

## New perspectives on Web search engine research

Dirk Lewandowski

Hamburg University of Applied Sciences, Germany

*This is a preprint of a book chapter to be published in*

*Lewandowski, Dirk (ed.): Web Search Engine Research. Bingley: Emerald Group Publishing, 2012*

<http://books.emeraldinsight.com/display.asp?K=9781780526362>

### Abstract

**Purpose** – The purpose of this chapter is to give an overview of the context of Web search and search engine-related research, as well as to introduce the reader to the sections and chapters of the book.

**Methodology/approach** – We review literature dealing with various aspects of search engines, with special emphasis on emerging areas of Web searching, search engine evaluation going beyond traditional methods, and new perspectives on Web searching.

**Findings** – The approaches to studying Web search engines are manifold. Given the importance of Web search engines for knowledge acquisition, research from different perspectives needs to be integrated into a more cohesive perspective.

**Research limitations/implications** – The chapter suggests a basis for research in the field and also introduces further research directions.

**Originality/value of paper** – The chapter gives a concise overview of the topics dealt within the book and also shows directions for researchers interested in Web search engines.

**Paper type** – Literature review

For most users, Web search engines are the central starting point for their exploration of Web content. Search engines lead us to new websites we have never heard of, help us re-encounter familiar websites and offer us a wide variety of content from the many sources of the Web, which we would not be able to discover with other tools. Most users use search engines every day, and the amount of queries entered into general-purpose Web search engines such as Google worldwide exceeds 100 billion queries per month (ComScore, 2009). Even though most users use search engines every day, they know very little about them (cf. Hendry & Efthimiadis, 2008).

Also, research on Web search engines and their impact is still in its infancy. While technical development is fast, and lots of research is published in that area, with regard to gaining a deeper understanding of the user, the searching process, and the societal impact of search engines (not to mention the combination of these), there is still only limited understanding. This book brings together researchers from different fields and

aims to stimulate research looking beyond the obvious research questions and methods of one's own discipline.

This introduction to the book is divided into two parts. The first part deals with the current state of Web search, and how the emerging field of Web search engine research—or Web search studies, or whatever the best label might be—is defined by researchers across disciplines. The aim thereby is not to give a complete literature review, but to show fruitful areas for research, especially in the Library and Information Science (LIS) field.

The second part then introduces the chapters of the book, which are grouped into three sections: emerging areas of Web searching; beyond traditional search engine evaluation; and new approaches to Web searching. The concluding section gives some suggestions for further research.

## The context of Web search engine research

### The Search engine market

When discussing Web search engines, in most cases one arrives quickly at a discussion of Google. In fact, Google is often seen as synonymous with Web search. However, the search engine market is richer than it might seem at first look. Smaller companies are active, even though they usually focus on niche markets or business applications. A major reason for this is that while search may be highly profitable for smaller companies in these specialised areas of search, the high costs of building and maintaining a search engine on the scale of the Web lead to a concentration on the search engine market, with just a few major players left (Buganza & Della Valle, 2010; for a historical perspective reaching back to 2000, see also the Search Engine Relationship Chart Histogram, Clay, 2011a).

It may be irritating to see that many search engines claiming to search the “whole of the Web” are available on the market; however, only a few of them have their own, Web-scale index. Outside of these few, most search engines license search results from other search engines, the most famous example being Yahoo using results from Microsoft's Bing search engine (Microsoft, 2009; also see the Search Engine Relationship Chart, Clay, 2011b).

Another point to consider is the market shares of the different search engines. While there may be at least a small variety of Web search engines, users' acceptance of these choices greatly differs among them. In the U.S., we can see that while Google dominates with a share of 65 percent (Sterling, 2011), as measured in the relative number of queries entered into this search engine, and that the Bing/Yahoo alliance follows with a considerable share of 31 percent, the market in most European countries is much more concentrated (Lunapark, 2011). In most countries, Google has a market share of around 90 percent.

When discussing the search engine market, it is often forgotten that while search engines are surely commercial enterprises, they also serve as facilitators of information, and therefore, that they serve the interests of the public (see Zimmer, 2010; van Couvering, 2008). When considering that mainly one search engine is used, one has to ask whether this one search engine does indeed serve these interests. While some researchers would agree with Peter Jacsó that “in the ideal world one perfect search engine would suffice” (Jacsó, 2008, p. 864), others argue for a plurality of search engines

to best serve users' interests (Zimmer, 2010; van Couvering, 2007). To agree with the former, one would have to assume that a user would be allowed to specify how the rankings of that one search engine should be produced. While it may be possible to give users tailor-made rankings through personalisation techniques, this tactic would not be transparent and therefore allow the search engine provider too much power over its users.

### **Challenges to information retrieval and the Library and Information Science research communities**

Web search engines are nowadays researched in many different disciplines, ranging from computer science to the humanities. The two research communities that were concerned with searching long before Web search engines emerged were the Information Retrieval (IR) community, and the Library and Information Science (LIS) communities. While information retrieval is both based on Computer Science and on LIS, the two disciplines have a distinct view on the topic, IR being more oriented towards technical developments and system-centred evaluation, while LIS is more focussed on user aspects and user-centred evaluation. With Web search engines, both communities are challenged, in that (1) other communities become more and more interested in search engine studies, (2) it becomes clear that only a deeper understanding of Web searching will suffice, which requires a combination of methods from different disciplines, and (3) the social impact of Web search engines, which is only sometimes the focus of both disciplines, is an important area to consider.

But even on a technical level, Web search engines cannot be treated as just another kind of information retrieval system. Lewandowski (2005, p. 140) divided the differences between "classic" IR and Web IR into four distinct areas: documents, Web characteristics, user behaviour, and IR systems. An important aspect here is the nature of queries entered into search engines: Queries are generally very short (2-3 words; see Jansen & Spink, 2006; Höchstötter & Koch, 2009) and the systems are designed to answer such short—and therefore usually very general—queries. This leads to search engines' focus on high-precision documents, while in traditional IR, a balance between a complete set of results and precise results must be found. Directly connected with user behaviour is the design of the search engines' user interfaces. Again, a "one size fits all" approach has to be followed. Interfaces must be very easy to understand and therefore cannot allow for complex interactions while building a query or viewing the results.

The challenges search engines pose to library and information practice are obvious: Users who are used to the comfort and fast response of Web search engines expect other information systems to deliver the same performance. It is not uncommon that patrons compare information systems to Web search engines, and state that where Google is able to deliver valuable results in an instant, another searching system should also be able to do so. On the other hand, search engines usually offer only limited search functions and do not allow for complex queries, a fact that makes it difficult for the information professional to build precise and complex queries.

### **Approaches to classifying Web search engine research areas**

Research on Web search engines reaches in scope from technical developments to studies on search engine quality, from investigations on the social impact of the Web search engine to approaches to using data from Web search engines for analytic approaches (e.g., Thelwall, 2004; Ginsberg et al., 2009).

It is difficult to define the field of “Web search engine research”, as most researchers see themselves more as part of a discipline-based research community (such as Information Science, Human-Computer interaction, Sociology, and so on) than as part of a topic-based, interdisciplinary research community. However, similar to the wider area of Web Science (Berners-Lee, Hall, J. A. Hendler, et al., 2006; Berners-Lee, Hall, J. Hendler, & Weitzner, 2006), where the Web should be researched in a multidisciplinary manner, we see search engine research as a multidisciplinary research area, and as an important part of Web Science, as well (Lewandowski, 2008a). Web search engine research (or “Web search studies”, as Michael Zimmer named the discipline) can be seen as a “meta-discipline” investigating search engines from different perspectives (Zimmer, 2010, p. 508). However, the question remains of which parts would constitute such a meta-discipline. Researchers from different fields have proposed frameworks for Web search engine research, taking different perspectives into account.

Bar-Ilan (2004) gives an overview of the different research areas of interest for Information Science, divided into the two main sections of (1) understanding the Web’s structure and processes, and (2) on the other hand of understanding users’ needs and behaviours. In this book, I will argue that only an integrated approach combining the two areas will lead to better understanding of the quality of Web search engines.

Machill, Beiler, and Zenker (2008) find “five topic fields considered to be central to future search-engine research from an interdisciplinary perspective” (p. 592). These are (1) search-engine policy and regulation, (2) search-engine economics, (3) search engines and journalism, (4) search-engine technology and quality, and (5) user behaviour and competence (p. 592).

Lewandowski (2008a) also differentiates between five sub-fields, but with a different angle: (1) information retrieval technology, (2) search engine quality, (3) information research, (4) user behaviour and user guidance, (5) and search engine economics.

Riemer and Brüggemann (2009, S. 116f.) see search engine research at the crossroads between the design-science paradigm and the behavioural-science paradigm. An integrated approach would consider both, and this would lead to a better understanding of existing systems and to the design of better systems in the future.

Zimmer (2010) sees Web search studies “centered around a nucleus of major research on web search engines from five key perspectives: technical foundations and evaluations; transaction log analyses; user studies; political, ethical, and cultural critiques; and legal and policy analyses” (p. 508), and finds that the following areas deserve particular attention: search engine bias, search engines as gatekeepers of information, values and ethics of search engines, framing the legal constraints and obligations (pp. 516-517).

In general, we found that many researchers dealing with Web search engines complain that Web search engine research is much too focused on technical aspects and that a wider perspective is needed. Hargittai (2007) stresses that especially research dealing with search engines’ impact on society is largely missing: “Despite their central role in how people access information, however, little social science work has focused on the non-technical dimensions of search engine tools, the companies that run them, or the practices of the users who rely on them” (p. 769). A conclusion from Spink and Zimmer (2008) goes in the same direction: “Until recently, most scholarly research on Web search engines have been technical studies originating from computer science and related disciplines” (p. 343).

So, while a large part of search-engine-related research is still on technical aspect, we now see a wider interest in the topic from researchers originating from different fields. This could lead to fruitful cooperation, and the combination of technical knowledge with methods and findings from the social sciences in particular could lead to a deeper understanding of Web search engines.

## Book outline

This book brings together researchers from various fields, ranging from Computer Science to Ethnography. Accordingly, the studies presented in the book are based on very different methods. We hope that especially readers more at home in the IR-related fields and familiar with system-centred retrieval effectiveness measures can benefit from the studies where user-centred, qualitative approaches are applied, and vice versa. The book is divided into three parts, and the following sections give an overview of what to expect from the individual chapters and from the book as a whole.

### Part 1: Emerging areas of Web searching

Part 1 of the book is devoted to emerging areas of Web search. The chapters give broad overviews of these areas. Researchers can benefit from these reviews, as they define the fields for research in emerging areas.

The first chapter is “The Many Ways of Searching the Web Together: A Comparison of Social Search Engines”, by Manuel Burghardt, Markus Heckner, and Christian Wolff. In recent years, a lot of interest has been generated by the rise of social media, which also led to search engines exploiting social data to improve rankings for individual users. However, as Burghardt, Heckner and Wolff show, the concept of social search is not limited to traditional search engines improving their rankings, but is instead multi-faceted. They present a taxonomy of social search, which first differentiates between people-powered search and social data mining—the former exploiting (either explicitly or implicitly) data generated by users, and the latter referring to search within social media or people search.

Regarding people-powered search, the authors explore the areas of social tagging, social question answering, collaborative search, collaborative filtering, personalized social search engines, the exploitation of click popularity and usage data, and the exploitation of the link topology of the Web, as well. The authors review all of these areas thoroughly and show that social information retrieval is much more than just searching on (or integrating data from) the well-known social networks. However, this review of social search also shows that we are far from having one central access-point to the Web (a search engine such as Google) that allows for searching all of the content available. Quite the contrary: The fact of social media networks not making their data available for indexing by general-purpose Web search engines leads to a situation where a user has to use different kinds of research tools to get a complete picture.

Another area that generated a lot of interest, is map-based search engines (e.g., Google Maps), also called local (Web) search engines. Their results are also included in the search engine results pages (SERPs) of the general-purpose Web search engines. The

chapter “Local Web Search Examined”, by Dirk Ahlers, deals with the concept of local search, its potentials and its challenges. Also, the major players in the field of local Web search are reviewed, and trends in the field are examined.

This author makes it clear that today’s map-based search engines have their foundations in earlier Geographic Information Retrieval (GIR) technologies, and that information needs expressed in these systems quite differ from the ones served by general-purpose Web search engines. Therefore, we need a deeper understanding of users’ intents towards map-based search engines. The single type of query accepted by local Web search engines today is limited to searching for a concept at a certain location (“Hotel Berlin”), while future systems should be able to richly interpret the geo-location and make new views of the already available data possible. Ahlers gives the example of a search for “a camping site near a river”. The data to answer such a query is already available today, as the concept “camping site” and the rivers are already included in map data. However, the spatial data included in the maps is not yet fully exploited. Also, users’ interactions with local Web search engines are not yet taken into account, even though data on the searching behaviour of users could greatly help improve the search engines, amongst other things through giving recommendations based on users’ location trails (Zheng, Zhang, & Xie, 2009).

Web search engines have not only been the object of research, but it also became clear that using their data is valuable for answering a variety of research questions (cf. Goel, Hofman, Lahaie, Pennock, & Watts, 2010). An important area of research is the analysis of query data (i.e., exploiting the large numbers of queries entered into a search engine to identify trends). Since 2006, Google has offered a free tool that allows for easily analysing search volumes (trends.google.com). All a user has to do is to enter one or more queries and select a time-span. The result is a graph showing the search volumes over time, even though only relative data is given, not exact numbers. There are already studies using search query statistics instead of traditional approaches to collecting data for forecasting (e.g., Ginsberg et al., 2009; Choi & Varian, 2009; Goel et al., 2010).

In his chapter, “The Computational Analysis of Web Search Statistics in the Intelligent Framework Supporting Decision Making”, Wiesław Pietruszkiewicz discusses possibilities and practical applications of query data for forecasting. The advantages of using search queries lie, apart from the low cost in collecting such data, in the amount of data building up the so-called database of intentions (Batelle, 2005), which allows for examining user intent not only with reference to popular topics, but in great depth. Also, the data allows for precise and accurate behavioural observations, and the analysis of search data can be used in many fields. Using examples from the field of economics, Pietruszkiewicz details the process of collecting and analysing search volume data.

However, it should also be mentioned that such an approach is not flawless.

Pietruszkiewicz discusses these flaws, using a variety of examples and also offering tips for reliable data collection.

## **Part 2: Beyond traditional search engine evaluation**

The chapters in the second section of the book deal with a variety of aspects concerning the evaluation of Web search engines. While evaluation has always been an integral part of information retrieval (IR) research (Robertson, 2008), traditional evaluation methods are challenged by the behaviour of Web search engine users, who differ greatly from the assumed user of traditional information retrieval systems, and by the properties of the

databases underlying the Web search engines. Here, issues of trust and reliability in the search results are of great importance.

In their chapter on “Evaluating Web Retrieval Effectiveness”, Ben Carterette, Evangelos Kanoulas, and Emine Yilmaz give an overview of retrieval effectiveness measures. They first review traditional measures, and then focus on measures developed in recent years. The authors claim that the main change in this topic is that older retrieval measures are not based on an explicit user model, but

they nevertheless imply a user model: a user will look at and derive utility from the full set of retrieved documents. Every relevant document is of equal value. Having more is better than having fewer, but only as long as the precision does not drop to unacceptably low levels.

Regarding user behaviour in Web search engines (cf. Machill, Neuberger, Schweiger, & Wirth, 2004; Jansen & Spink, 2006), it is obvious that such basic assumptions do not hold true, at least not in this particular case. The newer models reviewed by Carterette, Kanoulas and Yilmaz take into account typical user behaviour, but, as the authors note, still “The ‘users’ are highly simplified mathematical objects with no will or motivation of their own, and no ability to provide useful feedback that might inform future research directions”.

While retrieval effectiveness studies ask for the relevance of search results, other aspects of the results set can also be of importance to a searcher. While the concept of *diversity* is discussed briefly in the context of retrieval effectiveness tests in Carterette, Kanoulas and Yilmaz’s chapter, Kerstin Denecke devotes her chapter entirely to “Diversity-Aware Search: New Possibilities and Challenges for Web Search”.

Based on the definition of *diversity* by van Cuilenburg (2000), who writes that “diversity is the co-existence of contradictory opinions and/or statements (some typically non-factual or referring to opposing beliefs/opinions)”, Denecke gives a detailed overview on the concept and its applications in search.

Diversity in search results is a multi-faceted concept. Giunchiglia et al. (2009) define the following dimensions of diversity: diversity of sources (multiplicity of sources of texts and images); diversity of resources (e.g., images, text); diversity of topic; diversity of viewpoint; diversity of genre (e.g., blogs, news, comments); diversity of language; geographical/spatial diversity; and temporal diversity.

From the popular Web search engines, one can already see that the presentation of results on the search engine results pages (SERPs) has become more complex and diverse in recent years (Höchstötter & Lewandowski, 2009). This mainly concerns diversity of sources, diversity of resources, and diversity of genre. However, content-based diversity, such as the diversity of viewpoint, is not yet implemented, although it could be a valuable addition, if a user can clearly see how and why certain results are produced.

Denecke discusses the current diversification of results in the popular Web search engines, even as she shows the existing approaches to diversity and examines the presentation methods for representing diversity on the SERPs. She also discusses an exemplary application, a diversity-aware search engine for medical content (Denecke, 2009).

For future research, Denecke sees a focus on making the various dimensions of diversity accessible in the search results. Also, she sees the need for integrating diversity measures into the search engine evaluation methods. And finally, she holds that



diversity is not only important in textual Web search, but also in other areas, such as image search.

While search engine evaluation and measures try to measure aspects of usefulness of search engines for all users, or at least for a certain user group, Li, Wang, and Yu stress that the usefulness of a search engine for an individual user depends on the needs and wishes of that very user. In their chapter “Personalised Search Engine Evaluation: Methodologies and Metrics”, they develop a taxonomy of indicators for measuring the quality of a search engine. A user can give each indicator an individual weight, so that the evaluation results are adapted to his or her individual preferences. The model presented does take a considerable variety of aspects into consideration. It is therefore related to approaches aiming at more complex models for measuring Web search engine quality, such as Balatsoukas, Morris, and O’Brien (2009), Lewandowski and Höchstötter (2008), Zhu (2011), and Petter, DeLone, and McLean (2008). As the model comprises seventy features, it allows for detailed specifications. Among them are freshness measures, which are visualised in histograms, so that the user can easily compare them.

Some search engine evaluation studies (e.g., Bar-Ilan, 2005; Bar-Ilan, Mat-Hassan, & Levene, 2006) tested search engines through comparing their ranked results lists. The idea is that results are not independent of one another, but that the results *sets* produced by an engine determine its usefulness. Another factor to be considered is that when deciding upon using an additional search engine, or even a new search engine, it is important to the user whether this engine shows different results on the first positions. To measure this, one can apply rank correlations. With that regard, Massimo Melucci, in his chapter “Search Engines and Rank Correlation”, reviews the literature on rank correlations and shows the usefulness of the concept for conducting search engine studies. In this context, rank correlations are applicable to a variety of purposes:

To compare the rankings observed during an experiment with the rankings produced by (i) a competitor engine, (ii) the same engine but with different parameters or (iii) the engine which correctly ranks all the items (e.g. a human) and is then considered the best.

A major merit of Melucci’s chapter is that he introduces findings and measures from the statistics literature and shows how they can be applied in search engine research.

### Part 3: New perspectives on Web searching

The third part of the book comprises chapters that are dealing with search in a wider context and that expand the view from the traditional information retrieval disciplines to that of ethnography, psychology, and philosophy.

In recent years, it has become obvious that search would not continue to encompass only a user entering a query and then selecting results from a ranked list (cf. White & Roth, 2009). Since then, new approaches to interacting with Web content through search have been introduced (Schraefel, 2009).

The first chapter in this section, “Beyond Search: A Technology Probe Investigation”, by Erin Bryant, Richard Harper and Philip Gosset, introduces two new approaches—called *Cards* and *Pebbles*—to exploring the Web’s information. *Cards* show results as a card with a picture and some text, while *Pebbles* is built around the idea of a user “travelling the Web”. The basic idea of both probes is to go beyond query-based information retrieval and develop new metaphors that go beyond search yet still use search engine technology as their underlying basis. In the present case, data from Microsoft’s Bing



search engine was used, but the user experience is completely different from Bing's more traditional approach to search.

For evaluating the new tools, Bryant, Harper and Gosset conducted a study where households were given the probes to play with, and then were asked about their experiences. The study shows how valuable results can be achieved concerning a search system, going beyond results that can be achieved in retrieval tests or even in lab settings. Therefore, the uses of Bryant, Harper and Gosset's chapter are two-fold: On the one hand, we learn about two new metaphors for exploring Web content; on the other hand, we learn about methods for studying users that may not be familiar to most of the researchers in the IR/Information Science domain. One value of such a study design that must not be underestimated is that it can be used to generate new ideas; or, as the authors themselves say, "it became clear that the probes had successfully elicited some ideas and aspirations about how to engage with the web on the part of the participants who pointed towards new possibilities".

Due to the great variety of the quality of the Web's content and the low barriers of search engines for including content in their indices, the user is confronted with content of mixed quality, even though search engines try to determine the quality of individual web pages through formal criteria (cf. Lewandowski, 2008b), such as the number and quality of the links pointing to that page. A user has to select relevant and credible pages based on the information presented on the search engine results pages. As Yvonne Kammerer and Peter Gerjets show in their chapter titled "How Search Engine Users Evaluate and Select Web Search Results: The Impact of the Search Engine Interface on Credibility Assessments", this selection behaviour is heavily influenced by the position of a certain result within the ranked list. Additionally, search engines do not provide users with enough information on the (assumed) credibility of the results presented. Therefore, the credibility of the results cannot be adequately evaluated at this stage, but a user has to examine the result itself directly to make a judgement. Even so, aggregated information on the credibility of the result is not available, and the user is left to his own devices and has to apply his own criteria. New interfaces try to help the user to evaluate the credibility of the results that already appear on search engine results pages.

The chapter concluding the book, "What Would Kant Think? Testing Truth claims in Research Traditions, and Proposing Deeper Meanings for the Concept of 'Search'", by Denise N. Rall, introduces philosophical concepts to the area of Web search. The chapter deals with truth claims, where a truth claim should be understood as a claim that "examines the *relationship* between the type of question or inquiry that researchers ask, and the evidence found in response to that inquiry". Discussing the differing truth claims in science, social science, law and in judgements of excellence, Rall gives an overview of different approaches to claiming truth. Considering search engine results, an analysis of the truth claims presented could be used to improve the quality of the results. Again, it should be stressed that formal quality measurements such as exploiting the link structure of the Web are not sufficient to determine whether results are reliable or even truthful.

Another point Rall makes is that search engines assert the appropriateness of a result through its presence in the search engine's index or through its assignment of a good position in the ranked results list. Rall draws a comparison to the art world: "Like viewers in Danto's artworld [where "an artwork is merely something indexed in accord with artworld practices of indexing"], the searchers in webworld follow a similarly self-reflexive path that accepts any link as result by its ontological presence, and as a non-result (of course) by its absence".

One may be at first confused about the connections between such differing fields as Information Retrieval and Philosophy or the Arts, but Rall's text will be inspiring also for researchers usually more concerned with technical or more hands-on user issues.

## Suggestions for further research

All individual chapter authors offer suggestions for further research at the closing of their respective contributions. These suggestions should not be repeated here. Instead, two points should be stressed in this concluding section: (1) Web search engine research should be multi-disciplinary in nature, and (2) to gain a better understanding of users' interactions with Web search engines, search engine providers should make more such data available to the research community.

From the outline given above, one can see that research on Web search engines involves far more than developing new features or using traditional measures to evaluate their quality. Web search engines raise a multitude of questions, some of which are answered by the authors in this book. However, it is clear that Web search engine research is still in its infancy, but that building up on the richness of approaches and methods from various disciplines could lead to a thorough understanding of Web search engines, not only from a technical perspective, but also from a societal point of view. Recent discussions on search neutrality (cf. Edelman & Lockwood, 2011; Edelman, 2010; Granka, 2010), the investigation led by the European Commission on the market power (and its abuse) by Google (Commission, 2010), and discussions on users' privacy while they use search engines (cf. Poritz, 2007; Weber, 2009) have shown that Web search engine research has to consider much more than technical developments. As Web searching is, next to e-mail, the most-used activity on the internet (Purcell, 2011; Eimeren & Frees, 2011) and billions of queries are entered into search engines every day (ComScore, 2009), we should be aware that every search engine results page and every result clicked influences what users get to see and the way in which we, as a society, organize knowledge (Höchstötter & Lewandowski, 2009).

Some of the chapters in this book are the result of collaborations between researchers from academia and industry. Such collaborations are usually fruitful, as the different perspectives on Web searching complement each other. When the behaviour of real users must be researched using mass data (usually transaction-log data), there is no way around collaboration with a live search engine. However, it is often difficult to obtain such data from search engine providers. Part of the reason for that lies in privacy aspects, part of it in bad experiences in the past with making such data publicly available, and part of it simply in keeping business secrets. However, search engine providers would benefit from reconsidering these concerns and making cleared data sets available. This could leverage Web search engine research, foremost for researchers conducting studies on a smaller scale, who could broaden their studies and verify their results through the additional data.

## Acknowledgements

First and foremost, I would like to thank the chapter authors for their contributions, as well as the book series editor, Amanda Spink, for giving me the opportunity to edit this

book. I am also grateful to the chapter reviewers, especially to Friederike Kerkmann, for her suggestions for improving the chapters presented in this book.

## References

- Balatsoukas, P., Morris, A., & O'Brien, A. (2009). An evaluation framework of user interaction with metadata surrogates. *Journal of Information Science*, 35(3), 321-339.
- Bar-Ilan, J. (2004). The use of web search engines in information science research. In B. Cronin (Ed.), *Annual review of information science and technology* (Vol. 38, pp. 231-288). Medford, NJ: Information Today, Inc.
- Berners-Lee, T., Hall, W., Hendler, J. A., O'Hara, K., Shadbolt, N., & Weitzner, D. J. (2006). A framework for web science. *Foundations and Trends in Web Science*, 1(1), 1-130. Hanover, Mass.: Now Publishers Inc.
- Berners-Lee, T., Hall, W., Hendler, J., & Weitzner, D. J. (2006). Creating a science of the web. *Science*, 313(5788), 769-771.
- Buganza, T., & Della Valle, E. (2010). The search engine industry. In S. Ceri & M. Brambilla (Eds.), *Search computing: Challenges and directions* (pp. 45-71). Berlin, Heidelberg: Springer.
- Choi, H., & Varian, H. (2009). *Predicting initial claims for unemployment benefits*. Retrieved from [http://static.googleusercontent.com/external\\_content/untrusted\\_dlcp/research.google.com/en//archive/papers/initialclaimsUS.pdf](http://static.googleusercontent.com/external_content/untrusted_dlcp/research.google.com/en//archive/papers/initialclaimsUS.pdf)
- Clay, B. (2011a). Search engine relationship chart histogram. Retrieved from [http://www.bruceclay.com/serc\\_histogram/histogram.htm](http://www.bruceclay.com/serc_histogram/histogram.htm)
- Clay, B. (2011b). Search engine relationship chart. Retrieved from <http://www.bruceclay.com/searchenginechart.pdf>
- ComScore. (2009). Global search market draws more than 100 billion searches per month - comScore, Inc. Retrieved September 26, 2011, from [http://www.comscore.com/Press\\_Events/Press\\_Releases/2009/8/Global\\_Search\\_Market\\_Draws\\_More\\_than\\_100\\_Billion\\_Searches\\_per\\_Month](http://www.comscore.com/Press_Events/Press_Releases/2009/8/Global_Search_Market_Draws_More_than_100_Billion_Searches_per_Month)
- van Couvering, E. (2007). Is relevance relevant? Market, science, and war: Discourses of search engine quality. *Journal of Computer-Mediated Communication*, 12(3), 866-887.
- van Cuilenburg, J. (2000). On measuring media competition and media diversity. Concepts, theories and methods. In R. G. Picard (Ed.), *Measuring media content, quality and diversity. Approaches and issues in content research* (pp. 51-84). Turku: Turku School of Economics.
- Denecke, K. (2009). Assessing content diversity in medical weblogs. *Proceedings of the First International Workshop on Living Web at the 8th International Semantic Web Conference (ISWC)*. Retrieved from <http://livingknowledge.europarchive.org/images/publications/LivingWeb.pdf>
- Eimeren, B. V., & Frees, B. (2011). Drei von vier Deutschen im Netz – ein Ende des digitalen Grabens in Sicht? *Media Perspektiven*, (7-8), 334-349.
- Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, 457(7232), 1012-1014.
- Giunchiglia, F., Maltese, V., Madalli, D., Baldry, A., Wallner, C., Lewis, P., Denecke, K., Skoutas, D., and Marenzi, I. (2009). *Foundations for the representation of diversity, evolution, opinion and bias*. Retrieved from <http://eprints.biblio.unitn.it/archive/00001758/01/063.pdf>
- Goel, S., Hofman, J. M., Lahaie, S., Pennock, D. M., & Watts, D. J. (2010). Predicting consumer behavior with web search. *Proceedings of the National Academy of Sciences of the United States of America*, 107(41), 17486-90.

- Hargittai, E. (2007). The social, political, economic, and cultural dimensions of search engines: An introduction. *Journal of Computer-Mediated Communication*, 12(3), 769-777.
- Hendry, D., & Efthimiadis, E. (2008). Conceptual models for search engines. In Amanda Spink & M. Zimmer (Eds.), *Web searching: Interdisciplinary perspectives* (pp. 277-308). Berlin: Springer.
- Höchstötter, N., & Koch, M. (2009). Standard parameters for searching behaviour in search engines and their empirical evaluation. *Journal of Information Science*, 35(1), 45.
- Höchstötter, N., & Lewandowski, D. (2009). What users see – Structures in search engine results pages. *Information Sciences*, 179(12), 1796-1812.
- Jacsó, P. (2008). How many web-wide search engines do we need? *Online Information Review*, 32(6), 860-865.
- Jansen, B. J., & Spink, A. (2006). How are we searching the World Wide Web? A comparison of nine search engine transaction logs. *Information Processing & Management*, 42(1), 248–263. Elsevier.
- Lewandowski, D. (2005). Web searching, search engines and information retrieval. *Information Services & Use*, 25, 137-147.
- Lewandowski, D. (2008). Suchmaschinenforschung im Kontext einer zukünftigen Webwissenschaft. In K. Scherfer (Ed.), *Webwissenschaft – Eine Einführung* (pp. 268-282). Münster: Lit.
- 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. Lunapark. Retrieved from <http://www.luna-park.de/home/internet-fakten/suchmaschinen-marktanteile.html>
- Machill, M., Beiler, M., & Zenker, M. (2008). Search-engine research: A European-American overview and systematization of an interdisciplinary and international research field. *Media, Culture & Society*, 30(5), 591-608.
- Microsoft. (2009). Microsoft, Yahoo! Change search landscape. Retrieved September 26, 2011, from <http://www.microsoft.com/presspass/press/2009/jul09/07-29release.mspx>
- Petter, S., DeLone, W., & McLean, E. (2008). Measuring information systems success: models, dimensions, measures, and interrelationships. *European Journal of Information Systems*, 17(3), 236-263.
- Purcell, K. (2011). *Search and email still top the list of most popular online activities*. [http://www.pewinternet.org/~media/Files/Reports/2011/PIP\\_Search-and-Email.pdf](http://www.pewinternet.org/~media/Files/Reports/2011/PIP_Search-and-Email.pdf)
- Riemer, K., & Brüggemann, F. (2009). Personalisierung der Internetsuche - Lösungstechniken und Marktüberblick. In D. Lewandowski (Ed.), *Handbuch Internet-Suchmaschinen* (pp. 148-171). Heidelberg: Akademische Verlagsgesellschaft Aka.
- Schraefel, M. C. (2009). Building knowledge: What's beyond keyword search? *Computer*, 42(3), 52-59.
- Spink, A., & Zimmer, M. (2008). Conclusions and future research. In A. Spink & M. Zimmer (Eds.), *Web search: Multidisciplinary perspectives* (pp. 343-347). Dordrecht: Springer.
- Sterling, G. (2011). Google search share plateaus, BingHoo gains, AOL drops. *Search Engine Land*. Retrieved September 26, 2011, from <http://searchengineland.com/google-search-share-plateaus-binghoo-gains-aol-drops-92714>
- Thelwall, M. (2004). *Link analysis: An information science approach*. Library and information science. Amsterdam: Academic Press.
- Zheng, Y., Zhang, L., & Xie, X. (2009). Mining interesting locations and travel sequences from GPS trajectories. *Proceedings of the 18th World Wide Web Conference* (p. 791). New York: ACM Press.

- Zhu, Q. (2011). Using a Delphi method and the Analytic Hierarchy Process to evaluate the search engines: A case study on Chinese search engines. *Online Information Review*, 35 [in press].
- Zimmer, M. (2010). Web search studies: Multidisciplinary perspectives on web search engines. In J. Hunsinger, L. Klastrup, & M. Allen (Eds.), *International Handbook of Internet Research* (pp. 507-521). Dordrecht: Springer.