Real-World User Evaluations of a Visual and Interactive Web Search Interface

Orland Hoeber
Department of Computer Science
Memorial University
St. John's, NL, Canada A1B 3X5
hoeber@cs.mun.ca

Daniel Schroeder and Michael Brooks
Department of Computer Science
University of Regina
Regina, SK, Canada S4S 0A2
{schroedd, brooks1m}@cs.uregina.ca

Abstract

Although laboratory user studies are the most common method for validating the utility of information visualization systems, it may be difficult to determine if such studies accurately reflect the tasks of real users. In this paper, we describe a longitudinal real-world user evaluation of a visual and interactive Web search interface designed to support exploratory searching: $theHotMap.com$. Although we experienced some attrition during the course of the four-week study, the results indicate that many of the features of the system are beneficial and useful in real-world use. Although exploratory search comprised a small portion of the total search activities undertaken, the features of $theHotMap.com$ which support query refinement and search results exploration did not interfere with the more common targeted searching tasks. As a result, most participants indicated that they found the system useful and easy to use.

1 Introduction

In recent years, the technologies underpinning Web search engines have matured to the point that Web search is considered by many to be a solved problem. However, the interfaces used by the top search engines have changed very little in this time. While a simple query box and a ranked list of search results may be all that is needed for targeted searches, these interfaces do not adequately meet the needs of searchers in many other situations (e.g., when their goal is not well defined or ambiguous, or when there is a desire to explore the breadth of information available for a specific topic).

For such search tasks, providing visual and interactive support for the fundamental activities of crafting a query and exploring the search results has the potential to greatly improve the ability of searchers to find the information they are seeking. $TheHotMap.com$ is a publicly available Web search interface designed to support the human elements of searching through the use of lightweight visual enhancements to the commonly-used list-based representation of Web search results. Information that can assist searchers in their Web search tasks is presented in a visual manner. The system supports user interaction both during query refinement and search results exploration processes.

In this paper, we provide a brief overview of the features of TheHotMap.com (see Figure 1, and [11] for more details). The system is a re-implementation and extension of some of our previously published research prototypes for visual and interactive Web search [12, 13]. The goal is to employ easy-to-use interface elements that allow users to visually interpret and make sense of the fundamental Web search information, and take an active role in the Web search process, when necessary.

The primary contributions of this paper include the longitudinal research methodology, and the results from real-world user evaluations conducted over a four-week period. Our focus is not only on how the features of the system are being used, but also the participants' impressions of usefulness and ease of use. We contend that the utility of a visual element in an interactive interface is closely related to its effectiveness in visually representing the underlying data.

The remainder of this paper is organized as follows. An overview of related work is provided in Section 2. A brief summary of the features of $theHotMap.com$ are outlined in Section 3. Section 4 describes the longitudinal research methodology used in this study. Section 5 presents the results of the study. Conclusions and future work are discussed in Section 6.

2 Related Work

The system under evaluation in this study is based on a combination of two of our previous research prototypes: HotMap [13] and Wordbars [12]. Both of these prototypes were originally developed with the purpose of exploring visual representations and the use of various types of informa-
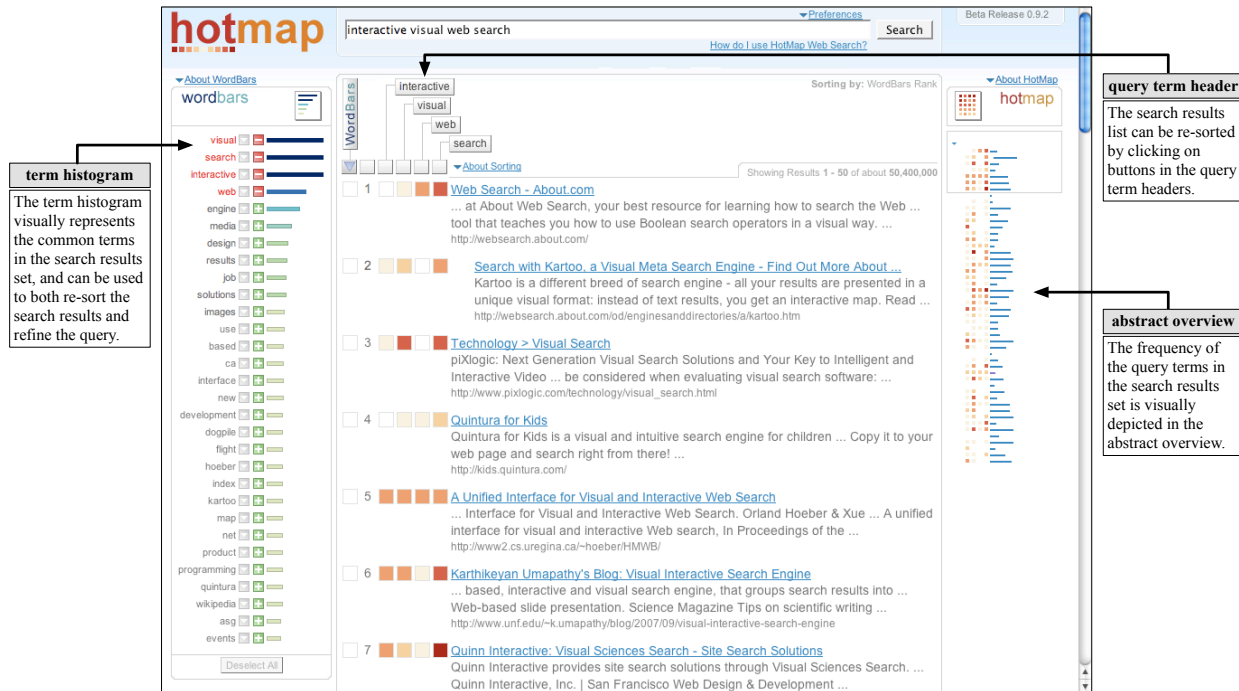
Figure 1. A screenshot of $TheHotMap.com$**. Note the lightweight controls labeled in the figure.**

tion to support Web search activities. As research tools, the prototypes were useful for validating the potential utility of the proposed techniques [14, 15]. However, they were not designed for public use. $TheHotMap.com$ is a complete re-implementation and extension of the methods described in these previous works.

Others have explored the use of visual interfaces to support the exploration of Web search results. Heimonen and Jhaveri [9] created an icon-based representation of the location of specific query terms within individual search results sets. Based on TileBars [8], this system allowed the searcher to see where in the resulting documents their search terms were being used together.

In VIEWER [2], the frequency of all combinations of the query terms were counted within the document surrogates. This information was presented in a histogram representation. Selection within the histogram allowed the searcher to filter the search results set based on specific combinations of the query terms.

Web search clustering systems, such as Clusty [4] and Grokker [7], dynamically identify and label clusters of documents discovered within the search results sets. Normally presented as a tree-based representation, users can expand and select clusters, resulting in a filtering of the search results set. Kules [18] extended the standard paradigm for clustering search engines by providing consistent naming of the clusters. The result is a system that allows users to learn the names and meanings of the clusters over time.

In terms of supporting users in crafting accurate Web search queries, very little work has been done. Joho et al. [17] developed a method for dynamically generating a hierarchy of potentially useful terms for query expansion, and presented these in a menu-like structure. Although the focus is on filtering RSS feeds, VisGets [6] provides a set of coordinated visual representations designed to support the searcher in dynamically refining their query. Some of the top search engines and Web browsers have begun to address this issue by suggesting potential queries based on their usage logs. This feature appears in both Ask [1] and Yahoo [22], as well the Firefox [19] Web search bar.

3 theHotMap.com

The version of $theHotMap.com$ tested in this study is implemented as a Web search interface layer overtop of the search results provided by the Yahoo API [21]. There are three main features that enable flexible interaction within the system: the *term histogram*, the re-sortable search results list using the *query term headers*, and the *abstract overview* of the full search results set (see Figure 1). These features were selected and refined based on the results of user studies conducted with the original HotMap and Word-Bars prototypes [14, 15].

The *term histogram* provides a visual representation of the most frequently appearing terms within the search results set (using the title, snippet, and URL), allowing the rel-

ative frequency of these terms to be easily observed. Users can interactively re-sort the search results set by selecting the arrow icon beside any term of interest. Interactive query refinement is supported by clicking the plus icon beside any term users wishes to add to their queries, or the minus icon beside any term users wish to remove from the query.

In addition to the re-sorting supported via the *term histogram*, searchers may also re-sort the search results based on the use of their specific query terms. Clicking on any of the *query term headers* above the search results list will cause the search results to be re-sorted based on the use of the selected term within the titles, snippets and URLs. Although the default sorting method is to perform single-term sorting, an advanced feature is available that supports nested sorting.

The *abstract overview* provides a compact visual representation of the entire set of search results that are presented in the list-based representation. In the current implementation, the system collects 50 search results per page. Colour coding is used to represent the frequency of the query terms within the search results overview. This same colour coding is also used in the search results list. The *abstract overview* supports an interactive exploration of the search results. As users visually identify documents of interest, they may click on the abstract representation of the search result to cause the search results list to scroll to that location. The system temporarily highlights the corresponding search result that was selected in the *abstract overview*, allowing users to easily relate their selection in the overview to the scrolled location in the search results list.

4 Real-World User Evaluations

Our previous studies on the research prototypes were conducted in a controlled laboratory setting using pre-assigned exploratory search scenarios [14, 15]. Although these tasks were designed to mimic a sub-class of real-world search activities (e.g., exploratory searches), and were presented as simulated work tasks [3], one must always question how well they represent the types of searches normally conducted. To address this concern, we have designed and conducted a longitudinal evaluation of $theHotMap.com$ in a real-world search setting.

4.1 Longitudinal Research Methodology

Longitudinal studies are designed to allow participants to engage in learning and using a system under investigation over an extended period of time. Such studies may span multiple days or weeks. During this time, the participants make exclusive use of the system for all activities related to the target tasks that are supported.

Data collection methods may include logging the usage activities of the participants, administering questionnaires, conducting interviews, and/or conducting focus groups. The usage logs allow the data to be collected remotely and automatically. Daily, biweekly, or weekly questionnaires (perhaps administered online) allow the subjective reactions of the participants to be measured throughout the course of the study. Interviews and focus groups can also be conducted throughout the study period to capture qualitative impressions on the utility of the system within the real-world setting.

Longitudinal studies are especially well-suited to the evaluation of visual Web search interfaces. Since searching the Web is the type of activity that occurs in the workplace, school, and home, and at various times throughout the day, laboratory evaluations and field trials may not capture all aspects of use. Longitudinal studies provide valuable insight into the utility of the visual representations and interaction throughout the full breadth of Web search activities. More details on benefits and drawbacks of longitudinal studies, and how they can be used as part of a comprehensive evaluation methodology, can be found in [10].

4.2 Study Design

This longitudinal study of the real-world use of $theHotMap.com$ spanned a four-week period. At the beginning of the study, an information session was held to introduce the participants (n=13) to the system with a short demonstration and explanation of the study procedures. A pre-study questionnaire was administered to measure demographics and Web search behaviours. Each participant was provided with a brochure outlining the key features of the system.

Immediately following the information session, the participants were asked to start using a custom version of TheHotMap.com as their primary Web search engine for the duration of three weeks. At the end of each week, an online questionnaire was administered to gauge the participants' impressions of the usefulness and ease of use of the system. These questionnaires were based on the *technology acceptance model* [5].

After the third week of searching was completed, the participants were informed that they were no longer required to use the system as their primary search engine. One week later, the participants were invited to an exit session, during which a final questionnaire regarding the acceptance of the system was administered. This final questionnaire also included questions regarding the types of searches that were conducted with the system, and whether the participants continued to use the system after the conclusion of the required use period.

All the participants were financially compensated for

their participation in the study. Half of the compensation was distributed at the information session; the other half was distributed at the exit session.

4.3 Usage Logs

During the three week period in which the participants were using the system as their primary Web search engine, all the queries and use of the interactive features of the system were logged. Whenever the participants activated or selected a feature of the system (e.g., submitting a query, re-ordering the search results, adding a term to the query, clicking on a search result, etc.), the relevant information was sent back to the server via an AJAX request, where it was matched to the activity session and stored in a usage log.

A study design goal was to ensure that the system would be used in as normal a manner as possible by the participants. To achieve this goal, we ensured the anonymity of the participants by not requiring them to login or identify themselves to the system. Activity session cookies were used to match an individual's queries and interactions to one another. However, by design, it was not possible to match the queries and interactions back to an individual participant.

5 Results and Analysis

5.1 Participant Demographics

Participants in this study were drawn from the undergraduate population of our university. Thirteen individuals attended the information session and began the study. Nine of these participants were male, four were female; nine were studying science, three were studying engineering, and one was studying arts. All participants self-reported a high degree of computer expertise, and all reported using Google as their primary Web search engine. Eight of the participants reported conducting four or more searches per day, three reported conducting three or more searches per day, and two reported conducting two or fewer searches per day. Over the course of the four weeks, two participants chose to discontinue their involvement prior to the completion of the study.

5.2 Usage Logs Analysis

5.2.1 Queries

Table 1 shows the number of unique queries submitted during each of the three weeks of the study. Although our goal was to keep the participants actively involved in the study through the weekly questionnaires, there is a clear case of attrition as the study progressed. However, even with this

Table 1. The number of unique queries submitted over the course of the study period.

	Week 1	Week 2	Week 3
Unique Queries	669	413	275
Average Queries per Participant	51	32	21

attrition, the participants still submitted a similar number of queries per day (two to four) to what they reported in the pre-study questionnaire.

5.2.2 Re-Sorting Operations

There were two methods by which participants were able to re-sort the search results: using the *term histogram*, and the *query term headers*. Table 2 shows how many times these re-sorting features were used during the study. Of the total number of unique queries during the three week search period, in only 7% did the participants choose to make use of the re-sorting operations. Clearly, the re-sorting features were not used as heavily as we had anticipated. This may have been due to a number of aspects including attrition, fewer exploratory searches, and participants reverting to using the system like a non-interactive Web search engine.

Of particular interest is whether those who did use the re-sorting features were able to bring more relevant documents to the top of the search results list. We measured this by investigating which search results were viewed by the participants when re-sorting was involved. For each of these search results, if it was moved higher in the list than the underlying search engine would have dictated, we deemed this a *positive change*. If it was moved lower in the list, we deemed this a *negative change*. Table 3 shows the results of this analysis.

In all but the third week, we can see a clear benefit to the re-sorting operations. In almost all cases, search results that were identified as potentially relevant by the participants (i.e., clicked) were moved to a higher location in the search results list as a result of the re-sorting operations. During the third week, the number of re-sorting operations dwindled to a total of nine, producing mixed results in this analysis.

5.2.3 Query Refinement Operations

The *term histogram* provides searchers with the ability to add or remove terms from the query with a single click. Table 4 shows the number of queries that were refined using the system, as well as the number of queries that were manually refined. The manually refined queries were identified as having at least one term in common with the previous

Table 2. The number of re-sorting operations initiated by the users, (in brackets are the number of unique queries that were affected).

	Week 1	Week 2	Week 3
Term Histogram	78 (22)	43 (16)	5 (3)
Query Term Headers	91 (38)	44 (15)	4 (4)

Table 3. The number of viewed search results that were affected by the re-sorting of the search results list.

	Week 1	Week 2	Week 3
Positive Change	29	34	2
No Change	6	0	0
Negative Change	1	3	2

Table 4. The number of refined queries using the term histogram and manual refinement.

	Week 1	Week 2	Week 3
Term Histogram	36	10	3
Manual Refinement	280	180	98

Table 5. The frequency of use of the scroll and jump operations using the abstract overview.

	Week 1	Week 2	Week 3
Scroll	74	9	0
Jump	89	6	7

query in the activity session, and with no use of the query refinement operations in between.

The large number of manually refined queries (41% of all the queries over the three week search period) is a noteworthy finding. This is nearly twice that of what has been reported in studies of Web search engine logs [16]. We suspect that participants were recognizing relevant terms in the *term histogram*, and manually adding these to their queries rather than using the tools provided. A further analysis of the phenomenon is warranted.

5.2.4 Visual Inspection and Scrolling Operations

The *abstract overview* feature provided searchers with the ability to visually inspect an abstract representation of the search results set, and easily scroll or jump to a location of interest. Table 5 shows the frequency of use of these operations during the study.

Clearly, the features were being used heavily in the first week, and then almost not at all in the remainder of the study. This change in participant behaviour is similar to what was seen in the other interactions: heavy use of the features in the first week, followed by reduced use in the second and third week. With respect to these scrolling operations, it is likely that the participants were trying them out in the first week, and then continued to scroll in the way they are accustomed (using the regular scroll bar or the scroll-wheel on their mouse) for the remainder of the study. It appears that the abstract visual representation used here may not convey enough information to support identifying individual documents to analyze further.

5.3 Questionnaire Analysis

5.3.1 Usefulness and Ease of Use

Our expectations while planning this longitudinal study was that we would see a change in the responses as the participants became more familiar with the system. As one can see from Figures 2 and 3, this was not the case. Clearly, the participants were able to learn to use the system and become effective with it within the first week of use (prior to completing the first questionnaire).

Of special note is the third question in Figure 2 pertaining to the productivity of the searcher. Here, we see a steady positive increase over the three weeks of the study, and then a significant drop once the participants were no longer required to use the system (in this case, nearly half of the participants indicated disagreement with the statement). A similar drop can be seen in the last questions in both Figures. Our intuition is that this drop is a result of participants recalling the negative aspects of their experiences with the system more readily than the positive aspects.

In general, the responses to the questionnaire were positive. There was a strong neutral bias to the questions regarding the usefulness of the system. However, this bias leaned more towards agreeing with the statements. With the ease of use questions, the responses almost uniformly ranged from neutral to strongly agree. Even though we saw the use of the features drop during the second and third week of the study, the participants continued to report similar responses regarding the usefulness and ease of use of the system.

5.3.2 Changes Over Time

The variability of the responses to the questions regarding the usefulness and ease of use were analyzed using a chi-squared test for independence based on the week-to-week
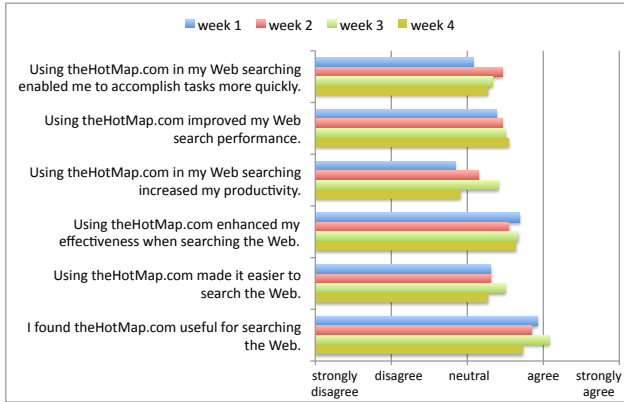
Figure 2. Average responses to the questions regarding the *usefulness* **of the system over the four-week study period.**
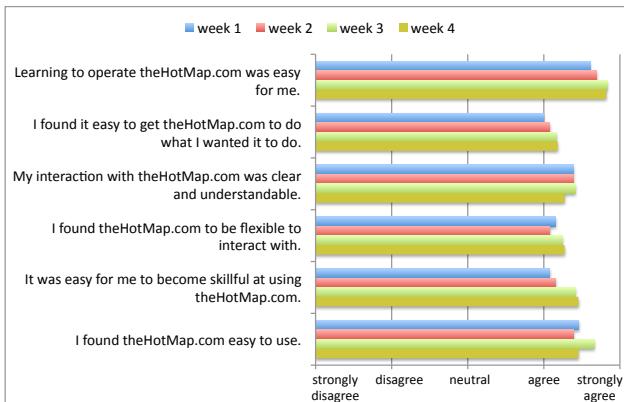

Figure 3. Average responses to the questions regarding the *ease of use* **of the system over the four-week study period.**

Figure 4. Participants with statistically significant changes in their responses to the *usefulness* **measures.**
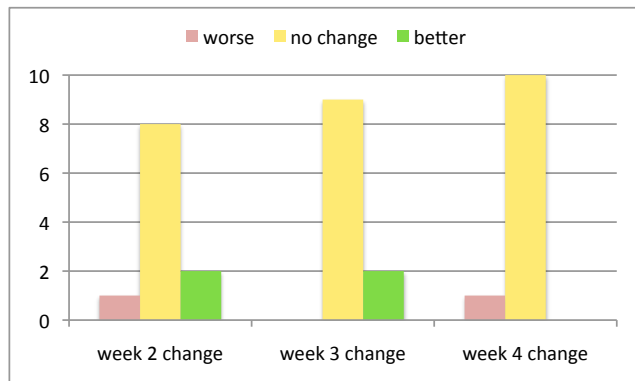

Figure 5. Participants with statistically significant changes in their responses to the *ease of use* **measures.**

5.3.3 Open-Ended Feedback

Participants were encouraged to provide general comments about the system after completing each of the online questionnaires as well as the post-study questionnaire. This open-ended feedback was analyzed using thematic content analysis. Although there were a number of positive comments and recommendations for new features, we focus here on the problems and difficulties participants reported.

The most frequent negative comment was with respect to the underlying search engine. The system used search results provided by Yahoo. However, all the participants reported using Google as their primary search engine. This lead to impressions of poor performance with respect to obtaining a good set of search results. In future studies, we must ensure that the underlying search engine matches the user's preference in order to avoid this negative impression.

A small portion of the participants reported slow load

responses of the 11 participants who completed the study. For all the scores that experienced a statistically significant change from one week to the next ($\chi^2 < 0.95$) we determined whether the change was positive or negative. The results of this analysis are provided in Figures 4 and 5.

For both the usefulness and ease of use measures, as time progresses, the responses stabilized resulting in fewer changes in impressions. It should be noted that for the usefulness measure, there were two participants who varied their responses between questionnaire submissions from being less useful, to more useful, to less useful. One of these participants also varied their responses to the ease of use questions in a similar manner.

times with the system. Certainly, there is some overhead in calculating and rendering the features of the system, when compared to the simple representations used by the top search engines. During the development process, we found the performance of the system to be acceptable using modern computer systems and browsers. It is possible that some participants were using older equipment and slower Internet connections. As network speeds increase and computers become faster, such problems will be reduced.

There were also a number of minor issues that were identified by the participants. These included problems with the removal of stop-words from the *query term headers*, the perception of clutter in the display, and the inability of the system to recommend better query terms for the user. The next version of $theHotMap.com$ will address these issues.

5.3.4 Continuation of Use

Part of the post-study questionnaire included a question regarding the use of $theHotMap.com$ after the conclusion of the required use period. Of the participants who saw the study through to its completion, 73% indicated that they continued to use the system. However, none of them indicated using it as their primary search engine due to the fact that the underlying search engine was different from their preferred search engine. This willingness to continue to use the system after the completion of the study is a strong indication of the benefits it provided to the participants.

5.4 Discussions

This real-world user evaluation provides a much richer view of how $theHotMap.com$ is being used than the previous laboratory studies conducted on the individual components. In this study, participants were able to conduct searches on whatever topic they wished, and used the system as part of their regular work, school, and home activities for an extended period of time. Unfortunately, it is difficult to untangle the benefits of the visual representations of the data from the interactions with those elements. Therefore, we base our conclusions on how the features of the system were used, under the assumption that the utility of a feature is an indication of the effectiveness of its visual representation.

While the re-sorting features of the system were not used heavily, this may not be as negative of a result as it appears. These features were designed to support exploratory search activities. It may be that many of the searches conducted by the participants were targeted searches, for which the underlying search engine performed well. In these cases, there would have been little need to explore the search results. Although the proportion of queries that were explored in an interactive manner appears to be low (7%), it is inline with results that have been reported on Web search user behaviour [20].

In the cases where the search results were re-sorted, this feature proved to be quite successful. Of the documents that were deemed to be relevant (i.e., clicked by the user), most were promoted to a more prominent location in the search results list than they would have occupied given the original order provided by the underlying search engine. As such, we conclude that the re-sorting features of $theHotMap.com$ are valuable and useful.

With respect to the query refinement features, a large portion of the queries were manually refined, with only a small number taking advantage of the interactive query refinement features. We conclude that the *term histogram* was valuable for guiding the user in refining their query. However, the visual indicator that supported the one-click addition or removal of query terms may have been too subtle.

Similarly, the ability to navigate within the *abstract overview* of the search results set may not have had sufficient visual cues to indicate its function. However, some participants did note that they found this overview useful for quickly determining the makeup of the search results set.

Over the course of the four-week study, there was some variance in the responses to the questions regarding the utility and ease of use of the system. As time went on, this variance was reduced as the participants learned how to use the system and increase the effectiveness of their searches. Most of the responses regarding usefulness were in the *neutral* to *agree range*; whereas most of the responses regarding the ease of use were in the *agree* to *strongly agree* range. As such, we conclude that the system can be useful (e.g., for searches that are exploratory), and that the system is easy to use.

6 Conclusion

In this paper, we outline and report on a real-world user evaluation of $theHotMap.com$. We found that while the search results exploration features were not used heavily, they were effective when the searchers chose to use them. The query refinement features did support the construction of new queries, although the automatic methods for doing so were underutilized. In the end, most participants agreed that the system was useful and easy to use.

Future work includes replicating these studies using search results provided by Google (since all the participants in this study used Google as their primary search engine), making refinements to the system based on the outcomes of this study, and exploring other visual representations that may support the user-centric activities of interactive search results exploration and interactive query refinement. We also wish to explore other search domains that may benefit from such visual and interactive search interfaces, such

as library systems and corporate intranets.

Acknowledgements

References

[1] Ask. Ask.com search engine. `http://www.ask.com/`, 2008.

[2] E. Berenci, C. Carpineto, V. Giannini, and S. Mizzaro. Effectiveness of keyword-based display and selection of retrieval results for interactive searches. *International Journal on Digital Libraries*, 3(3):249–260, 2000.

[3] P. Borlund. The IIR evaluation model: A framework for evaluation of interactive information retrieval systems. *Information Research*, 8(3), 2003.

[4] Clusty. Clusty: the clustering search engine. `http://www.clusty.com/`, 2008.

[5] F. D. Davis. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *Management Information Systems Quarterly*, 13(3):319–340, 1989.

[6] M. Dörk, S. Carpendale, C. Collins, and C. Williamson. VisGets: Coordinated visualizations of Web-based information exploration and discovery. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1205–1212, 2008.

[7] Grokker. Grokker - enterprise search management and content integration. `http://www.grokker.com/`, 2008.

[8] M. Hearst. Tilebars: Visualization of term distribution information in full text information access. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pages 59–66, 1995.

[9] T. Heimonen and N. Jhaveri. Visualizing query occurrence in search result lists. In *Proceedings of the International Conference on Information Visualization*, pages 877–882, 2005.

[10] O. Hoeber. User evaluation methods for visual Web search interfaces. In *Proceedings of the International Conference on Information Visualization*, 2009.

[11] O. Hoeber, M. Brooks, D. Schroeder, and X. D. Yang. TheHotMap.com: Enabling flexible interaction in next-generation Web search interfaces. In *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence*, pages 730–734, 2008.

[12] O. Hoeber and X. D. Yang. Interactive Web information retrieval using WordBars. In *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence*, pages 875–882, 2006.

[13] O. Hoeber and X. D. Yang. The visual exploration of Web search results using HotMap. In *Proceedings of the International Conference on Information Visualization*, pages 157–165, 2006.

[14] O. Hoeber and X. D. Yang. Evaluating WordBars in exploratory Web search scenarios. *Information Processing and Management*, 44(2):485–510, 2008.

[15] O. Hoeber and X. D. Yang. HotMap: Supporting visual explorations of Web search results. *Journal of the American Society for Information Science and Technology*, 60(1):90–110, 2009.

[16] B. J. Jansen, A. Spink, and T. Saracevic. Real life, real users, and real needs: A study and analysis of user queries on the web. *Information Processing and Management*, 36(2):207–227, 2000.

[17] H. Joho, C. Coverson, M. Sanderson, and M. Beaulieu. Hierarchical presentation of expansion terms. In *Proceedings of the ACM Symposium on Applied Computing*, pages 645 – 649, 2002.

[18] B. Kules, J. Kustanowitz, and B. Shneiderman. Categorizing Web search results into meaningful and stable categories using fast-feature techniques. In *Proceedings of the ACM/IEEE-CS Joint Conference on Digital Libraries*, 2006.

[19] Mozilla. Firefox Web browser. `http://www.mozilla.com/en-US/firefox/`, 2008.

[20] R. W. White and S. M. Drucker. Investigating behavioral variability in web search. In *Proceedings of the International World Wide Web Conference*, pages 21–30, 2007.

[21] Yahoo. Yahoo search Web searvices. `http://developer.yahoo.com/search`, 2008.

[22] Yahoo. Yahoo Web search. `http://www.yahoo.com/`, 2008.