought to remain as close as possible to the established vocabulary.
- 13 Furthermore, Human and Social sciences research also plays a part in carrying out more detailed analyses of the social consequences and ethical and cultural stakes of these applications. Certainly, processes are underway to rationalize and standardize online publication, particularly within the context of the Semantic Web. Nevertheless, there is still a lot of progress to be made in this sector, where, as is often the case, technological innovations precede any reflection about their consequences. Let us hope for more research in French to be published in this field – both as a strategic element to drive more French researchers to expand their research and for young researchers to become interested in these issues.

BIBLIOGRAPHY

Austin B., 2001, “Mooers’ Law: In and Out of Context”, *Journal of the American Society for Information Science and Technology*, 52 (8), pp. 607-609.

Bar-Ilan J., 2005, “Expectations versus reality - Search engine features needed for Web research at mid 2005”, *Cybermetrics*, 9 (1), <http://www.cindoc.csic.es/cybermetrics/articles/v9i1p2.html>, consulté le 25/08/08.

Chiaramella Y., Mulhem P., 2007, “La recherche d’information. De la documentation automatique à la recherche d’information en contexte”, *Document numérique*, 10(1), pp. 11-38.

CIBER, 2008, *Information behaviour of the researcher of the future*, Londres, UCL, <http://www.bl.uk/news/pdf/googlegen.pdf>, consulté le 26/08/08.

Heilmann É., “Surveiller (à distance) et prévenir. Vers une nouvelle économie de la visibilité”, *Questions de communication*, 11, pp. 303-322.

Ingwersen P., Järvelin K., 2005, *The Turn: Integration of Information Seeking and Retrieval in Context*, Dordrecht, Springer.

Jansen B.J., Booth D.L., Spink A., 2008, “Determining the informational, navigational, transactional intent of Web queries”, *Information Processing & Management*, 44(3), pp. 1251-1266.

Järvelin K., Ingwersen P., 2004, "Information seeking research needs extension towards tasks and technology", *Information Research*, 10(1), paper 212, <http://informationr.net/ir/10-1/paper212.html> (consulté le 22/08/08).

Jeanneney J.-N., 2005, *Quand Google défie l'Europe. Plaidoyer pour un sursaut*, Paris, Éd. Mille et une nuits.

Mesguich V., Thomas A., 2006, *Net recherche : le guide pratique pour mieux trouver l'information utile*, Paris, Éd. adbs.

O'Reilly T., 2005, "What is Web 2.0 - Design Patterns and Business Models for the next generation of software", *O'Reilly Network*, 30/09/05 <http://oreilly.com/web2/archive/what-is-web-20.html> (consulté le 27/09/06).

Robins D., Holmes J., 2008, "Aesthetics and credibility in web site design", *Information Processing & Management*, 44 (1), pp. 386-399.

Savolainen R., 1995, "Everyday life information seeking: Approaching information seeking in the context of "way of life", *Library & Information Science Research*, 17(3), pp. 259-294.

Thelwall M., 2008, "Quantitative Comparisons of Search Engines", *Journal of the American Society for Information Science and Technology*, 59 (11), pp. 1702-1710.

Tjondronegoro D., Spink A., 2008, "Web search engine multimedia functionality", *Information Processing & Management*, 44(1), pp. 340-35.

NOTES

1. Study carried out on a panel of over 2 million Internet users: "ComScore Releases June 2008 U.S. Search Engine Rankings", *Comscore Inc.*, <http://www.comscore.com/press/release.asp?press=2337> (accessed on 20/07/08).

2. See "comScore Releases May 2008 French Search Rankings", *Comscore Inc.*, 18/07/0 <http://www.comscore.com/press/release.asp?press=2327> (accessed on 20/07/08).

3. Data collected from a sample of 8,500 Internet users aged 2 years and over. See "L'audience de l'internet en France en mai 2008", [Internet audience in France in May 2008] *Mediamétrie/Netratings*, 30/06/08, www.mediametrie.fr/resultats.php?rubrique=net&resultat_id=557 (accessed on 20/07/08).

4. See, for example, (Tjondronegoro, Spink, 2008)

5. See V. Mesguich and A. Thomas (2006) for an overview.

6. Following the example of Canadians, who use the terms "*fouineurs*" [snoopers] or "*foreurs*" [rummagers] for web browsers, we could perhaps use the expression "*prospecteurs Web*" [Web prospectors] to describe these tools in order to better account for their real purpose.

7. See the definitions provided by the online French dictionary, *le Trésor de la Langue française informatisé* (TLFI), www.cnrtl.fr/definition/moteur

8. This contribution was received in March 2007 and the final version was accepted in November 2007.

9. For example, if exact statistics were made available on the links to a given site, any good mathematician could easily carry out a proxy search to deduce the weighting given to these links in page ranking calculations, which would enable a part of the algorithm arcana to be discovered.

10. Some studies show a correlation between the physical appearance of sites and their credibility with Internet users (Robins, Holmes, 2008).

AUTHORS

BRIGITTE SIMONNOT