Comparing Google to Ask-a-Librarian Service for Answering Factual and Topical Questions

Pertti Vakkari and Mari Taneli

University of Tampere, FIN-33014, Finland
Pertti.Vakkari@uta.fi

Abstract. This paper evaluates to which extent Google retrieved correct answers as responses to queries inferred from factual and topical requests in a digital Ask-a-Librarian service. 100 factual and 100 topical questions were picked from a digital reference service run by public libraries. The queries inferred simulated average Web queries. The top 10 retrieval results were observed for the answer. The inspection was stopped when the first correct answer was identified. Google retrieved correct answers to 42 % of the topical questions and 29 % of factual questions. Results concerning the characteristics of queries and retrieval effectiveness are also presented. Evaluations indicate that public libraries' reference services answer correctly 55 % of the questions. Thus, Google is not outperforming Ask-a-Librarian service, although it seems to perform relatively satisfactory in retrieving answers to topical questions.

1 Introduction

The Internet has brought vast amounts of information and search tools within the reach of people. Search engines have enhanced the accessibility of those resources by providing easy means to search, identify, locate and download information items needed. Finding information and answers to requests is only a few clicks away.

The number of Internet users is growing steadily especially in industrialized countries. The use of search engines has gained popularity world-wide. They are considered easy to use. Easy access to information resources on the Internet from one's own desk with the help of search engines is challenging the functions of libraries, public libraries in particular (D'Elia & al. 2002, Huysmans & Hillebrink 2008). Material and services provided by libraries are also available on the Internet. This may result in a potential change in consumer demand for the types of services provided by libraries. People can compare the suitability and performance of the services of the library and the Internet and make their choice based on that evaluation.

There is not much empirical evidence on this change. D'Elia & al. (2002) showed that in the year 2000 the use of public libraries and the Internet in the US were complementary. Most of those who used the library used the Internet and vice versa. Using the Internet for various purposes was associated neither with the reasons why people use the library or their frequency of use.

It is evident that the use of the Internet influences differently the use of various library services. Borrowing novels or reference services are probably influenced

differently by the Internet (Huysmans & Hillebrink 2008). It is likely that the demand of the reference services is changing due to use of search engines. Numminen & Vakkari (2009) have shown that of the questions sent to an Ask-a-Librarian digital reference service the proportion of topical questions has decreased and that of factual questions increased significantly between the years 1999 and 2006. They suggest that this change may be due to the way users are searching on the Internet.

Although there have been several evaluations of search engines, they are compared with each others (Spink & Jansen 2004, Lewandowski 2008). Similarly, digital Ask-an-Expert-services - including librarian - are evaluated by comparing these services with each other (Janes & al. 2001). It is very rare to do across search space evaluations, e.g. how search engines function compared to digital library services like question answering or even retrieving information from a digital library (cf. McCown & al. 2005). However, it is necessary to evaluate search engines with appropriate services provided by traditional or digital libraries, because people may increasingly use search engines as surrogates for library services. The aim of this study is to explore to what extent Google retrieves correct answers to questions addressed to a digital Ask-a-Librarian reference service. Specifically, we analyze what is the proportion of correct answers Google retrieves as responses to queries inferred from factual and topical requests in an Ask-a-Librarian service.

2   Related Research

2.1   Evaluation of Search Engines

Search engines can be evaluated on several dimensions, but retrieval effectiveness has been the major theme (Lewandowski 2008). By effectiveness is meant systems' ability to retrieve relevant information items. It is measured typically by precision and recall. Due to the differences between search engines and digital Ask-a-Librarian services, the reasonable feature to compare their functioning is their ability to provide correct answers to questions. In our study we analyze how effectively Google retrieves answers to queries inferred from questions to an Ask-a-Librarian service. Therefore, retrieved correct items are the indicator on which we base the evaluation.

There are several studies analyzing the effectiveness of search engines. The major studies in the 1990s (Chu & Rosenthal 1996, Gordon & Pathak 1999, Leighton & Srivastava 1999) have used queries drawn from real reference questions. The number of queries used is small varying in those studies from 5 to 15. The number of queries in later studies has varied from 25 to 70 (Lewandowski 2008).

The number of retrieval results analyzed for the evaluation has typically been either 10 or 20 first information items (Lewandowski 2008). This is consistent with the findings of empirical Web search studies, which show that the majority of searchers view only the first result page. A typical user views on average 8 Web documents per session (Spink & Jansen 2004). The studies also show that the number of words in queries is typically 2.5-3 (Spink & Jansen 2004), and that topical queries are shorter compared to factual ones. Kellar & al. (2007) have shown that on average topical queries consist of 3.3 words, whereas factual queries consist of 4.7 words.

The precision of search engines in studies varies somewhat. Precision is the proportion of relevant items among those retrieved. In Gordon & Pathak (1999) precision

within the top ten varied from 40.6% to17.6%, whereas in Chu & Rosenthal (1996) the range was from 78 % to 55 %. In a more recent study precision within the top 20 varied from 49 % to 34 % (Lewandowski 2008).

There is one study, which has compared a search engine and a digital library service in their ability to produce useful results. McCown & al. (2005) measured the pedagogical usefulness of results returned by the National Science Digital Library (NSDL) and Google. Teachers evaluated search terms and search results based on the Standards of Learning (SOL) for schools. They ranked on a five-point scale the links returned by Google as more relevant to the SOL than the links returned by NSDL. On average neither of the search engines generated search results that would satisfy the educational requirements in the SOL. However, the proportion of Web pages rated as adequate of all results (precision) was in Google 38 % and in NSDL 17 %. Thus, Google seemed to outperform NSDL in retrieving educationally useful results.

2.2   Web Queries

Web queries can be divided into informational, navigational and transactional (Broder 2002). The aim of an informational query is to obtain information about the topic of the query (Rose & Levinson 2004). Navigational queries aim at reaching a particular site. They can be described as known item searches (Broder 2002, Rose & Levinson 2004). Transactional queries seek to obtain something else than information, typically resources from the Web (Rose & Levinson 2004).

Most of the queries on the Web are informational. The proportion of informational queries was about 60 % in Rose & Levinson (2004) and 40 % in Broder (2002). Topical queries were distinguished only in Rose & Levinson by the name "undirected queries". Their share varied between 23-31 %. Fact-finding i.e. directed queries comprised only 3-7 % of all queries in Rose & Levinson (2004). Navigational queries are equal to known item questions. Their share was about a quarter in Broder (2002) and about one seventh in Rose & Levinson (2004).

Thus, topical queries seem to be much more common than factual queries and somewhat more common than known item queries among Internet searches. Based on the proportions of these query types, we may expect that the use of search engines challenges reference services in public libraries most in answering topical questions, but not so much in answering factual or known item questions.

2.3   Evaluation of Digital Reference Services

There is a long tradition of evaluating reference services in libraries. Unobtrusive observer studies are used to assess the quality of question responses. In unobtrusive observation evaluations, trained questioners act as real clients to pose reference questions in the actual reference setting (Katz 2002). The accumulated results show that about 55 % of the questions are answered accurately (Katz 2002). This approach has also been used in the evaluations of digital reference services. Kaske & Arnold (2002) found that in 36 libraries, chat reference services answered correctly 55 % and email services 60 % of the 12 questions asked. Kwon (2007) showed that public libraries' chat services answered completely 78 % of factual questions and 70 % of topical questions. Arnold & Kaske (2005) found an even greater proportion (92 %) of correctly answered questions among 3000 questions to a public library's chat service.

It is difficult to assess validly the trends emerging from these results due to the varying nature of the data used. The 55 % rule seems to apply to those studies using unobtrusive testing, whereas higher figures are derived from the answers to users' real questions as assessed by researchers. Although the findings hint that a growing proportion of users' requests are answered correctly especially through digital reference services, this hypothesis requires further empirical evidence.

3   Research Design

The aim of this study is to explore to which extent Google retrieves correct answers to questions addressed to a digital Ask-a-Librarian reference service. The specific research questions are as follows:

- What is the proportion of correct answers Google retrieves as responses to queries inferred from factual and topical requests in an Ask-a-Librarian service?
- How are query characteristics related to search effectiveness in question types?

3.1   Questions

We focus on factual and topical questions. A factual question has typically one definite answer and it can be answered by consulting one or two reference tools (Katz 2002). We categorized the factual questions into sub-categories based on our data. The following kind of factual question types emerged from the data: definition of a concept (What is the cost of living index?), origin or meaning of a name or saying (What is the meaning of the name Maximus?), instruction how to make a product (Instruction to make a broom?), translation of a word (What is depression (in Finnish) in Swedish?), and other facts (How many state officials are there in Finland?).

A topical question requests materials or information on a certain subject (Katz 2002). We divided topical questions into questions concerning individual persons, and other topical questions. The former mostly included questions concerning e.g. artists or writers (I need information about the novelist Anni Swan). The latter typically focused on a subject matter (I need materials about the early history of globalization).

We selected questions from an Ask-a-Librarian digital reference service (http://www.libraries.fi/en_GB/ask_librarian). This service is run by about 50 public libraries. Questions are sent via a Web form and answered by email. The staff running the service typically directs questions to the local public library of the requesters. The question answer pairs are stored in the public archive of the service.

Our data consists of 100 factual and 100 topical questions. They were selected from the archive of the service from September 2008 backwards so long that the required number of both question types was achieved. Both researchers classified the questions. The inter-rater reliability among factual questions was 94 % and among topical questions 90 %, which can be considered as high (Krippendorf 2004).

3.2   Queries

We formulated queries of the factual and topical questions. We sought to simulate typical users of search engines, who formulate short queries of 2.5-3 words and view

only the first result page (Spink & Jansen 2004). In queries we used only words, which were used in the questions. For selecting the expressions for the queries, we made conceptual query plans for each question (Lancaster & Warner 1993). A facet is a concept identified from, and defining one exclusive aspect of a request (Sormunen 2000). We identified the major facts of the requests. The maximum number of facets per query was three. Each facet was expressed typically by one word in the basic form picked from a question. There were some exceptions when a facet was expressed by more than one word. Those exceptions consisted mainly of phrases from requests concerning the origin of sayings like the first line of a poem or other necessary phrases for expressing a facet. Synonyms were not used. On average factual queries consisted of 2.2 facets and 3.5 words and topical queries of 1.8 facets and 2.5 words. Thus, on average the length of both factual and topical queries seem to resemble typical Web queries (Spink & Jansen 2004). Our factual queries contained more words than topical ones, which corresponds with the findings in Kellar & al. (2007).

We calculated between both researchers the inter-rater reliability of the queries formulated. The reliability rate among factual queries was 92 % and among topical queries 90 %, which can be considered as high (Krippendorf 2004).

3.3   Answers and Evaluation

The answers given by the Ask-a-Librarian service formed the basis for assessing the correctness of the answers retrieved. The information provided in the pages and links retrieved should correspond to the answer given to a request by the service but not necessarily be exactly the same. For instance, links and document lists or information in items retrieved to topical questions should not correspond totally with the information in the original answer, but to cover its main content. The point was that the content of the information item provided a satisfactory answer to the question in line of the answer by the reference service.

We categorized the correctness of the information item as an answer as follows: It was considered as correct if it contained the main content of the answer. It was partially correct if it contained some aspects of the answer. It was categorized as "mention" if the information item provided only a mention of the answer.

To identify the answer we observed the first ten information items on the result page. First we assessed whether the answer could be found in the results descriptions on the result page containing the first ten items. If it was not identified on the result page, the first link from each of the results was inspected from the first to the tenth. The inspection was stopped when the first correct answer was found.

We calculated for all queries the proportion of answers found, and for those queries retrieving an answer the average rank of the answer and its average degree of correctness.

We supposed that the answers of the reference service were correct, but this is not necessary the case. For assessing the quality of the answers we inspected them. Although we did not check the sources of the answers, they however, seemed reliable to us. This is the major limitation of our study, which may bias the results.

4 Results

4.1 The Answers Retrieved

By using Google answers were found to 47 % of the questions. Significantly more answers were retrieved for topical than factual questions (t=3.3, df=198, p=.000). The proportion of answers for the former was 65 % and for the latter 29 %. Thus, Google produced over twice as many correct replies to topical than factual queries.

Table 1. The average rank of answers and the correctness of answers by question type

	Factual Q (n=29)	Topical Q (n=65)	p
Rank of answers	2.1	2.0	.820
Degree of correctness	3.0	2.5	.000

If we observe only answered questions there were no significant differences in the rank of references providing the correct answer to factual or topical questions (Table 1). In both groups on average the second item on the result list contained the correct answer. The first item included the correct answer in 62 % of factual questions and in 65 % of topical questions. The three first items contained the correct answer in 83 % of factual questions and 86 % of topical questions. Thus, in most of the cases in both groups the correct answers were ranked in the very beginning of the result list.

We scored the correctness of the answers as follows: totally correct=3, partially correct=2, and mention=1. The correctness of the answers differed significantly between factual and topical questions (t=.3.7, df=92, p=.000) (Table 2). All the answers to the factual questions totally covered the answer, whereas 65 % of answers to the topical questions were totally correct. In 35 % of topical requests the answers did not include all the necessary information. Thus, it seems that if Google was able to retrieve an answer to a factual query, it was always totally correct, whereas in about one third of the cases the answer to a topical query was incomplete.

Table 2. The degree of correctness in answers by question type (%)

Correctness	Factual Q (n=29)	Topical Q (n=65)	Total (n=94)
Totally	100	65	76
Partially	0	15	11
Mention	0	20	14
Total	100	100	101

4.2 Characteristics of Queries

The average number of facets in factual queries was 2.2. and in topical queries 1.8 (Table 3). A two-way ANOVA indicated that this difference was marginally significant (F=3.1, p=.08).

Table 3. The average number of facets in queries

Answer	Factual Q (n=100)	Topical Q (n=100)	Total (n=200)
Found (n=94)	1.9	1.7	1.7
Not found (n=106)	2.3	2.2	2.3
Total (n=200)	2.2	1.8	2.0

The queries retrieving correct answers consisted of significantly fewer facets than queries producing no answers (F=18.3, p=.000). The former consisted of 1.7 facets and the latter of 2.3 facets. It seems that queries representing requests by fewer facets retrieve more correct answers. The significant correlation between the number of facets and whether the answer was found (-.35, p=.000) confirms that.

On average queries contained three words (Table 4). Factual queries consisted of 3.5 words and topical queries of 2.5 words. This one word difference in the length of query was significant (F=8.4, p=.004). Results also indicate that queries that retrieve correct answers were significantly shorter (2.5 vs. 3.5 words) than those that did not retrieve correct answers (F=9.7, p=.002). This was regardless of the question type (factual: t=2.1, df=98, p=.04; topical: t=2.3, df=98, p=.009).

Table 4. The average number of words in queries

Answer	Factual Q (n=100)	Topical Q (n=100)	Total (n=200)
Found (n=94)	2.9	2.3	2.5
Not found (n=106)	3.8	2.9	3.5
Total (n=200)	3.5	2.5	3.0

It seems that queries representing requests by fewer words retrieve more correct answers. The significant correlation between the number of words and whether the answer was found was -.30 (p=.000) confirming that. It is likely that those queries containing fewer words are simpler in the sense that they consist of fewer facets. Simple requests containing fewer facets are likely to produce more correct answers as response to queries than complex requests with several facets.

The number of facets was significantly associated with the correctness but not with the rank of the answers found. The correlation between the number of facets and the rank of correct answers was -.04 (p=.74) and between facets and the degree of correctness -.28 (p=.006). The corresponding coefficients between the number of words and rank order was .09 (p=.40) and the level of correctness -.15 (p=.14). Thus, the fewer facets the question contained the higher the correctness of the answer re-trieved. The number of words in queries was associated neither with the rank or the correctness of the items retrieved.

4.3 Categorizing Factual and Topical Questions

Based on data we categorized both factual and topical questions to more specific groups. Factual questions were divided into questions concerning definition of concepts, origin of words, or phrases and sayings, instructions to make a product, translation of a

word or phrase, and typical factual questions. Topical questions were divided into questions concerning persons, and other topical questions. We first analyze factual question types and then topical question types.

Although the number of cases in groups other than typical facts is small, the figures hint marginally at certain trends (Chi2=2.8, df=5, p=.091) (Table 5). It seems that Google retrieves a high proportion of definitions of concepts (71 %), whereas it is not successful in locating origins of sayings or phrases, or providing searchers with translation of words. These categories produced zero hits. About one third of typical facts, origin of words, and instructions were found by Google.

Table 5. Answers found and the number of words in the queries by factual question type (%)

Type of factual question (n)	% found	# words
Definition of a concept (n=7)	71	1.7
Origin of a word (n=6)	33	2.3
Origin of a saying or a phrase (n=10)	0	6.2
Instructions to a product (n=6)	33	3.7
Translation (n=6)	0	4.5
Typical fact (n=65)	31	3.3
Total (n=100)	29	3.5

As we showed earlier there was a significant inverse correlation (-.30, p=.000) between answers found and the number of query words. An ANOVA confirmed that there are significant differences in the length of queries between the groups (F=7.4, p=.000) (Table 5). The number of words in queries concerning concepts was significantly smaller than concerning typical facts, origins of sayings or instructions (Dunnett C: p<.05). The proportion of correct answers in this question type was higher than in other factual questions, and its query length was the shortest of all, on average 1.7 words. The length of queries in those question types with zero hits was the greatest, from 4 to 6.2 words. There seems to be a inverse rank order between the proportion of correct answers and the length of the query across question types.

The proportion of correct answers in queries concerning persons was 72 %, and in other topical queries 60 %. The difference is not significant (t=1.5, df = 63 p=.14). Neither did the rank of answers differ between these groups (t=0.5, df=63, p=.61). Thus, Google tends to produce to the same degree correct answers to queries in both groups and rank them about equally.

The average degree of correctness in answers to questions concerning persons was significantly greater than in other topical question (t=2.2, df=63, p=.032) (Table 6). In the former the answers tend to be correct whereas in the latter they were on the partially correct side.

The number of facets was significantly greater in other topical queries compared to queries concerning persons (t=4.8, df=63, p=000). The number of words was about the same (2.2 vs. 2.4) in both groups (t=1.4, df=63, p=.16). The name of a person comprises one facet, which can be expressed by two words (i.e. first name and family name). This explains in part the difference between the groups in the number of facets and not in the number of words.

Table 6. Means of correctness and query variables by topical question type

	Persons (26)	Topics (39)	p
Rank of answers	1.9	2.1	.610
Degree of correctness	2.7	2.3	.032
Number of facets	1.2	2.0	.000
Number of words	2.2	2.4	.160

It seems that it is easier to represent questions concerning persons as queries compared to other topical questions, and consequently the proportion of correct answers and the correctness of the answers is higher in the former group.

In all, a more specific categorization of question types revealed that the positive associations between short queries and the higher proportion of answers found, and the higher correctness of answers were in part produced by certain question types (definitions of concepts, and persons), which required fewer facets and words.

5   Discussion and Conclusions

Digital libraries are challenged by the increasing provision of information and tools for accessing information on the Internet. It is important to analyze to which degree services provided by libraries are able to compete with the Internet. Topical queries are popular on the Web compared to factual queries (Rose & Levinson 2004). There are signs that the proportion of topical questions in the digital reference services has decreased, and the proportion of factual questions increased (Numminen & Vakkari 2009). It is proposed that this is produced by the way people are searching on the Net.

We have compared the functioning of one particular digital library service to one typical service provided by the Internet. We have analyzed to which degree Google retrieves correct answers to queries inferred from topical and factual questions answered by an Ask-a-Librarian digital service. We formulated queries from requests by simulating average Web searchers. It is known that they use short queries of about three words and look only at the first result page (Spink & Jansen 2004). We analyzed whether the queries produced a correct answer within the top ten retrieval results.

Our study extends our knowledge of how an Ask-a-Librarian service performs compared to Google in answering topical and factual questions. We have used 200 queries, which is much more than in typical studies assessing search engines (Lewandowski 2008) and not less than on average in evaluations of public libraries' reference services. Also the high inter-rater reliability of classifying question types and formulating queries enhance the validity of our findings.

5.1   Search Results

On average Google retrieved correct answers to 47 % of the questions. The correct answer was retrieved to 65 % of topical questions and to 29 % of factual questions. Thus, Google retrieved twice as many correct answers to topical than factual requests. This result supports our hypothesis, that the use of search engines challenges reference services in public libraries most in answering topical questions, but not so much in answering factual questions (cf. Numminen & Vakkari 2009).

On average the second item on the result list provided the correct answer regardless of the question type. The correct answers were typically ranked in the very beginning of the retrieval results, in most of the cases among the three first items.

The level of answers' correctness differed between the question types. All correct answers to factual questions covered the answer totally, whereas 35 % of the answers to topical questions did not include all the necessary information.

It is not possible to compare our findings with the earlier studies on the effectiveness of search engines due to the differences in the measurement of effectiveness. We calculated the proportion of correct answers retrieved, whereas the typical measure of effectiveness has been precision (eg. Gordon & Pathak 1999, Lewandowski 2008).

Evaluation of question answering in public libraries' reference services has produced the rule that 55 % of the questions are answered accurately (Katz 2002). In digital reference services this regularity varies to some extent. Kaske & Arnold (2002) found that 60 % of questions asked by e-mail from 36 public libraries were answered correctly. Kwon (2007) showed that public libraries chat services answered correctly78 % of the factual questions and 70 % of the topical questions.

Compared to these evaluation results it seems that Google performs relatively well when it retrieves correct answers to 65 % of topical questions. However, if we take into account only totally correct answers retrieved, then the proportion of accurate answers drops to 42 %. This is a somewhat poorer performance than average in answering topical questions in digital reference services.

The digital reference services in public libraries seem to outperform Google as provider of answers to factual questions. Google retrieved totally correct answers to 29 % of factual questions. This is clearly less compared to 55 % rule or to the finding in Kwon (2007), which indicated that 78 % of factual questions in public libraries' chat service were answered correctly.

If we suppose that searchers are willing to invest more effort in inspecting the results, the picture changes somewhat in favor of Google. By allowing that each information item on the result list can be inspected three clicks away instead of the one click we used as criterium, the proportion of totally correct answers retrieved to topical questions was 52 % and to factual questions 40 %.

In all, it seems that Google is not severely challenging public libraries' reference services. It does not reach the 55 % proportion of correctly answered questions, which is regular in public libraries. However, if searchers are willing to invest more effort on inspecting the retrieval results, then Google may approach the regular performance level of public libraries' reference services in answering topical questions.

5.2 Characteristics of Queries

Our factual queries contained more facets and words than topical ones. This resembles the finding in Kellar & al. (2007), that factual Web queries contain more words than topical ones. Our results also indicated that queries with fewer facets and words retrieved significantly more correct answers than longer queries regardless of question type. Thus, the fewer facets and words queries contained, the more correct answers they retrieved.

An elaboration of question types showed that factual questions concerning definition of concepts and topical questions concerning persons retrieved more correct

answers compared to other question types within the major question categories. The former retrieved 71 % correct answers, and the latter 72 % correct answers. Both types of queries also contained the lowest number of words and facets compared to other question types in the respective major question categories.

The association between the low number of facets and words, and success in retrieval seem to hint that those questions, which are relative unambiguous and can be expressed by few words, also perform best as queries. Questions concerning the names of persons and definition of concepts are such clear questions. This explains why shorter queries tend to perform best in our study.

5.3   Conclusions

We have shown that Google challenges public libraries' digital reference services most in answering topical questions. It retrieved twice as many correct answers to topical than factual questions. This finding supports the hypothesis that people increasingly make topical searches on the Web (Rose & Levinson 2004), and this is likely to lead to the decrease in the proportion of topical questions addressed to digital reference services (Numminen & Vakkari 2009). In particular, Google was successful in retrieving answers to unambiguous questions, which can be expressed by few facets and words. Although search engines are challenging the role of public libraries' reference services, on average they still outperform search engines. It seems that the engines are currently not able to reach the 55 % level of correct answers, which public libraries regularly provide to their customers according to evaluations.

Google's performance figures improved, when results were inspected following further links they provided. If people are willing to invest more effort on inspecting search results, this increases the likelihood of finding correct answers. Thus, also in this respect the future of libraries' reference service is in the hands of people looking for information.

References

Arnold, J., Kaske, N.: Evaluating the quality of a chat service. Portal 5(2), 177–193 (2005)

Broder, A.: A taxonomy of web search. SIGIR Forum 36(2), 3–10 (2002)

Chu, H., Rosenthal, M.: Search engines for the world wide web: A comparative study and evaluation methodology. In: ASIS 1996 Annual Conference Proceedings (1996)

D'Elia, G., Jörgensen, C., Woelfel, J.: The impact of the Internet on public library use. JASIST 53(10), 802–820 (2002)

Gordon, M., Pathak, P.: Finding information on the world wide web: The retrieval effectiveness of search engines. Information Processing & Management 35(2), 141–180 (1999)

Huysmans, F., Hillebrink, C.: The future of the Dutch public library: Ten years on. SCP, The Hague (2008)

Janes, J., Hill, C., Rolfe, A.: Ask-an-Expert services analysis. JASIST 52(13), 1106–1121 (2001)

Kaske, N., Arnold, J.: An unobtrusive evaluation of online real time library reference services. Paper presented at the Library Research Round Table, ALA Annual Conference, Atlanta, GA. (2002), `http://www.lib.umd.edu/groups/digref/LRRT.html`

Katz, W.A.: Introduction to reference work, vol. 1. McGraw-Hill, Boston (2002)

Kellar, M., Watters, K., Shepherd, M.: A field study characterizing web-based information seeking tasks. JASIST 58(7), 999–1018 (2007)

Krippendorf, K.: Content analysis. Sage, Thousand Oaks (2004)

Kwon, N.: Public library patrons' use of collaborative chat reference service: The effectiveness of question answering by question type. Library & Information Science Research 29, 70–91 (2007)

Lancaster, W., Warner, A.: Information retrieval today. Information Resources Press, Arlington (1993)

Leighton, W., Srivastava, J.: First 20 precision among world wide web search services. JASIST 50(10), 870–881 (1999)

Lewandowski, D.: The retrieval effectiveness of web search engines: considering results descriptions. Journal of Documentation 64(6), 915–937 (2008)

McCown, F., Bollen, J., Nelson, M.: Evaluation of the NSDL and Google for obtaining pedagogocal resources. In: Rauber, A., Christodoulakis, S., Tjoa, A.M. (eds.) ECDL 2005. LNCS, vol. 3652, pp. 344–355. Springer, Heidelberg (2005)

Numminen, P., Vakkari, P.: Question types in public libraries' digital reference service in Finland: Comparing 1999 and 2006. JASIST 60 (2009),
http://www3.interscience.wiley.com/journal/122219062/abstract
DOI: 10.1002/asi.21047

Rose, D., Levinson, D.: Understanding user goals in web search. In: WWW 2004, New York, May 17-22, 2004, pp. 13–19 (2004)

Sormunen, E.: A method of measuring wide range performance of Boolean queries in full-text databases. Acta Universitatis Tamperensis 748. Tampere University Press (2000)

Spink, A., Jansen, B.: Web search: Public searching of the Web. Kluwer, Dordrecht (2004)