A UNIFIED INTERFACE FOR VISUAL AND INTERACTIVE WEB SEARCH

Orland Hoeber and Xue Dong Yang
Department of Computer Science
University of Regina
Regina, Saskatchewan, Canada
email: {hoeber, yang}@cs.uregina.ca

ABSTRACT
The interfaces used by the top Web search engines have changed very little since the early days of Web search. These interfaces follow the traditional model of information retrieval in which users first formulate their queries and then evaluate the search results. Little support is provided for the users' tasks of crafting and refining their queries, and subsequently exploring the search results. In this paper, we describe HotMap+WordBars, a next-generation interactive Web search interface that allows users to take an active role in their Web search tasks through visual exploration and manipulation features.

KEY WORDS
Web search, user interface, interactive query refinement, interactive search results exploration.

1 Introduction

In recent years, significant research has gone into the improvement of Web search engines, both in the sizes of their indexes, as well as in the algorithms used to match users' queries to these indexes. A prime example of these advances are those made by Google [2, 3]. In the course of these improvements, Web search engines have become the primary tool people use to find information on the Web.

Web search engines generally perform very well when users are able to craft good queries that accurately capture their information needs. In these cases, there are often many relevant documents within the first few pages of the search results. Presented in a list-based format, the users can easily view the top search results one-by-one and decide the relevance of the documents with respect to the information they are seeking.

However, when users are unable to effectively craft a query, or when their information needs are inherently ambiguous or poorly defined, Web search engines provide little support to assist the users in their primary Web search tasks. In these cases, the results of an initial search will often contain many non-relevant documents. The options available to the user are to either spend time looking through many pages of search results, or try to construct a better query.

Evidence of the inability of users to craft effective queries has been highlighted in studies on user behaviour with respect to Web search engines [9, 14, 15]. A recurring theme in these studies is that search queries commonly consist of only one to three terms, and that users seldom make subsequent modifications to their queries, even when the search results are poor. Part of the difficulties that users experience can be attributed to the lack of support provided by the Web search engines for the task of crafting a query. Often, queries must be modified manually, based on information the users may have learned by viewing the results of the current or previous searches.

A related problem identified in these studies is that users seldom venture past the third page of search results [14, 15]. Even when users are able to craft a high quality query that effectively captures their information needs, the search results may be a mixture of relevant and non-relevant documents. The static list-based representation that is common among Web search engines requires the users to evaluate the search results one-by-one, and to some degree, in the order provided.

Given the shortcomings of the interfaces of the top Web search engines, an opportunity exists to use interactivity and visualization as a means for supporting the users in their primary Web search tasks. Our approach provides visual representations of features of the initial set of search results, supporting the users as they interactively refine their queries and interactively explore the search results sets. This represents an example of what we believe will be the next generation of Web search interfaces: systems which allow the users to become actively involved in their Web search tasks through the visual exploration and manipulation of Web search results.

2 HotMap+WordBars

Our prior research in the area of interactive Web search interfaces has resulted in a number of systems to support users in performing their Web search tasks. Two such systems are HotMap [8] and WordBars [7]. User evaluations of HotMap have shown that the visual and interactive features it provides can effectively support users in finding relevant documents [6]. A user evaluation with WordBars was also very positive (the results are currently under review for publication).

Two observations about how some people search the Web have guided this research. The first observation is

that users sometimes scan the search results looking for documents which make frequent use of their query terms. This observation led to the exploration of methods for visually displaying query term frequencies from the top search results (HotMap), and subsequently, the investigation of methods for visually displaying the frequencies of other terms that also appear in the top search results (WordBars). The second observation was that there are commonly many relevant documents buried deep in the search results list; accessing these documents often requires the users to consider many non-relevant documents in the process. From this observation, techniques were explored for re-sorting the search results to bring these relevant documents to the top of the list, a feature that is present in both HotMap and WordBars.

There are benefits to combining the positive features of these two systems into a single system that supports both interactive query refinement and interactive search results exploration in a single interface. HotMap is an effective tool for the visual exploration and interactive manipulation of the search results set when a high-quality query is used. However, it provides no support for crafting or refining a query. The WordBars features address this shortcoming, and provides additional methods for exploring the search results. As a result, the users of the *HotMap+WordBars* system can take advantage of the features of both systems in a seamless unified interface that supports both fundamental Web search tasks.

In this paper, we describe the features drawn from HotMap and WordBars, along with the benefits provided by combining these features into a single Web search interface. A usage scenario is provided to illustrate these benefits. The subjective reaction results from user studies highlight the ease of use and the degree to which users are willing to accept the visual and interactive features of HotMap+WordBars.

3 HotMap

The primary goal in the development of HotMap was to support the users in exploring the search results and ultimately finding relevant documents. After a user submits a query, a set of the top search results are obtained from the underlying search engine (the Google API [4]), and the frequencies of the query terms are counted within the title and snippet of these search results. The search results are presented in a scrollable document list. Along with each document in this list, a group of colour-coded indicators are provided (one for each query term) that visually represent the frequency of the query terms.

This supplemental information about the query term frequencies is present both in the *document list*, which shows the details of a subset of the search results (see Figure 1a), and in the *overview map*, which shows a compact and abstract representation of all the documents in the search results set (see Figure 1b). Coordinated scrolling between these two views of the search results allows the users to easily remain aware of the location of documents they are currently viewing with respect to the entire set of documents retrieved.

In the original implementation of HotMap, the search results were presented in the *document list* in a compact format, where 20 to 30 documents could be viewed at one time. This compact format only showed the title of the document in a persistent manner; the snippet and URL of the document were available as a tool tip. Although many participants in our user studies indicated that they preferred a compact representation which promotes scanning within the titles, others indicated that they did not like the compact representation.

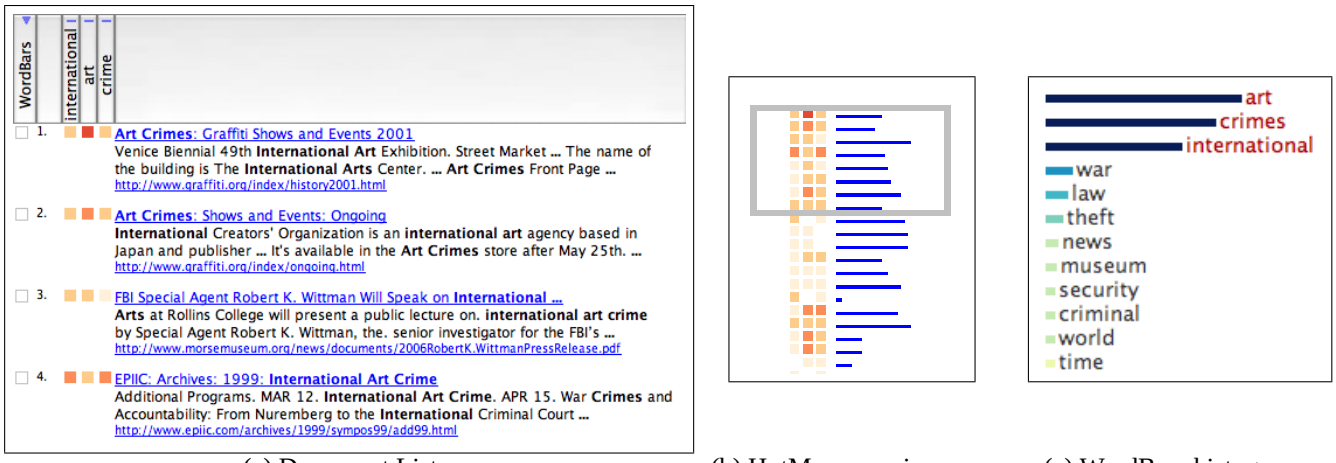
In the development of WordBars, we departed from this compact representation, and provided the title, snippet, and URL in a format similar to that used by the major Web search engines. In the combined HotMap+WordBars system, the user has a choice between the expanded and compact format, and can easily switch between the two views as as they perform their search tasks.

In the *overview map*, a compact visual representation of the set of retrieved search results is displayed. This includes colour-coded boxes to represent the query term frequencies in the search results, as well as lines whose lengths are proportional to the lengths of the titles of the documents. A grey box indicates which documents are currently being displayed in the document list. These features provide perceptual cues that make the relationship between the overview map and the document list apparent to the user [1].

Although the overview map can easily represent several hundreds of documents at a time, the current implementation limits the search results set to 100 documents. This limit represents 10 pages worth of search results, and was chosen as a balance between two constraints: collect more documents than users would normally have the patience to view; and return the search results set to the users in near real-time. Collecting fewer documents may result in missing some highly relevant documents buried deeper in the search results; collecting more documents may result in the user waiting longer for the search results to finish loading.

The features of HotMap provide two interaction methods to support users as they explore the search results. First, users can visually inspect the overview map to identify documents which make frequent use of the query terms. A simple glance at the overview map provides users with an immediate impression of the use of their query terms within the search results set. The users can jump to the location of documents of interest by clicking on the overview map, which scrolls the document list to the corresponding location in the search results. The documents can then be investigated further in the document list and considered for relevance.

A second method for interacting with the search results is through a nested sorting feature in the document list. Each query term is provided at the top of a column in



(a) Document List (b) HotMap overview map (c) WordBars histogram

Figure 1. Zoomed-in views of the three main features of HotMap+WordBars.

the document list, below which are the colour-coded query term frequency indicators. Clicking on the query term column header sorts the search results based on the frequency of this term. Control-clicking can add additional query terms in a nested sorting fashion. This allows the users to easily specify their order of preference for their query terms, resulting in a re-sorting of the top search results.

Since the re-sorting of the search results is an important feature in this system, a method for keeping track of which documents have been viewed is necessary. The de facto standard for Web link colours is employed, with blue being used for the links of documents that have not yet been viewed, and purple being used for the links of viewed documents. Accessing a document simply requires a click of the title link of the document, resulting in the document being opened in a new browser window.

Through these two interactions methods, the users can interactively and visually explore the set of search results returned from the query. Doing so can lead the user to discover relevant documents buried deep in the search results that would have been difficult to view using the static list-based representation that is commonly employed by the top search engines. User evaluations of HotMap indicate that the visual inspection of the overview map and the nested sorting in the document list window based on query term frequencies can allow the users to become more efficient and effective in finding relevant documents within the search results [6].

4 WordBars

The goal in the development of the WordBars features was to provide support to users both in their tasks of crafting better queries, and their tasks of evaluating and exploring the Web search results. The technique used is to count the frequencies of all the terms that appear in the title and snippet of the top search results, and presents the top 20 of these terms in a vertically-oriented colour-coded histogram (see Figure 1c).

Rather than using exact term matches, Porter's stemming algorithm [12] is used in the counting process so that words with the same root are matched as the same term. The terms displayed in the histogram use the format of the first occurrence of the term from the search results. Common terms, and those that are less than three characters long are ignored, since they are often words with little descriptive value for the Web search process.

By visually inspecting this histogram, users are able to determine the general makeup of the top search results. The commonly used terms are clearly identifiable, as are the relative differences in frequency of use. The terms that appear in this histogram can assist the users in determining whether their query is very specific (with many of the terms being relevant to the information need) or very vague (with a wide range of terms on a number of different topics).

The user interaction within the histogram was designed to be very easy to use. Clicking on a term in the histogram selects it, and causes the search results to be re-sorted based on the frequency of this term. Clicking multiple terms in this manner causes the search results to be re-sorted based on their collective use of the selected terms. Double-clicking adds or removes the target term from the query, allowing the user to submit a newly refined query to the underlying search engine.

Providing a list from which the users can choose additional terms to add to their query is not a new technique. A commonly cited method [5] provided three simple lists of terms from which the users could choose. However, it has been suggested that users may have difficulties making effective selections from these lists when given the task of expanding their queries [13, 10]. In WordBars, the histogram representation allows the users to easily see the relative frequencies of the terms in the top search results. In addition, the users can experiment with re-sorting the search results based on some of these terms before committing to adding

them to the query. These additional features make the utility of the interactive WordBars histogram superior to a simple list of terms.

The histogram containing potential terms to add or remove from the query allows the users to recognize terms that are relevant to their information need, rather than requiring them to recall all the relevant terms. *Recognition rather than recall* is a primary usability principle, suggesting that users should be able to see the information they need as they conduct their tasks, rather than having to remember the information [11]. As a result, users can start with a query consisting of a few relevant terms, and then easily identify additional terms to add or remove from the query.

A new feature in HotMap+WordBars is a colour-coded indicator in the document list which visually represents the frequencies of the set of selected terms in WordBars for each of the documents in the search results set. These colour codes use a yellow-green-blue colour scale so as not to interfere with the HotMap colour scale. They communicate to the users the relative differences in the frequencies of the terms selected from the WordBars histogram within the search results set.

It is interesting to note that simply selecting the most frequently used terms to add to the query may not produce better search results. Often, the significant words are those that appear in the middle frequencies [16]. WordBars assists the users in recognizing terms that may be good descriptors of their information needs, and allows the users to easily see how these terms are being used in the search results before choosing to add them to their query.

5 Combined Benefits

Although both the HotMap and WordBars features provide visual representations of the properties of the Web search, they do so in support of two different user tasks. The primary benefit of the WordBars features is to support the users in interactive query refinement processes. By re-sorting the search results, the users can consider the potential relevance of candidate terms, and easily add or remove these from the query. By contrast, the primary benefits of the HotMap features are to explore the search results (both visually and through interactive nested re-sorting) once a high-quality query has been provided. Together, these features support the users in their Web search task within a unified user interface.

The single interface is an important aspect of HotMap+WordBars. Within the Web search process, once the user is satisfied that they have constructed a quality query for their information need, they can readily turn their attention to the HotMap features as they explore the search results. Similarly, if a user decides that they would like to return to the query refinement stage, they can do so without the cognitive overhead of initiating a change of state in the interface. All they need to do is direct their attention back to the WordBars histogram.

6 Usage Scenario

The HotMap+WordBars system supports the users in their Web search tasks through visual representations of features of the current search, along with methods for interactively manipulating both the query and the search results set. To illustrate how the this system can support the users as they conduct their Web search tasks, a typical usage scenario is provided. Suppose a user wishes to search for information on the same topic as the previous example: “international art crime”. The results of the first stage of this search are depicted in Figure 2a. By inspecting the top few search results, the user can see that there are a lot of non-relevant documents mixed in with the relevant documents. As such, the user may wish to refine their query further.

Inspecting the WordBars histogram, the user can see that there are a number of terms that appear in the search results that are relevant to their information need. By selecting a few of these, those documents that make use of these terms the most are moved to the top of the search results, as shown in Figure 2b. These documents can readily be considered for relevance.

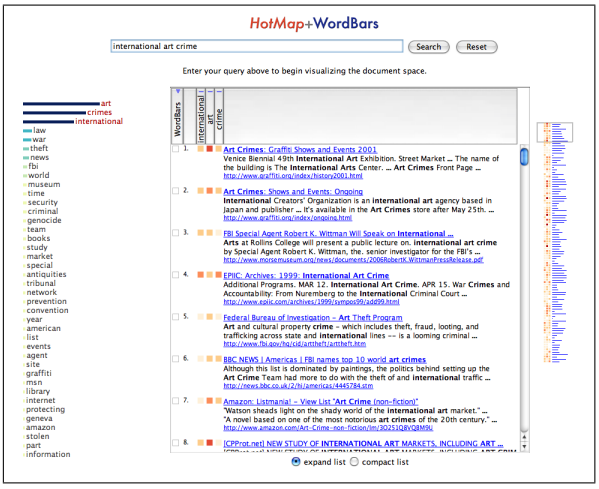
With the query refined based on terms recognized as relevant in the WordBars histogram, the user can retrieve a new set of search results by clicking the search button. The system presents a new set of search results, along with a new WordBars histogram and a new HotMap representation, both based on the new search results provided by the Google API. The results of this refined search are displayed in Figure 2c.

At this point, the user may wish to explore the top 100 search results by visually inspecting the HotMap. Within this compact and abstract representation, the user can easily identify documents that make frequent use of the query terms. Clicking or scrolling to the vicinity of one such document allows the user to view the corresponding document in the document list and determine its relevance. For example, consider document 52 in Figure 2d.

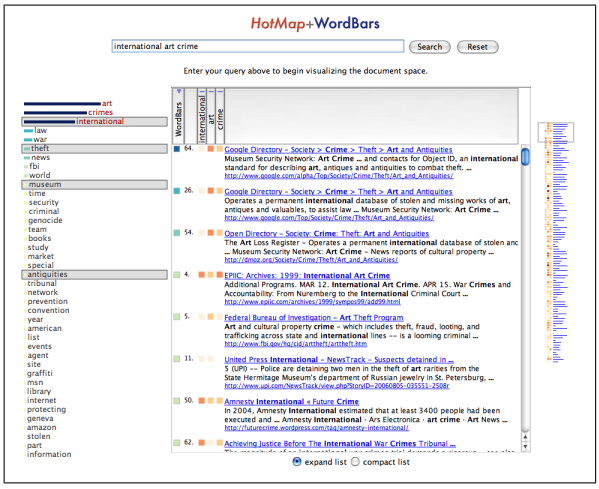
Alternately, the user may wish to browse the search results in the document list in a more compact manner. Selecting the “compact list” option, the search results are provided with one document title per row. Tool tips are used to access the snippet and URL of the document, as illustrated in Figure 2e.

In addition to the visual inspection, the user can also re-sort the search results based on the importance they place on the terms in our query. For example, the user may wish to explore the documents that make frequent use of the terms “international” and “theft”. A nested sort is specified by control-clicking on these terms in the column header of the document list, the results of which are illustrated in Figure 2f.

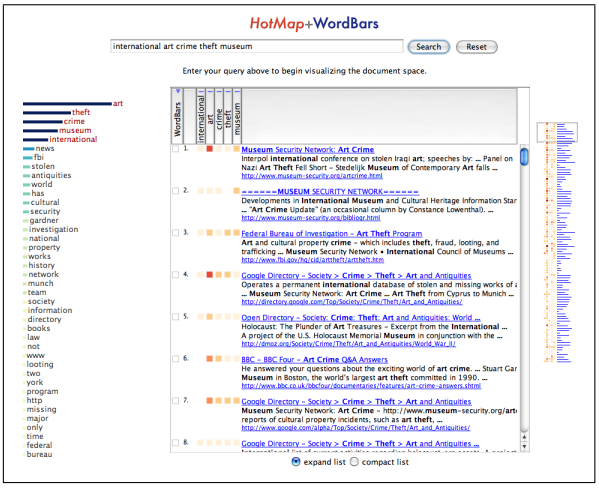
From this scenario, the support this system provides to the user in interactively refining their query, as well as interactively exploring the search results is clear. In the query refinement process, the user was able to recognize additional terms that might be relevant to their informa-



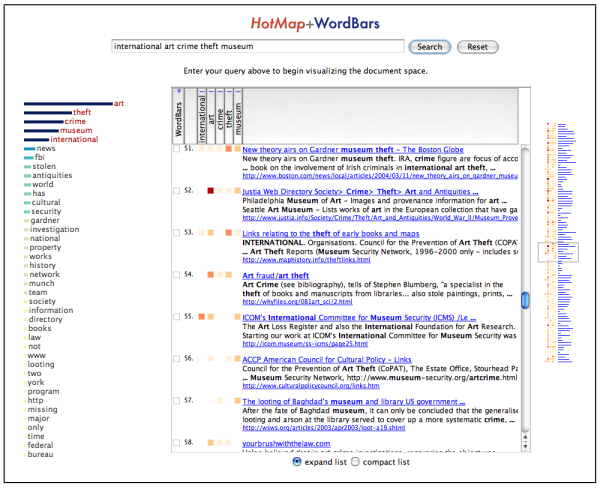
(a) Initial search for *international art crime*.



(b) Selecting relevant terms in the WordBars histogram.

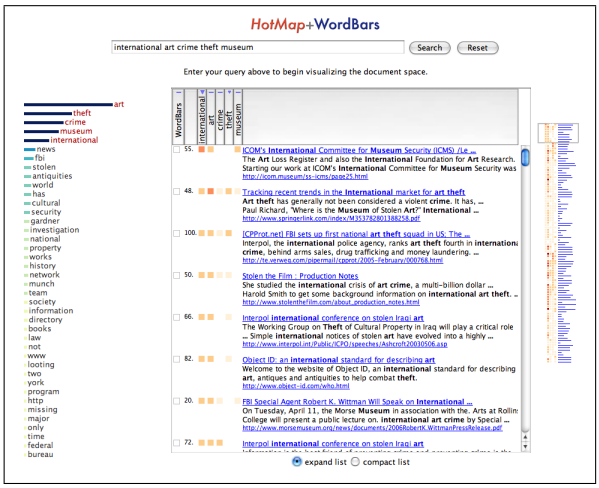


(c) Refining the query based on the information in the WordBars histogram.



(d) Visual inspection of the HotMap

(e) Compact representation of search results list.



(f) Re-sorting the search results based on query terms.

Figure 2. Screenshots from HotMap+WordBars illustrating a usage scenario.

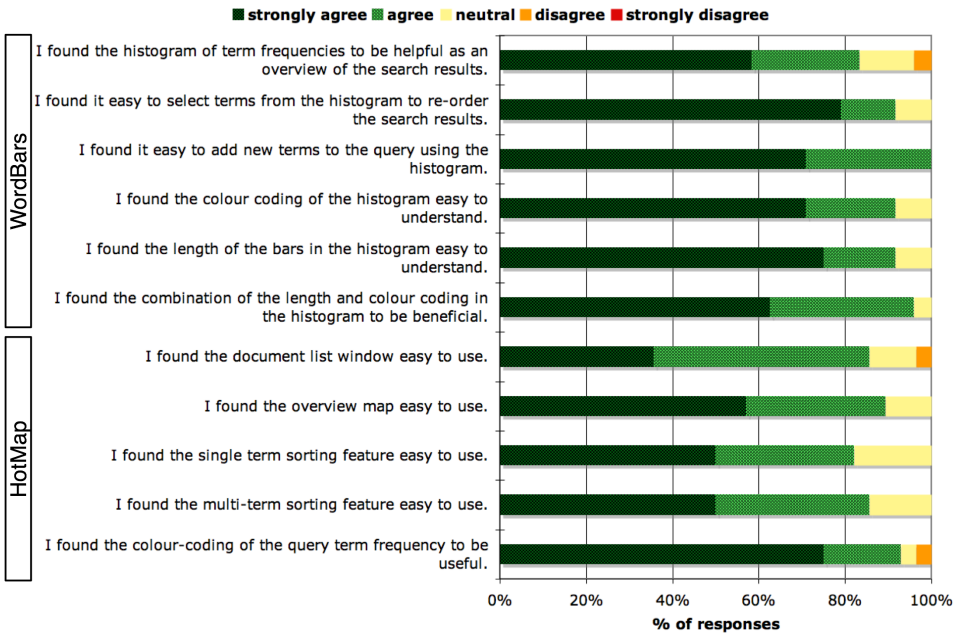


Figure 3. Responses to subjective reaction questions for HotMap and WordBars features.

tion needs, and experiment with bringing the documents that make use of those terms to the top of the search results list. Adding new terms to the query simply required a double-click. Once a good query was developed, the HotMap representation supported the user in visually exploring the search results, as well as re-sorting the search results based on the importance placed on the query terms by the user. Because each of the operations that re-sorts the search results in the document list does so instantly, the user was able to interactively explore the search results. A video depicting a similar scenario is available on the author's Web site [1].

7 Subjective Reactions

User studies were conducted to evaluate the effectiveness of the HotMap and WordBars features separately. Preliminary results from HotMap were reported in [6]; here we report the subjective reactions with respect to the various interface features from the extended studies. There were 28 participants in the HotMap evaluation, and 24 participants in the WordBars evaluation.

Figure 3 illustrates the subjective reactions of the participants to the various features of both HotMap and WordBars after having completed a number of assigned tasks using the interfaces. Clearly, many of the participants found the features easy to use and useful in their Web search tasks.

When asked to rank their preference for a Web search interface, 68% of the participants ranked HotMap higher than Google. In the WordBars study, 96% of the participants ranked the search results generated using the features of WordBars as preferable to the original set of search results provided by the Google API. These subjective reactions to the visual and interactive features present in HotMap+WordBars indicate that users are willing to accept and use new tools to support their Web search tasks.

8 Discussion

Wise et al. noted that "the need to read and assess large amounts of text that is retrieved through even the most efficient means puts a severe upper limit on the amount of text information that can be processed by any analyst for any purpose" [17]. This statement provides a clear motivation for the use of visualization methods to support Web search tasks. Allowing the searchers to *see* the information without having to *read* the information may allow this upper limit to be exceeded.

Visual representations of the features of a Web search can be further exploited through interactive methods. While a static visual representation may be of some value in evaluating the search results, an interactive visual representation allows the users to take an active role in the information retrieval process. Systems such as HotMap+WordBars can provide support for users in their tasks of interactive query refinement and interactive search results exploration.

The primary drawback in this work is that there is little ability to support the users in their Web search tasks when a very poor initial query is provided. If no relevant documents are present within the top search results, the ability to explore the search results or refine the query based on frequently used terms is of little value

[1] `http://www.cs.uregina.ca/~hoeber/HMWB/`

to the user. However, if at least some of the documents that are returned from the initial search are relevant, then HotMap+WordBars can be very beneficial in supporting the users as they conduct their Web searches, as illustrated by the usage scenario in this paper.

9 Conclusion

In this paper, the methods by which HotMap+WordBars makes use of interaction and visualization features to support Web searchers in their tasks of interactive query refinement and interactive search results exploration are highlighted. The system allows the users to take an active role in their information seeking tasks, and interact with the information present in the search results both to build better queries, and to find relevant documents that are buried deep in the search results. All of the interaction occurs within a single unified interface, allowing the users to easily switch between their two primary tasks of interactive query refinement and interactive search results exploration.

HotMap+WordBars represents an example of what we believe will be the next generation of Web search interfaces: tools that provide visual representations of features of the search, and allow the users to interact within the search process in order to improve their abilities and performance. Our future plans include further evaluations of this work in real-world Web search settings, exploring other metrics which may assist the users in selecting relevant terms which are of value for query refinement and search results exploration, as well as investigating the utility of personalization features to support the users as they conduct their Web search tasks.

References

[1] Michelle Q. Wang Baldonado, Alliston Woodruff, and Allan Kuchinsky. Guidelines fo using multiple views in information visualization. In *Proceedings of the ACM Advanced Visual Interfaces*, 2000.

[2] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. In *Proceedings of the Seventh International World Wide Web Conference*, pages 107–117, 1998.

[3] Sanjay Ghemawat, Howard Gobioff, and Shun-Tak Leung. The Google file system. In *Proceedings of the Nineteenth ACM Symposium on Operating System Principles*, pages 29–43, 2003.

[4] Google. Google web API. www.google.com/apis/, 2005.

[5] Donna Harman. Towards interactive query expansion. In *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 321–331, 1988.

[6] Orland Hoeber and Xue Dong Yang. A comparative user study of web search interfaces: HotMap, Concept Highlighter, and Google. In *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence*, 2006.

[7] Orland Hoeber and Xue Dong Yang. Interactive web information retrieval using WordBars. In *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence*, 2006.

[8] Orland Hoeber and Xue Dong Yang. The visual exploration of web search results using HotMap. In *Proceedings of the International Conference on Information Visualization*, pages 157–165, 2006.

[9] Bernard J. Jansen and Udo Pooch. A review of web searching studies and a framework for future research. *Journal of the American Society for Information Science and Technology*, 52(3):235–246, 2001.

[10] Mark Magennis and Cornelis J. van Rijsbergen. The potential and actual effectiveness of interactive query expansion. In *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval*, 1997.

[11] Jakob Nielsen. Enhancing the explanatory power of usability heuristics. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 1994.

[12] Martin Porter. An algorithm for suffix stripping. *Program*, 14(3), 1980.

[13] Ian Ruthven. Re-examining the potential effectiveness of interactive query expansion. In *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval*, 2003.

[14] Craig Silverstein, Monika Henzinger, Hannes Marais, and Michael Moricz. Analysis of a very large web search engine query log. *SIGIR Forum*, 33(1):6–12, 1999.

[15] Amanda Spink, Dietmar Wolfram, B. J. Jansen, and Tefko Saracevic. Searching the web: The public and their queries. *Journal of the American Society for Information Science and Technology*, 52(3):226–234, 2001.

[16] C. J. van Rijsbergen. *Information Retrieval*. London: Butterworths, 1979.

[17] James A. Wise, James J. Thomas, Kelly Pennock, David Lantrip, Marc Pottier, Anne Schur, and Vern Crow. Visualizing the non-visual: Spatial analysis and interaction with information from text documents. In *Proceedings of IEEE Information Visualization*, 1995.