



JD
60,3

Searching for people on Web search engines

Amanda Spink and Bernard J. Jansen

School of Information Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

Jan Pedersen

Overture Web Search Division, Palo Alto, California, USA

266

Received 20 October 2003
Revised 21 November 2003
Accepted 23 November 2003

Keywords *Information retrieval, Information searches, Worldwide web*

Abstract *The Web is a communication and information technology that is often used for the distribution and retrieval of personal information. Many people and organizations mount Web sites containing large amounts of information on individuals, particularly about celebrities. However, limited studies have examined how people search for information on other people, using personal names, via Web search engines. Explores the nature of personal name searching on Web search engines. The specific research questions addressed in the study are: "Do personal names form a major part of queries to Web search engines?"; "What are the characteristics of personal name Web searching?"; and "How effective is personal name Web searching?". Random samples of queries from two Web search engines were analyzed. The findings show that: personal name searching is a common but not a major part of Web searching with few people seeking information on celebrities via Web search engines; few personal name queries include double quotations or additional identifying terms; and name searches on Alta Vista included more advanced search features relative to those on AlltheWeb.com. Discusses the implications of the findings for Web searching and search engines, and further research.*

Introduction

People search the Web for many things, including shopping, sex and information on a near infinite array of topics. Major focuses of people's lives are other people and communicating with other people. The Web now provides another source of information about people from every country of the world. Ordinary people and celebrities can use the Internet to distribute and seek information. For example, some Web services (e.g. www.classmates.com) provide a way of connecting with your former high school classmates. People also use list serves, chat rooms and e-mail to communicate constantly with each other (e.g. Yahoo! groups). Seeking information on other people is also a key activity in the business and intelligence world (Ayers, 2000; Lane, 1996).

Owing to the growth of information about people on the Web, personal information privacy is now a key issue for individuals, organizations and governments (Federal Trade Commission, 2002). Web users often provide personal information through the Web that is processed and sold. Sholtz (2000) provides an overview of the privacy,

The authors thank the anonymous reviewers for their helpful comments, and AlltheWeb.com and Alta Vista.com for providing the Web query data sets without which they could not have conducted this research.



property rights, and economic issues associated with personal information on the Web. He concludes that personal information has become the new currency of e-commerce.

Recent studies of public Web searching have provided preliminary data on the degree to which people use Web search engines to seek information about other people. Spink *et al.* (2002a) studied Web search queries from 1997 to 2001. They found that despite some high frequency terms (e.g. free, sex, games, weather, maps, etc.), an unusually large number of terms are not repeated or used with low frequency. They suggest that these low frequency terms may include personal names, spelling errors, non-English terms, or Web specific terms (e.g. URLs).

Research also shows that few Web queries include advanced search features and Boolean operators, such as double quotations, e.g. "John Smith", that one may assume are required to effectively process a personal name query (Jansen *et al.*, 2000; Spink *et al.*, 2002a).

This paper provides results from a study exploring how people look for information on other people using Web search engines. We sought to examine if personal names form a major part of queries submitted to Web search engines and the characteristics of personal name queries. In the next section of the paper, we discuss related studies in personal name searching on information systems.

Related studies

Name recognition, matching and searching is an important research area of information retrieval (IR) and Web studies (Carroll, 1985; Hermansen, 1985; Pfeiffer *et al.*, 1996). Due to the interest in people searching by many businesses, such as headhunting companies, detective companies, and the intelligence community, researchers have developed and tested name finding IR systems (Hayes, 1994; Navarro *et al.*, 2003). Borgman and Seigfried (1992) categorized various name-searching applications, as have commercial organizations (Search Software America, 2003). Thomson and Dozier (1997) studied name searching in various domains, including legal and news documents, and suggested the need for more effective name searching and matching techniques.

User studies have specifically explored the human aspects of name searching on IR systems and online catalogs, and the development of search features and techniques to improve name searching (Drabenstott and Weller, 1996; Everett and Pilachowski, 1986; Molto, 1993; Spink and Leatherbury, 1994).

Recent studies have also examined the terms submitted to Web search engines and suggested that personal names may form part of the low frequency terms submitted (Jansen *et al.*, 2000). Spink *et al.* (2002a) also suggest that personal names form part of the unusually large number of low frequency terms used in Web searching. However, limited studies have specifically examined personal name searching on Web search engines.

In the next section of the paper, we present the research questions addressed in this study, followed by the research design for our study, the findings of the study of Web queries from Alta Vista and AlltheWeb.com, our discussion and conclusions related to our key findings.

Research questions

Our goal in this study was to examine the degree that people use personal names as Web queries. The specific research questions we address in the study are:

- Do personal names form a major part of queries to Web search engines?
- What are the characteristics of personal name Web searching?
- How effective is this mode of personal name Web searching?

This study is important and timely, as we seek to understand more about how and why people search the Web and the nature of Web-based information seeking within specific domains. Few studies have examined personal name searching on the Web.

Research design

Data collection

We were interested in how users are searching for information on other people or personal name searching via Web search engines. To address these research questions, we obtained, and qualitatively analyzed, actual queries that searchers submitted to two commercial Web search engines: AlltheWeb.com and Alta Vista in 2002. Alta Vista and AlltheWeb.com are major Internet companies that offer free Web searching and a variety of other services. Alta Vista users in the query set were largely from the US and AlltheWeb.com users were largely Europeans from Germany and Norway.

Table I shows the basic data from the study.

The researchers were not able to secure data from other Web search engines, such as Google or Microsoft Search.

AlltheWeb.com queries. We obtained a data set of 1.2 million queries, and qualitatively analyzed, actual queries submitted to AlltheWeb.com, a major European Web search engine at the time of the study owned by FAST. Since the study, an outside company has purchased the FAST corporation (Kane, 2003; Liedtke, 2003). AlltheWeb.com supports several query operators including AND, OR, NOT, NEAR, MUST APPEAR, MUST NOT APPEAR, and PHRASE (AlltheWeb.com, 2003). The queries examined for this study were submitted to AlltheWeb.com on May 28, 2002, spanning a 24-hour period.

The queries were recorded in a transaction log and represent a portion of the searches executed on the Web search engine on that particular date. The transaction log held a large and varied set of queries (over one million records). We examined a random sample of 10,000 queries grouped into sessions from the larger data set using Poisson sampling techniques (Ozmutlu *et al.*, 2002).

Alta Vista.com queries. In 2002, Alta Vista was the ninth most popular search engine (Sullivan, 2002) with a content collection of 550 million Web pages (Sullivan, 2000). Alta Vista supported several query operators including AND, OR, NOT, NEAR,

Table I.
Characteristics of the AlltheWeb.com and Alta Vista Web query data sets analyzed from 2002

	AlltheWeb.com	Alta Vista.com
Million query data set	1.2	1.2
Sessions	153,848	365,350
Queries	451,551	1,073,388
Terms per query	2.4	2.9
Random sample	10,000	10,000

MUST APPEAR, MUST NOT APPEAR, and PHRASE (AltaVista, 2003). In 2003, Yahoo! purchased Overture Services, which has previously purchased AltaVista (Morrissey, 2003; Liedtke, 2003).

The queries examined for this study were submitted to Alta Vista on September 8, 2002 and span a 24-hour period. We took a random sample, using Poisson sampling techniques (Ozmutlu *et al.*, 2002), of 10,000 queries grouped in sessions from a data set of 1.2 million Alta Vista Web queries from September 8, 2002.

Each record within the AlltheWeb.com and Alta Vista transaction logs contained three fields:

- (1) Time of day: measured in hours, minutes, and seconds from midnight of each day as recorded by the respective Web server.
- (2) User identifier: an anonymous user code assigned by the respective server.
- (3) Query terms: terms exactly as entered by the given user.

Data analysis

We qualitatively examined random samples of 10,000 Alta Vista and 10,000 AlltheWeb.com Web queries grouped into sessions.

Personal name query classification. Web queries were qualitatively examined for personal name queries by two researchers. We classified each query as either a personal name query or a non-personal name query.

To judge the query intent as a personal name, the query must contain at least a two word personal name in any language (e.g. John Smith, Misha Pavlovsky). Each personal name query was further classified into the following categories:

- *Celebrity or non-celebrity name.* Celebrity personal name queries were broadly interpreted to include movie stars, singers, actors, politicians, well-known people, etc.
- *Double quotations.* We classified each personal name query as within or not within double quotations, e.g. "John Smith". If the user does not enter a personal name in double quotations, the query is processed as a two terms query John Smith, not "John Smith". Neither the AlltheWeb.com nor Alta Vista search home page highlights the need to use double quotations when searching for personal names.
- *Additional terms.* We classified each personal name query as other terms were or were not included with the personal name in the query, e.g. John Smith Atlanta. User may add an additional identifier term(s) to refine their search.

We did not classify the personal name queries by gender, language nor correctness of the personal name's spelling.

To check coding consistency, each researcher recoded 2,000 queries previously classified by the other researcher. The researchers met again in order to make final decision about how the classifications. The two researchers discussed each disputed query until a classification consensus for that query was reached.

Name retrieval. A total of 100 queries containing names and no operators were randomly selected from the Alta Vista data set. We submitted each of these one hundred queries to four search services (AOL, Google, MSN, and Yahoo!). We retrieved the results page containing the top ten retrieved results from each search service. The

entire process of query submission and results retrieval took approximately 90 seconds.

An independent evaluator then viewed each results page and made a binary judgment on whether or not the name, as it appeared in the query was also present on the Web page. The terms in the name had to be adjacent on the Web page for the evaluator to rate the Web page a match. The evaluator repeated this process for all Web pages retrieved for each query from all search services. Once judgments were made on all pages, we calculated the matching rate for each query, along with various other statistics.

Double quotations. Using the 100 queries of personal names from the data set without double quotations, we also recorded the number of results retrieved, as reported by the specific search engine. We then enclosed each of these queries in double quotations, resubmitted them to the same four search engines, and again recorded the number of results retrieved. The entire process of submitted the queries and retrieving the results took less than five minutes, so there was little opportunity for the indexes to changes during the process.

Results

Table II shows the results of our query analysis.

Personal name queries

Overall, one in 25 (4 percent) of queries submitted to AlltheWeb.com and Alta Vista were personal name queries. Personal name queries were not a large or major proportion of all Web queries. However, seeking information electronically about other people is a common feature of the information society.

Celebrity name queries

Overall, one in four (25 percent) personal name queries submitted to AlltheWeb.com and Alta Vista were celebrity name queries. The number of celebrity personal name queries was not a large or major proportion of all Web queries, but these queries were a significant portion of the queries searching for persons. The most common celebrity names searched for in our query sample sets were Britney Spears, Kylie Minogue and Milla Jovic.

Variables	AlltheWeb.com 10,000 queries	%	Alta Vista 10,000 Queries	%
Personal name queries	489	4	360	3.6
Celebrity name queries	120	24.5 of PNQ	89	24.7 of PNQ
Personal name queries in double quotations	54	11 of PNQ	87	24.1 of PNQ
Personal name queries with additional terms	55	11.4 of PNQ	63	17.5 of PNQ

Note: PNQ = personal name queries

Table II.
Study results

Personal names in double quotations

Some one in ten (11 percent) of personal names searches on AlltheWeb.com, and one in four (24.1 percent) of personal name searches of Alta Vista, included double quotations. Alta Vista users were much more likely to submit their personal name queries with double quotations. However, most users enter personal name queries, both celebrity and non-celebrity, without double quotations. The search engine processes the personal name queries without double quotations as john smith. The query may not be sufficiently precise and this increases the number of results the search engine retrieves. This finding would seem to suggest a lack of understanding by users of the need to use double quotations to achieve the correct processing of query submitted as a personal name.

To test this assumption, we compare the number of results retrieved by personal names queries without double quotations to those same queries enclosed in double quotations on four major Web search services. Table III shows the results of this analysis.

From Table III, the use of the double quotations generally decreased the number of results on all search engines; however, this was the consistent case for all queries. On all four search engines the use of double quotations sometimes counter intuitively increased the number of retrieved results, ranging from 3 percent of the queries on MSN to 27 percent of the queries on Google.

We performed a paired *t*-test to evaluation whether or not the use of double quotations resulted in any significant difference in retrieved results. The results of this analysis are displayed in Table IV.

The use of the double quotations resulted in statistically significant differences in the number of retrieved results on both AOL and MSN. There was no statistical difference in using the double quotations on Google or Yahoo! compared with not using them.

	AOL		MSN		Google		Yahoo!	
	Without	With	Without	With	Without	With	Without	With
Average	129,674	30,190	25,590	1,584	178,817	84,798	5	4
St. dev.	446,672	91,821	76,787	5,125	562,996	251,077	11	9
Max	3,310,015	738,010	569,378	43,059	3,820,000	1,870,000	53	49
Min	1	-	-	-	1	-	-	-
Decrease	89		97		72		93	
Increase	10	10 percent	2	2 percent	27	27	6	6 percent
No change	0		13	13 percent	0		67	67 percent

Table III.
Retrieved results without
and with double
quotations

Search engines	<i>t</i> value	<i>p</i>
AOL	2.34	0.05
Google	1.73	0.08
MSN	3.19	0.01
Yahoo!	1.81	0.07

Table IV.
Paired *t*-test of search
results without and with
double quotations

Personal name queries with additional terms

Some one in ten (11.4 percent) of personal name searches on AlltheWeb.com, and 1 in 5 (17.5 percent) on Alta Vista, included additional term(s). Alta Vista users were more likely to submit their personal name queries with an additional term(s). Some users add one or more additional identifying terms to refine their search. For example, a user entered the query David Klein Vancouver lawyer. In this case, the user did not use double quotations, but did add two additional terms to further refine the personal name query.

Some advanced users followed a personal name in double quotes with an AND then an identifier term. In these cases, the AND is redundant if the default search engine operator is also AND. Additionally, the use of the AND would decrease the effectiveness of the query if the search engines did not treat it as a stop word.

Some users entered names such as F. Scott Fitzgerald, including the dot after the F, or they used a dash – between the personal name and an additional identifier term. Some users entered a personal name and then included an additional identifier term in double quotes.

Name retrieval effectiveness. The results of the name retrieval searches on four major Web search engines are displayed in Table V.

The 100 queries retrieved 3,131 results within the top ten ranked results for each query of which 2,574 contained a name match. The matching rate across the four search services varied from a low of 0.77 for Yahoo! to a high of 0.84 for Google. Some 55 of the queries retrieved zero results on Yahoo! and 6 of the queries retrieved zero results on MSN. The mean matching rate across all queries was 0.81. We also conducted a paired *t*-test to evaluate if there was any significant difference among the search engines. The results are displayed in Table VI.

There was no significant difference between AOL and Google. However, there were significant differences among all other search engines.

Discussion

In this section of the paper, we address each of the research questions examined in our study.

Do personal names form a major part of queries to Web search engines?

Overall, our study shows that personal name searching is common, but not a major topic of Web searching. Users are more likely to enter a noun phrase into a Web search engine or be looking for information on topics such as business, than seeking information about other people. In many ways, personal name searching is more specific than noun phrase searching. For example, a user looking for more general information on types of European cars must determine the correct search terms to use. However, person searching usually involves a known name.

Some one in four people who entered a personal name into a Web search engine were looking for celebrities. The other three in four queries were seeking information on non-celebrities, who may or may not be people they know from some related context, such as a work or social environment.

	Match	No match	No results	Final
<i>Overall</i>				
Total	2,574	557	829	
Average	26.00	5.63	8.37	0.81
St. dev.	10.74	8.75	5.15	0.30
Mode	30	0	10	1.00
Sessions with zero results	61			
<i>By search engine</i>				
AOL				
Total	825	156	9	
Average	8	2	0	0.83
St. dev.	3	3	1	0.31
Mode	10	0	0	1.00
Sessions with zero results				0
Google				
Total	831	150	9	
Average	8	2	0	0.84
St. dev.	3	3	1	0.30
Mode	10	0	0	1.00
Sessions with zero results				0
MSN				
Total	724	197	69	
Average	7	2	1	0.78
St. dev.	4	3	3	0.34
Mode	10	0	0	1.00
Sessions with zero results				6
Yahoo!				
Total	194	54	742	
Average	2	1	7	0.77
St. dev.	3	2	4	0.38
Mode	0	0	10	1.00
Sessions with zero results				55

Table V.
Name searching queries

Search	Service pairs	<i>t</i> value	<i>p</i>
AOL	Google	-0.58	0.56
AOL	MSN	4.17	0.01
AOL	Yahoo	9.38	0.01
Google	MSN	4.34	0.01
Google	Yahoo	9.80	0.01
MSN	Yahoo	-6.83	0.01

Table VI.
Paired *t*-test of search results

What are the characteristics of personal name Web searching?

When searching for information on other people, users enter few terms and queries in their search sessions. Personal name searches were generally short and involved little query reformulation. This finding is similar to previous research results showing the short nature of Web searches in general (Jansen *et al.*, 2000; Spink *et al.*, 2002a).

This finding suggests that many users have not cognitively grasped that the huge size of the Web will affect the effectiveness of their queries. By entering a personal name as a query with no further identifiers, such as location or occupation, users may be overestimating the amount of personal information about individuals on the Web, or they may be underestimating how many people share the same name.

Effectiveness of Web name searching

The most common personal name Web search includes one query with two terms with no double quotations around personal names, and no additional identifier terms. The vast majority of people who submitted personal name queries did not use double quotations. Only some 5 percent of those who entered personal name queries submitted a query with double quotations with additional identifier terms. Double quotations are a common phrase searching feature of many IR and Web search engines, but the general public of Web searchers appears to utilize them infrequently.

Given our evaluation of the 100 personal name queries, mean matching rate was quite high at 0.81. This is substantially higher performance than reported in other studies of Web effectiveness (Jansen and Spink, 2003). It indicates that double quotations may not be critical for effective performance. With many search engines now using the AND operator by default and using some proximity algorithm, the use of double quotations does not appear to be critical for effective personal name searching. Additionally, since many people go by nicknames, the use of the double quotations may be overly restrictive in some cases. However, this does not mean that double quotations are without value in this type of Web searching.

We submitted some of the personal names used in Web queries in double quotations and non-double quotations format into for major search engines. The double quotations personal name queries generally produced more specific retrieval in the form of Web sites with the first and surname names together. Not using double quotations generally lead to more retrieved results and less specific retrievals. For example, the personal name Stephen Jones was submitted as a query with no double quotations or identifier terms. This query retrieved three sponsored links and 925,744 Web sites. Many of these Web sites contained the words Stephen and Jones adjacent to some other term, not Stephen Jones.

Our study findings suggest that many users did not seem to appreciate that not using double quotations will trigger the searching of a personal name with and between the first and last names. The lack of double quotations usage, limited use of query reformulation, and little use of additional identifiers suggests that users are generally only looking at the results of the first personal name query as a form of known item searching

However, a fair percentage of the time, the use of double quotations in personal name queries increased the number of results retrieved, sometimes very substantially (10 percent on AOL and 27 percent on Google). Perhaps because of this counter intuitive effect of double quotations; Web searchers do not see the benefit of the double quotations as an effective method to reduce the number of retrieved results.

Another factor contributing to the low usage of double quotations is that the matching is fairly good without using them. With a matching rate ranging from 77 percent for Yahoo! to 0.84 percent for Google, the searching effectiveness may be good enough for most Web searchers. Of course, we used a very low standard of matching

based solely on whether or not the personal name appeared on the retrieved Web document. We could not determine if the Web document contained information on the particular person the user was searching. It would seem that the use of an additional identifier would be beneficial in this regard.

Use of query reformulation

Few personal name searches included query reformulation. We saw some personal name search sessions where the user first entered a query such as john smith and then entered the reformulated query John Smith. However, this was not a common occurrence. Alta Vista now provides the Alta Vista Prisma reformulation tool at the top of the results page. This tool may increase the level of users' query reformulations. However, this feature does not address the double quotations or additional identifier issues.

Web search engine differences

Previous studies have examined differences in Web search engines querying (Spink *et al.*, 2002b). Our data analysis suggested some differences between AlltheWeb.com and Alta Vista.com users. An interesting difference was the Alta Vista users' greater use of double quotations and additional identifier terms in personal name queries. The reason for this difference is not obvious from the query analysis. We can suggest that a larger group of Alta Vista users appear to have an understanding of advanced search features.

Alternatively, AlltheWeb.com users had a slightly higher use of personal name queries and were more likely to be looking for celebrities. Spink *et al.* (2002b) found that searches on people, places and things was a more major category of European AlltheWeb.com search topics, then compared to users of the Excite US-based Web search engine.

The mean matching of the personal name queries without double quotations may seem surprisingly. The Web search services are compensating for the lack of use of operators by searchers, probably by using some proximity retrieval algorithm and ranking methods that presented Web document with the terms near each other higher in the results listing. This is in line with other studies (Eastman and Jansen, 2003) showing that operator usage cannot always be implemented in a straightforward manner without some understanding of the underlying IR system.

Summary

Our study contributes to the Web searching literature in several important ways. In summary, people have trouble with personal name searching and the correct and effective formatting of such queries that will result in the effective system performance. The use of double quotations appears not to be a well-understood feature of Web search engine techniques. Additionally, the use of double quotations appears to have increased the number of retrieved results a fairly high percentage of the time.

The strengths of this study are the use of data from users submitting real queries and viewing actual Web pages. Accordingly, it provides a realistic glimpse into how users search, without the self-selection issues or altered behavior that can occur with lab studies or survey data. In addition, we analyzed a large sample of queries from a quite large group of users. We also obtained data from two popular search engines as

measured in by both document collection and number of unique visitors to ensure that our results were generalizable.

As with any research, there are limitations that should be recognized, including the use of query data from only two commercial Web search engines. We also do not have information about the browsing patterns of the users once they leave the search engine to visit a Web document. Similarly, we do not have information about the demographic characteristics of the users who submitted queries, and there is no knowledge of the underlying cognitive motivation concerning the searcher's information.

Implications

Our study has implications for Web searchers and search engine designers. Web searchers need to put some effort into understanding how Web search engines work, their commands and features, and techniques for more effective searching. People also need to put more effort into improving their information behaviors. Web users need to understand the importance of investing time in developing their information behaviors and searching skills to use complex Web searching tools more effectively. New computing and interface techniques may not provide the complete solution to the problem of ineffective Web searching.

Web search engine and interface designers have known for quite a while that many Web searchers are not using even the most basic of search commands and Boolean operators available to them. Alternatively, how and when to use double quotations and Boolean operators is not mentioned, nor featured, on most Web search engine homepages. Most users do not seem to know these techniques exist. Previous Web researchers have arrived at this finding (Spink *et al.*, 2002a).

However, users' interactions with Web search engines are still relatively short and limited. A new generation of Web searching tools is needed that are designed to support people seek to resolve their information problems. Improved Web search will come with a combination of greater user effort in use and the design of more effective search tools that support user information behaviors. However, it is still a hard row to hoe till we see major improvement at a societal level in IR by systems and users.

Conclusion and further research

Our results provide important insights into the current state of Web searching and Web usage. The non-use of double quotations during personal name searching, combined with short queries, does not seem to be a successful strategy to maximize searching effectiveness. However, it appears that Web search engine users are finding relevant information with this searching strategy even though they usually increase their retrieval output to include many potentially not relevant Web sites.

This study raises a range of research questions to be pursued with further user studies exploring why people search for information about people on the Web. We need to conduct more detailed analysis of users engaged in personal name Web searching and other specific domains to study the nature and effectiveness of their searches. Further research is also need to examine why users do not implement phrase and other advanced search techniques.

References

- AlltheWeb.com (2003), "Advanced search", available at: www.alltheweb.com/advacned?cs=utf-8 (accessed 21 July 2003).
- AltaVista (2003), "Special search terms", available at: www.altavista.com/help/adv_search/syntax (accessed 16 May 2003).
- Ayers, M. (2000), "Gathering information on people using the Internet", available at: www.jenkinslaw.org/collection/researchguide/publications/michelle-gathering.shtml (accessed 22 July 2003).
- Borgman, C.L. and Seigfried, S.L. (1992), "Getty synonyme and its cousins: a survey of applications of personal name algorithms", *Journal of the American Society for Information Science*, Vol. 43, pp. 459-76.
- Carroll, J.M. (1985), *What's in a Name? An Essay in the Psychology of Reference*, W.H. Freedman, New York, NY.
- Drabenstott, K.M. and Weller, M.S. (1996), "Improving personal name searching in online catalogs", *Information Technology and Libraries*, Vol. 15 No. 3, pp. 135-55.
- Eastman, C. and Jansen, B.J. (2003), "Coverage, relevance, and ranking: the impact of query operators on Web search engine results", *ACM Transactions on Information Systems*, Vol. 21 No. 4, pp. 383-411.
- Everett, D. and Pilachowski, D.M. (1986), "What's in a name? Looking for people online – humanities", *Database*, Vol. 9 No. 5, pp. 26-34.
- Federal Trade Commission (2002), "Privacy initiatives", available at: www.ftc.gov/privacy/ (accessed 20 July 2003).
- Hayes, P. (1994), "Namefinder: software that finds names in text", *Proceedings of RIAO 94, October 11-13, 1994, New York*, Vol. 1, pp. 762-74.
- Hermansen, J.C. (1985), "Automatic name searching in large data bases of international names", PhD thesis, Georgetown University, Washington, DC.
- Jansen, B.J. and Spink, A. (2003), "An analysis of Web information seeking and use: documents retrieved versus documents viewed", *IC'03: Proceedings of the 4th International Conference on Internet Computing, Las Vegas, Nevada, 23-26 June*, pp. 65-9.
- Jansen, B.J., Spink, A. and Saracevic, T. (2000), "Real life, real users, and real needs: A study and analysis of user queries on the Web", *Information Processing and Management*, Vol. 36 No. 2, pp. 207-27.
- Kane, M. (2003), "Overture to buy search services", *CNET News.com*, 23 February, available at: <http://rss.com.com/2100-1023-985850.html?type=pt&part=rss&tag=feed&subj=news> (accessed 1 March 2003).
- Lane, C.A. (1996), "Naked in cyberspace: how to find personal information online", *Information Today*.
- Liedtke, M. (2003), "Yahoo! buys overture in \$1.6 billion deal", available at: http://story.news.yahoo.com/news?tmpl=story&u=/ap/20030714/ap_on_hi_te/yahoo_overture_19 (accessed 21 July 2003).
- Molto, M. (1993), "Improving full text search performance through textual analysis", *Information Processing and Management*, Vol. 29 No. 5, pp. 615-32.
- Morrissey, B. (2003), "Overture to buy AltaVista [Web site]", *Internet Advertising Report*, 18 February, available at: www.internetnews.com/IAR/article.php/1587171 (accessed 16 May 2003).

- Navarro, G., Baeza-Yates, R. and Arcoverde, J.M.A. (2003), "Matchsimile: a flexible approximate matching tool for searching proper names", *Journal of the American Society for Information Science and Technology*, Vol. 54 No. 1, pp. 3-15.
- Ozmutlu, H.C., Spink, A. and Ozmutlu, S. (2002), "Analysis of large data logs: an application of Poisson sampling on Excite Web queries", *Information Processing and Management*, Vol. 38 No. 4, pp. 473-90.
- Pfeiffer, U., Poersch, T. and Fuhr, N. (1996), "Retrieval effectiveness of proper name search methods", *Information Processing and Management*, Vol. 32, pp. 667-79.
- Search Software America (2003), *The Man, Myth and Magic of Name Searching and Matching*, Search Software America, Old Greenwich, CT.
- Sholtz, P. (2000), "Economics of personal information exchange", *First Monday*, Vol. 5 No. 9.
- Spink, A. and Leatherbury, M. (1994), "Name authority files and humanities database searching", *Online and CD-ROM Review*, Vol. 18 No. 3, pp. 143-7.
- Spink, A., Jansen, B.J., Wolfram, D. and Saracevic, T. (2002a), "From e-sex to e-commerce: Web search changes", *IEEE Computer*, Vol. 35 No. 3, pp. 107-11.
- Spink, A., Ozmutlu, S., Ozmutlu, H.C. and Jansen, J. (2002b), "US versus European Web searching trends", *ACM SIGIR Forum*, Vol. 36 No. 2.
- Sullivan, D. (2000), "Search engine sizes", available at: <http://searchenginewatch.com/reports/sizes.html> (accessed 30 August 2000).
- Sullivan, D. (2002), "Nielsen/NetRatings search engine ratings".SearchEngineWatch.com available at: www.searchenginewatch.com/reports/netratings.html (accessed 6 January 2002).
- Thomson, P. and Dozier, C.C. (1997), "Name searching and information retrieval", *West Group Report*.

Further reading

- Jansen, B.J. (2003), "Operators not needed? The impact of query structure on Web searching results?", in *IRMA'03: Proceedings of Information Resource Management Association International Conference, Philadelphia, PA, 18-21 May 2003*, pp. 814-7.
- Jansen, B.J. and Eastman, C. (2003), "The effects of search engines and query operators on top ranked results", *IEEE ITCC'03: 4th International Conference on Information Technology, Coding and Computing, Las Vegas, Nevada, 28-30 April 2003*, pp. 135-9.