Search-Logger
Analyzing Exploratory Search Tasks

Georg Singer, Ulrich Norbisrath, Eero
Vainikko, Hannu Kikkas
Institute of Computer Science, University of Tartu
J. Liivi 2, Tartu, Estonia
{FirstName.LastName}@ut.ee

Dirk Lewandowski
Hamburg University of Applied Sciences
Hamburg, Germany
Dirk.Lewandowski@haw-hamburg.de

ABSTRACT

In this paper, we focus on a specific class of search cases: *exploratory search tasks*. To describe and quantify their complexity, we present a new methodology and corresponding tools to evaluate the user behavior when carrying out exploratory search tasks. These tools consist of a client called *Search-Logger*, and a server side database with frontend and an analysis environment. The client is a plug-in for Firefox web browsers. The assembly of the Search-Logger tools can be used to carry out user studies for search tasks independent of a laboratory environment. It collects implicit user information by logging a number of significant user events. Explicit information is gathered via user feedback in the form of questionnaires before and after each search task. We also present the results of a pilot user study. Some of our main observations are: When carrying out exploratory search tasks, classic search engines are mainly used as an entrance point to the web. Subsequently users work with several search systems in parallel, they have multiple browser tabs open and frequently use the clipboard to memorize, analyze and synthesize potentially useful data and information. Exploratory search tasks typically consist of various sessions and can span from hours up to weeks.

1. INTRODUCTION

"The ultimate search engine would basically understand everything in the world, and it would always give you the right thing. And we're a long, long ways from that." – Larry Page[1]). Search engines like Google, Bing and Yahoo have become the means for searching information on the Internet, supported by other information search portals like Wikipedia and Ask.com. Although search engines are excellent in document retrieval, this strength also imposes a limitation. As they are optimized for document retrieval, their support is less optimal when it comes to ex-

---

[1]Larry Page (Google Founder), in an interview with Business week `http://www.businessweek.com/magazine/content/04_18/b3881010_mz001.htm`

ploratory search tasks. Those are usually described as open-ended, abstract and poorly defined information needs with a multifaceted character [18, 29]. Exploratory search tasks are accompanied by ambiguity, discovery and uncertainty. When performing exploratory search tasks and if "users lack the knowledge or contextual awareness to formulate queries or to navigate complex information spaces, or the search task requires browsing and exploration, system indexing of available information is inadequate" [26]. Such exploratory search tasks fulfill needs like learning, investigating or decision making. Usually exploratory search tasks require a high amount of interaction. A report by Microsoft [25] states that during exploratory searches only 1 in 4 queries is successful. In addition, complex queries have a 38% share and yield "a terrible satisfaction". Search engines also quickly reach their limits when the information seeker is entering a new domain [15]. Along with the increasing popularity of search engines, the areas of their application have grown from simple look-up to rather complex information seeking needs. Look-up searches follow the "query and response" retrieval paradigm [27]. Each time users enter a query, they get a list of possibly relevant results. Look-up searches are among the most basic types of search tasks. Usually they are happening in context with *question answering* and *fact finding*. Typically they are needed to answer who, when and where questions. They are not the means to answer why, what and how questions which can be classified as exploratory search tasks. To cope with those information needs, where the available technologies do not directly produce a solution by query-answer only, users have adopted an exploratory search like behavior (multiple queries, follow links selectively, explore the retrieved document space interactively). The more complex and exploratory a task-based information need becomes, the less a search engine as a single means to fulfill the task appears to be appropriate [27].

Search engine quality measurement initiatives have widely contributed to enhancing search engine quality for look-up searches, but the same is not yet true for exploratory search tasks [27]. The evaluation methodologies focus predominantly on the search system itself, not on the search process that humans need to follow in order to fulfill their search need. Only implicit information is gathered during the classic search engine quality measurement experiments while explicit user feedback is seldom collected [17]. As White et al. stated [27], it is important to also integrate the behavior of users into the evaluation of exploratory search systems that have expanded beyond simple look-up. Although search systems are supporting exploratory search tasks better today,

their evaluation is still limited to those systems that rely on minimal human-machine interaction [15].

In this paper, we provide a method and a set of tools that allow to carry out evaluation experiments for exploratory search tasks. Our approach tries to close the gap between evaluation methods purely focused on technical aspects and expensive laboratory methods. In comparison to classic search engine quality measurement experiments with focus on non-user aspects, we will investigate how search engines and other information sources like Wikipedia [10], quality of results and usability are perceived by the information seeker when carrying out exploratory search tasks. Therefore we will look at the whole search task, not only at the individual query (for an explanation of the concepts task, query and session please refer to Section 3).

2. RELATED WORK

This section gives an overview of the scientific work in the areas of classical search engine performance measurement and user search experience research. Both fields are strongly connected when it comes to search process quality measurement as an integrated approach (machine and user). As B. Jansen stated, there is a tension between information searching and information retrieval. Despite their partly contradictory constructs, a trend towards convergence of both can be noticed [12]. Retrieval measures to measure the performance and quality of information retrieval systems have been used for more than 50 years. A set of new web specific measures has been developed, yet those are still limited in giving a comprehensive quality indication of present search services [23]. Lewandowksi and Hochstötter [17] proposed a search engine quality measurement framework that both reflects the more system-centric approach and the user-centric approach by taking index quality, quality of the results, quality of search features and search engine usability into consideration. Their main point of reasoning is that for measuring the user experience, empirical studies are indispensable. One side of the research methods to capture the user experience are laboratory experiments. A user test group is collected and the experiment is carried out in a laboratory environment. During these experiments sample sizes are typically small and therefore they are not very representative. A complementary approach is log file analysis. Although the just mentioned shortcomings of small sample sizes do not exist in the log file analysis, there exists another one instead: Log files are anonymous data files that lack any sort of additional demographic information about the user like gender, age or profession [9]. In addition, log files only include data gathered from a specific web site, e.g. when considering a search engine, one can collect data on all the interactions with the search engine, but not the interactions taking place after the user leaves the search engine and examines the results. An approach that is positioned in between log file analysis and surveys is experience sampling via a browser plug-in. The plug-in is installed locally and logs the behavior of the user (in search engines, as well as on other websites visited) but can also be used to gather explicit user feedback by integrating questionnaires in a structured way. Fox et al. did experiments with a browser plug-in in 2005 [7]. They also added some automated questions to their experiments to collect user feedback. With this approach they got explicit feedback about the search engine performance on query level as well as on session level. From the session log files Fox et

al. gathered implicit measures like time spent on page and time to first click. Each of the approaches of classical search engine evaluation covers some necessary aspects of an integrated search engine quality measurement framework, but none covers all. The shortcomings are mostly due to too little focus on the user. Therefore we also screen the research on evaluating exploratory search systems.

Exploratory search covers a different class of activities than classic information retrieval (IR). The latter mainly follows the query-response paradigm and (only considers a query and its resulting documents). Exploratory search comprises more complex activities like investigating, evaluating, comparing, and synthesizing, where new information is sought in a defined conceptual area. As exploratory search systems (ESSs) strongly rely on human interaction, the evaluation of ESSs is an even more complex task [14]. Although an adequate set of measures has been found for search engine evaluation, the same is still missing for ESSs. In IR the Cranfield methodology [4] was used to objectively compare IR systems. It was also later used by the Text Retrieval Conference (TREC) as the main paradigm for its activities of large-scale evaluation of retrieval technologies [24]. While TREC is good at measuring technical aspects like underlying algorithms, it does not sufficiently take the human interaction factor into account. Researchers are still struggling to integrate the user into the measuring process (as in TREC Interactive Track [6]). The main issue so far is the repeatability and comparability of experiments on different sites. When compared to classical IR (where relevance is the main performance measure), covering as many aspects of a topic as possible is equally important as relevance in an exploratory search context [21]. As exploratory search sessions can span over days and weeks, long term studies are indispensable [14]. Sample sizes in ESS evaluation studies are usually small. This limits the generalizability of the findings. Measures as in classic IR are not applicable to ESSs. Researchers have suggested to develop special measures for ESSs [1, 19] as the users cannot be neglected due to their high involvement in the search process. A workshop held by Ryen White et al. in 2006 [28] produced the following measures as appropriate for evaluating exploratory search systems: *Engagement and enjoyment*, *information novelty*, *task success*, *task time* and *learning and cognition*. As a consequence, our combined approach aims to cover both, the technical side as well as the user experience aspects.

3. IMPLEMENTATION

Search-Logger is an experimentation environment especially designed to carry out exploratory search task experiments. During the development phase our goal was to optimize the process (machine plus user interaction) as a whole. Measuring exploratory search tasks is a more complicated endeavor then measuring search at the query level [14]. We have created Search-Logger's architecture around the definition of a search task as an open-ended, abstract and poorly defined information need with a multifaceted character [18].

Such an exploratory search task typically consists of various sessions. A session is "a series of interactions by the user toward addressing a single information need" [13]. A query is a string of terms typed into a search engine and a list of documents is retrieved. As mentioned, the major part of the search engine quality research is focused on the query level and this is a limitation that we want to resolve with our
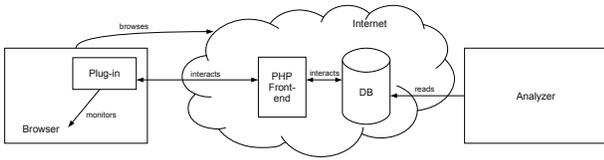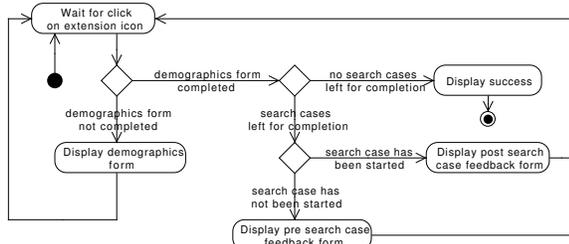
Figure 1: Architecture

Figure 2: Activity diagram

Search-Logger. As proposed by White et al. in 2006 [28] we will use the measures *learning and understanding*, *task success* and *task time* in our studies to extend the scope from IR related search needs to broader exploratory search tasks. Our main parameters will be search task complexity and its impact on the search process performance.

Architecture. The architecture of the Search-Logger framework is shown in Figure 1. Similar to the approaches by Fox et al. [7] and Lemur's Toolbar [5], the Search-Logger also consists of a browser plug-in for Firefox, a remote log storage database and an analysis environment. After activation, the plug-in monitors actions accessing the Internet in the browser and sends these to the PHP front-end of the database. Also initial information for starting new search tasks is read via this front-end from the database. The analyzer component accesses the database directly to allow the evaluation of the search results. The coarse behavior of the plug-in component is visualized in Figure 2. Depending on if the Search-Logger has been used before, if demographics has been filled in, and what search cases were started and finished, it displays different forms and monitors the ongoing search task. It fulfills the following three main tasks: (i) deliver the precompiled search tasks to the users, (ii) gather implicit information about the search process by logging various browser events as outlined in the next paragraph, (iii) gather explicit user feedback via standardized questionnaires supplied before and after each search task. With this approach we manage to log the search process on the search task level. Each logged event is ear-marked for a certain search task. Therefore the task performance can be analyzed and evaluated.

Data. With the Search-Logger experimentation framework we can automatically log implicit user data and also gather explicit user feedback at the same time.

Implicit user data about the search process is logged on all sort of standard user events like links clicked, queries entered, tabs opened and closed, bookmarks added and deleted and clipboard events as illustrated in Table 1. In terms of

Type of event	Explanation
Search task started (a)	Log entry for start of search task
Search task stopped (b)	Log entry for end of search task
Search experiment started	Start of search experiment
Search experiment stopped	End of search experiement
Web page visited	User visited a web page
Query entered	User entered a query
Link clicked	User followed / clicked on a link
Tab opened	User opened a tab
Tab closed	User closed a tab
Bookmark added	User created a bookmark
Bookmark deleted	User deleted a bookmark
Text copied	Copy/Paste event
Search logger paused	Pausing the search logger extension
Introducing new user	Extension started for the first time
Demographics displayed (c)	Demographics form displayed
Demographics submitted (d)	Demographics form submitted
Pre search task form submitted (e)	Pre search task feedback form submitted
Post search task form submitted (f)	Post search task feedback form submitted

Table 1: User events logged by Search-Logger

implicit logging, all comparable tools log similar events and therefore we will not go into every detail about that here. We focus on the implementation of the exploratory search task evaluation support instead. When the proband selects the first search case, an entry consisting of time and date is logged into the database. This also marks the start of the experiment. Until this search task is finished by the user, each log entry carries the token of this case and is identified accordingly. Next, the proband can start to carry out the search task. The user can always pause the experiment. This will also pause the Search-Logger till the user resumes the experiment. Those pause and start events are logged and allow the creation of realistic search scenarios that can span days and weeks.

Explicit data is gathered by asking the probands to fill in questionnaires at the beginning of the experiment and before and after each search task. Those questionnaires can be designed freely and will most often contain a set of standardized questions together with free form fields. Gathering demographic information about the user at the beginning of the experiment enables us to classify the users according to gender, Internet usage patterns and previous search experience. The explicit user feedback gathered before and after each case is aimed towards gathering information regarding task success, learning, cognition and enjoyment as the main measures for evaluating exploratory search systems [28] as outlined in Section 2. Table 1 lists the events that are logged by Search-Logger. Especially the task defining events (a)-(f) are unique to Search-Logger's task based
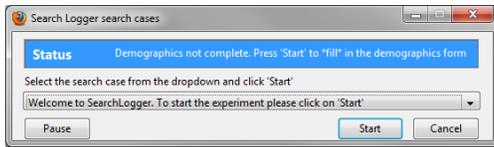
Figure 3: Dialogue window

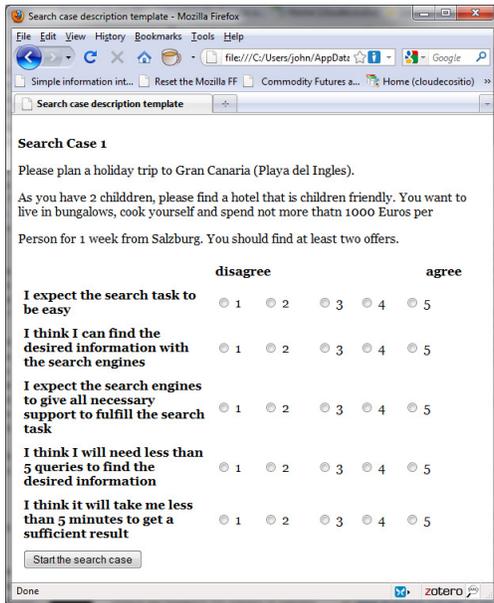| Study sample	10 participants
Gender	6 men, 4 women
Age	24-36 years
Backgrounds	academic staff(5), highschool teachers(2), students(2), sales reps(1)

Table 2: Sample description

Figure 4: Screenshot of questionnaire

measuring approach. This functionality is not built into any of the comparable tools discussed in the related tools sub section.

User Interfaces

The users interact with Search-Logger via a little icon that appears in the status bar of the browser at bottom right. A click on this icon opens up a window as illustrated in Figure 3. In this window the user can start, stop and pause search tasks. Depending on the chosen state, the icon in the browser status bar either blinks if in logging mode or shows a green pause otherwise.

The questionnaires as illustrated in Figure 4 are implemented as HTML pages and can be edited with a standard HTML editor.

Related tools

We have identified several other tools designed mainly for the same purpose of logging user events. The first tool to look at, is called "Wrapper" by Bernard J. Jansen [11]. The Wrapper tool was developed for the purpose of logging events of all applications used by an information seeker (including applications like Microsoft Office or email clients). As Wrapper is a desktop application it can log any event, i.e. what programs are started, which processes are running or what was copied to the clipboard. It was designed with the focus on evaluating exploratory search systems

Although this broad logging functionality is an advantage, the software does not have the possibility to supply precompiled search cases to a group of probands and does not collect explicit user feedback. Wrapper also does not have the functionality built in, to relate a set of logs to a certain search task. The concept of search task does not exist in this approach. Another tool is a browser plug-in that was created by Fox et al. in 2005 [7]. Fox's approach was implemented as an add-on for the Internet Explorer. It was the first tool to gather explicit as well as implicit information during searches at the same time. It evaluates the query level and gathers explicit feedback after each query. Fox's approach also came with a sophisticated analysis environment for the logged data. Unfortunately Fox's IE add on is not publicly available. A third tool is Lemur's Query Log Toolbar" [5] which is a toolkit implemented as a Firefox plug-in and Internet Explorer Plug-in. It logs implicit data on the query level. Logging exploratory search tasks is not implemented. Further tools, we will not discuss here are: The HCI browser [2], The Curious Browser [3], WebTacker [22] and Weblogger [20]. Some of those are either discontinued, or do not fulfill our requirements enough to be included in our comparison. All tools mentioned, have some features with the Search-Logger (to be described in the next section) in common (like the logging of user triggered events), while some aspects are not covered at all. The main and foremost difference comes from the different purposes the tools were designed for. Most of the tools were developed for evaluating the query level while the Search-Logger was purely developed for evaluating the exploratory search task level. None of the tools have the built-in functionality to have a precompiled set of exploratory search tasks carried out by a test group in a non laboratory environment without any time constraints. We hypothesize that realistic user study results for exploratory search experiments will only be possible if probands do not have any time constraints and can search in a way and in a surrounding they are used to. Only then new phenomena like social search (e.g. asking peers on Facebook) will be showing up in the observations.

4. EVALUATION

We carried out a pilot study to gauge the effectiveness of the Search-Logger framework. We compiled an experiment consisting of seven search tasks. The non look-up tasks were designed in accordance with the rules about designing exploratory search tasks for user studies as stated in a paper by Kules and Capra [16]. The plug-in was distributed to a convenience sample of 10 people as illustrated in Table 2.

They could install the plug-in wherever they wanted and were free to work on the experiment at their choice. They were not tied to any location, narrow time frame or system. The study was conducted over a 4 week period. Of the seven tasks, three were intended to be look-up tasks and four were exploratory search tasks of increasing complexity. Due to

many wedding checklists directly retrievable with one query, we reclassified task (iv) as look-up task. The tasks were as follows:

(i) "Please find the date when Mozart was born."

(ii) "What is the corresponding value of 20.000 Estonian Kroons in USD?"

(iii) "When was Penicillin invented and by whom?"

(iv) "How do you plan a budget but still nice wedding; write down the 20 most important points to consider."

(v) "Your daughter Anna is graduating from high-school and wants to study abroad in the field of either political sciences or architecture. You can support her with 2000 Euro per month. Which universities can you recommend that are affordable and will offer Anna the best career perspectives after graduating? Please compile a list of 10 universities that you can recommend."

(vi) "Mr Johnson and his family are planning a trip to Paris. Please write a summary of the following: What cultural events are there (for example in August 2010)? What sightseeing to do? Cover also hotels, flights, traveling to Paris, weather."

(vii) "Barbara has been offered a well paid position in Kabul, Afghanistan. She wants to know, how risky it is to live in Kabul at the moment. Compile enough information to have a good indication for the risk. Please compile at least half a page of compelling facts."

Due to the qualitative nature of this study and the small sample size we will only present rough numbers and interesting qualitative results. The longest experiment was 15 days and the longest net search time (when the search logger was recording) was 3.4 hours. The shortest net search time was 1h and 8 minutes. All users visited a maximum of 2 web pages and needed a maximum of 3 queries for each of the look-up tasks and performed the search with opening a maximum of 2 additional tabs in their browsers. Each of the tasks could successfully be fulfilled with search engines within a little over three minutes. The explicit user feedback (gathered before and after the tasks) showed no inconsistencies (apart from a slight underestimation of the effort of task (iii)). This reflects that users are able to judge the search effort for simple tasks quite well. As expected this combination of implicit and explicit feedback confirms that search engines are effective and efficient tools for look-up tasks.

The log for the tasks (v), (vi) and (vii) tells a different story. Per task users visited a maximum of 157 web pages and submitted a maximum of 28 queries to search engines. In addition users added a maximum of two bookmarks and opened and closed up to 54 tabs per search task. 88% of searches in search engines were performed in Google, 9% in Bing and the rest in local search engines. More experimental data can be retrieved from our webpage `http://www.search-logger.com/tiki-read_article.php?articleId=1` upon registration.

It is clearly visible that search engines are the means of choice to start a search as also observed in [29]. After being directed to a potentially interesting site, users spend most of the time for analyzing, synthesizing and information gathering away from search engine territory. We also compared the time users spent on searching with search engines to the time users spent on searching with other information sources. For each of the complex tasks less than ten percent of the time was spent on entering and rephrasing queries and residing on the search engine results pages. In terms of

userID	Action	Information	IP	Date
10	User viewed a related link	Location: http://www.facebook.com/...	90.191.164.56	2010-08-11 11:37:56
10	User viewed a related link	Location: http://0.75.channel.facebook.com/iframe/11?r= http%3A%2F%2Fstatic.ak.fbcdn.net%2Frsrc.php%2 FzBNLI%2Fh...	90.191.164.56	2010-08-11 11:37:56
10	User viewed a related link	Location: http://static.ak.facebook.com/common/redirect iframe.html...	90.191.164.56	2010-08-11 11:37:55
10	User viewed a related link	Location: about:blank...	90.191.164.56	2010-08-11 11:37:55
10	User viewed a related link	Location: about:blank...	90.191.164.56	2010-08-11 11:37:55
10	User performed a search	Location: http://www.google.ee/search?hl=et&q=college+% 22political+sciences%22+tuition+comparison&aq	90.191.164.56	2010-08-11 11:37:19

Figure 5: Switch to social search in Facebook

explicit feedback, we noticed that users tend to underestimate the time needed for exploratory search tasks and overestimate the power of search engines. Most users strongly agreed "I think I can find the desired information with search engines only" before the tasks and corrected their judgement after the task. One user spent almost 3 hours on case (v) and out of frustration navigated to Facebook (as shown in Figure 5) to shortcut his search and ask for recommendations from friends. Overall, users found the Search-Logger "easy to use" and liked the freedom to carry out the experiment in a non laboratory environment. Two users commented that they would maybe search a bit differently (with more engagement) if they really depended on the reliability of the information. We must therefore assume that the data does not fully reflect perfectly realistic user behavior.

Six users out of ten finished all search tasks, one user finished 6 tasks, and three users only finished the first 3 look-up tasks.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we are presenting an experimentation toolkit called Search-Logger for the analysis of exploratory search tasks. We introduce the difference between search queries, search sessions, and search tasks. For the proposed experimentation toolkit, we have already developed a Firefox plug-in and a database installer and have described them and their features in this paper. Furthermore, we have shown the results of a first pilot study carried out with the Search-Logger. The main discoveries are that new phenomena like using various search systems in parallel over an extended period of time and social search [8] (e.g. in Facebook) are poorly covered by classic search engine quality experiments. Especially when search tasks become complex, users obviously have a tendency to rather ask an expert or a peer than spending too much time searching themselves. With this experiment we could also prove that the Search-Logger approach enables us to gauge the main measures for exploratory search systems, engagement and enjoyment, task success and task time as suggested by R. White [28]. We have lined up several field-tests (with bigger statistically relevant sample sizes). The results of these experiments will be published in several follow-up papers. Using the demographic information, we hope to find relations between search tasks and specific users to make suggestions for improvements for targeted domain specific search.

6. REFERENCES

[1] P. Borlund and P. Ingwersen. Measures of relative relevance and ranked half-life: performance indicators for interactive IR. In *Proceedings of the 21st annual*

international ACM SIGIR conference on Research and development in information retrieval, page 324–331. ACM New York, NY, USA, 1998.

[2] R. Capra. HCI browser: A tool for studying web search behavior. Available from: `http://uiir-2009.dfki.de/papers/uiir2009_submission_18.pdf` [cited September 10, 2010].

[3] M. Claypool, P. Le, M. Wased, and D. Brown. Implicit interest indicators. In *Proceedings of the 6th international conference on Intelligent user interfaces*, page 40, 2001.

[4] C. W. Cleverdon, J. Mills, and E. M. Keen. An inquiry in testing of information retrieval systems.(2 vols.). *Cranfileld, UK: Aslib Cranfield Research Project, College of Aeronautics*, 1966.

[5] P. Clough and B. Berendt. Report on the TrebleCLEF query log analysis workshop 2009. In *ACM SIGIR Forum*, volume 43, page 71–77, 2009.

[6] S. T. Dumais and N. J. Belkin. *TREC Experiment and Evaluation in Information Retrieval, chapter The TREC Interactive Track: Putting the User Into Search*. MIT Press, 2005.

[7] S. Fox, K. Karnawat, M. Mydland, S. Dumais, and T. White. Evaluating implicit measures to improve web search. *ACM Transactions on Information Systems (TOIS)*, 23(2):147–168, 2005.

[8] J. Freyne, R. Farzan, P. Brusilovsky, B. Smyth, and M. Coyle. Collecting community wisdom: integrating social search and social navigation. In *Proceedings of the 12th international conference on Intelligent user interfaces*, pages 52–61, Honolulu, Hawaii, USA, 2007. ACM. Available from: `http://portal.acm.org/citation.cfm?id=1216295.1216312`, `doi:10.1145/1216295.1216312`.

[9] C. Grimes, D. Tang, and D. M. Russell. Query logs alone are not enough. In *Workshop on Query Log Analysis at WWW*. Citeseer, 2007.

[10] M. Hu, E. Lim, A. Sun, H. W. Lauw, and B. Vuong. On improving wikipedia search using article quality. In *Proceedings of the 9th annual ACM international workshop on Web information and data management*, pages 145–152, Lisbon, Portugal, 2007. ACM. Available from: `http://portal.acm.org/citation.cfm?id=1316926`, `doi:10.1145/1316902.1316926`.

[11] B. J. Jansen, R. Ramadoss, M. Zhang, and N. Zang. Wrapper: An application for evaluating exploratory searching outside of the lab. *EESS 2006*, page 14.

[12] B. J. Jansen and S. Y. Rieh. The seventeen theoretical constructs of information searching and information retrieval. *Journal of the American Society for Information Science and Technology*, pages n/a–n/a, 2010. Available from: `http://onlinelibrary.wiley.com/doi/10.1002/asi.21358/abstract`, `doi:10.1002/asi.21358`.

[13] B. J. Jansen, A. Spink, C. Blakely, and S. Koshman. Defining a session on web search engines. *Journal of the American Society for Information Science and Technology*, 58(6):862–871, 2007. Available from: `http://onlinelibrary.wiley.com/doi/10.1002/asi.20564/full`, `doi:10.1002/asi.20564`.

[14] D. Kelly, S. Dumais, and J. Pedersen. Evaluation challenges and directions for information seeking support systems. *IEEE Computer*, 42(3), 2009.

[15] W. Kraaij and W. Post. Task based evaluation of exploratory search systems. *EESS 2006*, page 24.

[16] B. Kules and R. Capra. Designing exploratory search tasks for user studies of information seeking support systems. In *Proceedings of the 9th ACM/IEEE-CS joint conference on Digital libraries*, pages 419–420, Austin, TX, USA, 2009. ACM. Available from: `http://portal.acm.org/citation.cfm?id=1555400.1555492`, `doi:10.1145/1555400.1555492`.

[17] D. Lewandowski and N. Höchstötter. Web searching: A quality measurement perspective. *Web Search, Information Science and Knowledge Management*, 14, 2008.

[18] G. Marchionini. Exploratory search: from finding to understanding. *Communications of the ACM*, Volume 49(4):Pages 41–46, 2006.

[19] H. O'Brien. *Defining and Measuring Engagement in User Experiences with Technology*. Unpublished doctoral dissertation, Dalhousie University, Halifax, Canada. 2008.

[20] R. W. Reeder, P. Pirolli, and S. K. Card. Webeyemapper and weblogger: Tools for analyzing eye tracking data collected in web-use studies. In *CHI'01 extended abstracts on Human factors in computing systems*, page 20, 2001.

[21] D. Tunkelang. Precision AND recall. *IEEE Computer*, 42(3), 2009.

[22] D. Turnbull. Webtacker: A tool for understanding web use. *Unpublished report. Retreived on May*, 18(2009):19–23, 1998.

[23] L. Vaughan. New measurements for search engine evaluation proposed and tested. *Information processing and management*, 40(4):677–691, 2004.

[24] E. M. Voorhees and D. K. Harman. *TREC: Experiment and evaluation in information retrieval*. MIT Press, 2005.

[25] S. Weitz. Search isn't search. Microsoft company report, Microsoft Corporation, SMX 2009, Munich.

[26] R. W. White, B. Kules, and S. M. Drucker. Supporting exploratory search, introduction, special issue, communications of the ACM. *Communications of the ACM*, 49(4):36–39, 2006.

[27] R. W. White, G. Marchionini, and G. Muresan. Evaluating exploratory search systems: Introduction to special topic issue of information processing and management. *Information Processing & Management*, 44(2):433–436, Mar. 2008. Available from: `http://www.sciencedirect.com/science/article/B6VC8-4R2H21N-3/2/9c2f50439d974562ab4fa0504e0bb865`, `doi:10.1016/j.ipm.2007.09.011`.

[28] R. W. White, G. Muresan, and G. Marchionini. Report on ACM SIGIR 2006 workshop on evaluating exploratory search systems. In *ACM SIGIR Forum*, volume 40, page 52–60. ACM New York, NY, USA, 2006.

[29] R. W. White and R. A. Roth. Exploratory search: Beyond the Query-Response paradigm. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, 1(1):1–98, 2009.