

The Visual Exploration of Web Search Results Using HotMap

Orland Hoerber and Xue Dong Yang
Department of Computer Science
University of Regina
Regina, Saskatchewan, Canada S4S 0A2
{hoerber, yang}@cs.uregina.ca

Abstract

While the information retrieval techniques used by web search engines have improved substantially over the years, the search results have continued to be represented in a simple list-based format. Although this list-based representation makes it easy to evaluate a single document, it does not support the users in the broader tasks of manipulating the search results, comparing documents, or finding a set of relevant documents. HotMap provides a compact visual representation of web search results at two levels of detail, and supports the interactive exploration of web search results. User studies have shown that HotMap can result in fewer low-relevance documents being considered, and generates a higher level of confidence, ease of use, and satisfaction than a Google-like interface.

1 Introduction

While it is clear that significant effort has gone into creating web search engines that can index billions of documents and return the search results in fractions of a second [4], the presentation of these search results has remained essentially unchanged since the early days of web search. The search results generated from a user's query consists of a collection of document surrogates, each of which contains summary information, attributes, and other meta-data about the matched documents. These document surrogates are often presented in a simple list-based format, displaying the title of the document, a snippet containing the query terms in context, and the URL.

Even though this simple list-based representation provides the search results in a clear and effective manner for determining the relevance of individual document surrogates, it requires that each document surrogate be evaluated in turn, and to some degree, in the order provided. Further, there is little support for determining the overall properties of the search results, nor for manipulating and exploring the

search results set.

Requiring users to evaluate each document surrogate individually, often with only ten documents per page, leads to a common user search trait of evaluating only one to three pages of search results before either re-formulating their query or giving up [18, 19]. While Spink et al. noted that "the public has a low tolerance of going in depth through what is retrieved" [19], we suggest that this is an indication of user frustration both with the process of evaluating the search results, as well as with the lack of relevant results in the first few pages. Often, re-formulating a poorly defined query is needed. However, in many cases there may be high quality relevant documents buried in the search results set that were missed because the users did not look at enough search result pages.

To address these shortcomings, we have developed a system for visualizing and interactively manipulating web search results, called *HotMap*. This meta-search system retrieves the top search results returned by the Google API [6] and presents these results in a compact visual manner that supports both visual information processing, and user-directed exploration. Although the current prototype uses the first 100 document surrogates from the search results, the visual representation is compact and flexible enough to support the display of thousands of document surrogates in a single display screen.

Our motivation for this work is based on a common method for evaluating the relevance of a document surrogate: identifying which of the query terms appear in the title and snippet. The list-based representation of the document surrogates often uses query term highlighting (i.e., bolding the query terms) which helps the users to identify their query terms in the title and snippet, but still requires the users to read or at least recognize the terms. *HotMap* represents the query term frequencies in the form of a colour on a heat scale. Multiple occurrences of a query term result in a dark red colour; fewer occurrences are represented by progressively lighter shades of red and orange. As shown in Figure 1, this colour coding allows the users to *see* the "hot"

documents easily (and provides the inspiration for naming our system *HotMap*).

In *HotMap*, the search results are presented in a grid-based layout at two levels of detail: an overview map provides a compact representation of the top search results, and a detail window provides a focused view of approximately 20 documents at a time. Coordinated scrolling between these two views allows the user to easily track the location of the detail window with respect to the overview map, as well as easily jump to a location of interest. Of the document surrogate information provided by the Google API, only the title of the document is displayed in the detail window; information hiding is used to hide and dynamically show the additional details available in the document surrogate, such as the snippet and the URL. To promote the manipulation and exploration of the search results, the users can initiate a nested sorting based on the query term frequencies, which automatically updates the order of the document surrogates in both the overview map and the detail window.

The remainder of this paper is organized as follows: An overview of text and document surrogate visualization is provided in Section 2. In Section 3, the details of the *HotMap* system are given. Section 4 describes the framework for the user evaluation, with the results presented in Section 5. Conclusions and future work are presented in Section 6.

2 Background

Many systems have been developed in recent years that apply information visualization techniques to information retrieval problems. While it is not feasible to provide a complete analysis of all these systems in this paper, we do provide a brief review of some relevant and interesting techniques. We categorize these according to whether they attempt to visualize the entire document or an abstraction of the document (i.e., the document surrogate).

2.1 Text Visualization

Text visualization can be defined as the process of converting textual information into graphical representations that can be processed visually rather than read. Since preattentive processing of certain types of graphical information is significantly faster than the non-preattentive processing required for reading [21], there is a great opportunity for taking advantage of the human visual processing capabilities when presenting textual information.

However, the representation of textual information in a visual manner is by no means a simple task. At the most fundamental level, we can think of a document as a collection of terms, represented by a high dimensional vector. Di-

mensional reduction techniques can be used to map a set of such document vectors into two or three dimensional space, resulting in each document being mapped to some point in space. The spatial proximity of two documents implies similarity, resulting in a visual clustering of documents.

These vector-based techniques have been used for the visualization of collections of documents in systems such as Galaxy of News [17], and ThemeScape [22]. However, since each document is a point in space, accessing and viewing the information on specific documents is not well supported. Hearst noted that “although intuitively appealing, graphical overviews of large document spaces have yet to be shown to be useful and understandable for users” [9].

Other techniques have retained the linear structure of the documents, and provide abstract representations of their contents. For example, in SeeSoft [5], each line of text is abstracted to a single horizontal line in the visual representation, retaining the general layout of the document. Colour is used to highlight the lines containing specific terms within the document. While useful in some situations, the resulting visual representation does not make efficient use of display space.

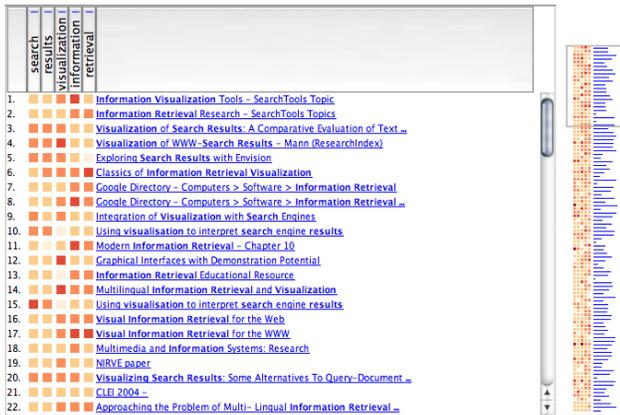
Rather than retaining the layout of the document, the contents can be divided into fixed blocks, and the frequency of the query terms can be represented by colour coding in each block, as in TileBars [8]. The result is a set of bars (one for each document) whose widths are relative to the length of the documents, and whose heights are relative to the number of query terms (or sets of query terms). This results in a more compact representation than the previous example.

In the work by Heimonen and Jhaveri [10], each document is divided into four equal sized blocks. The occurrences of all of the query terms within a 20-word window for each block are counted and depicted in a visual indicator. This indicator is displayed beside each document surrogate in the list-based representation of the search results.

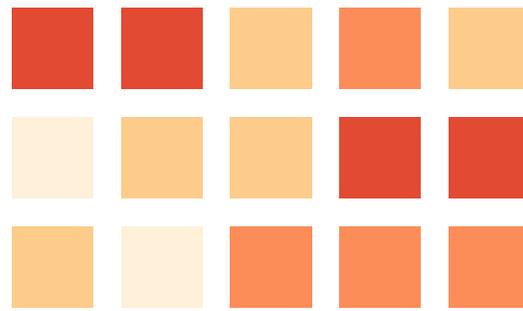
One common theme among these systems is that they all require access to the textual contents of all the documents to be displayed in order to generate the visual representation. Since meta-search systems are only provided with document surrogates from the underlying web search engines, to apply these techniques to web search would require retrieving each document individually. The additional time required to do this supplemental document retrieval would result in a meta-search system that is unable to display the search results in real-time.

2.2 Document Surrogate Visualization

Since access to the textual contents of each document is not feasible for meta-search visualization, the visual representation of document surrogates is a viable alternative. A



(a) HotMap screen shot



(b) magnification of overview map

Figure 1. The visual representation of the web search results consists of a detail window and an overview map (a). The magnified view of the overview map shows that the first document surrogate is “hot” with respect to the first two query terms; the second document surrogate is “hot” with respect to the last two query terms; and the third document surrogate is “warm” with respect to the last three query terms (b). These search results were returned from the query “search results visualization information retrieval”.

document surrogate consists of summary information, attributes, and other meta-data that represent the document in the search results. Document surrogates are the primary data objects in the list-based representation used by search engines, where they commonly consist of the title of the document, the URL, a snippet showing the query terms in context, as well as other information.

Envision [13] uses a highly customizable scatterplot and iconic visualization to represent the many different attributes available as part of their information retrieval system. Although this visual representation was shown to be very powerful, it makes use of information that is not commonly available in the document surrogates returned by web search engines. Further, there is an added level of complexity in this highly customizable interface that may make it too difficult for the general public to use effectively.

In VIEWER [2], all possible combinations of the query terms are generated and searched for in the document surrogates returned by the AltaVista search engine. A histogram of these query term combinations is provided to the user, which can be used to select subsets of the search results for further investigation. Although this system provides valuable information to the user in terms of how the query terms are used in the search results set, the information is provided in the context of the query terms, with little additional information provided with respect to specific document surrogates. While this information is of value in narrowing the search results to smaller subsets, and perhaps reformulating queries, it is not much value when evaluating the relevance of specific documents.

xFind [1] provides three different interfaces to a custom web document indexing system: a simple list-based representation; a scatterplot representation similar to that in Envision[13]; and a vector-based spatial clustering representation similar to that in ThemeScape [22]. While these representations of the search results take advantage of the extra information that is available through their indexing system, this information is not available with other search engines. Further, the spatial layout of the two visual representations maps the document surrogates to points in the two-dimensional display, making it difficult to view the additional information present in the document surrogate, or to make comparisons between document surrogates.

WaveLens [14] provides a focus+context representation of the search results allowing the users to dynamically zoom into document surrogates of interest. The results are provided in the traditional list-based representation. However, as users move their mouse over a document surrogate, its font size is increased as the font size of the other document surrogates is decreased. This results in a fisheye lens effect. Additional text from the document is dynamically added or removed from a document surrogate by clicking the mouse. While this technique may make it easier for the users to read the contents of the list of search results, it continues to promote the sequential evaluation of the document surrogates and provides little support for manipulating and exploring the search results.

Many other visualization systems exist for representing the document surrogates returned by web search engines, including a number of publicly accessible meta-search en-

gines such as Kartoo [11], Mooter[12], and Grokker [7]. An evaluation of the merits and problems with these systems is beyond the scope of this paper, although one review indicated that some of these systems do not add any support for the users in assimilating or processing the information [16].

3 HotMap

HotMap is a meta-search system that retrieves the document surrogates for a given user query from the Google API, and presents these web search results using visual representations at two levels of detail. The interactive exploration of the web search results is supported both via the inspection of the visual representations, and via the nested sorting features provided by *HotMap*.

Although the Google API can support advanced search features, Spink et al. noted that only a small portion of web searches make use of these advanced features [19]. In order to simplify the interaction with *HotMap*, only queries consisting of a list of terms are supported.

3.1 Document Surrogate Attributes

HotMap augments the document surrogate information provided by the Google API by calculating an additional set of attributes for each document surrogate representing the frequency of each of the query terms within the document title and snippet. These attributes corresponds to the users' question of "how often do my query terms appear in these documents?".

The query term frequency attributes are calculated by counting the occurrences of each of the query terms in the title and snippet for each of the document surrogates. Porter's stemming algorithm [15] is used to calculate the stems or roots of each of the query terms, as well as each of the words in the title and snippet. Matching based on these stems is more effective than exact word matches, since it takes into account different variations of the same root word. Therefore, given a query consisting of n terms, query term frequency attributes $\{q_1, q_2, \dots, q_n\}$ are added to each document surrogate.

3.2 Visualization of Search Results: HotMap

Information visualization takes advantage of the human visual information processing systems by generating graphical representations of data or concepts [21]. The cognitive activity involved in viewing and processing a visual representation allows the users to gain understanding or insight into the underlying data. With respect to the visualization of search results, the ultimate goal is to allow users to *see* the information without having to *read* the information. While

this is a difficult goal to meet using only the textual information present in search result document surrogates, the query term frequency attributes calculated by *HotMap* can be easily represented in a visual manner.

While some argue in favour of three-dimensional layouts, not only are there problems with occlusion, but judging the relative positions of objects can be difficult [21]. Instead, *HotMap* uses a two-dimensional grid layout, where each row represents a document surrogate in the search results, and each column represents one or more attributes or elements in the document surrogate.

The first column contains the document surrogate number, allowing the user to easily identify the degree of importance placed on each document surrogate by the underlying search engine algorithms. The next n columns hold the frequencies of the query terms, represented by a colour value. The final column contains the title of the document and tool tips to access the hidden information including the snippet and the URL.

Since the spatial position of an object and its colour can be perceptually separated, colour coding of the query term frequencies can be used without interfering with the spatial layout of the data [21]. Further, since colour is pre-attentively processed, this information is absorbed by the users far faster than if the users were required to read the numerical values [21].

The choice of a colour scale is not as simple as it might seem. Since we need to represent an ordered sequence of values, a colour sequence that varies monotonically on at least one colour channel is required [20, 21]. A set of perceptually distinct colours on the red end of a red-green colour scale were chosen to represent the term frequencies. This colour scale varies on both the luminance channel and the red-green colour channel. Visually, this colour scale appears to be a heat scale, resulting in high frequency terms appearing hot, and low frequency terms appearing neutral or warm. The colour scales used in *HotMap* were generated using the ColorBrewer application [3].

As illustrated in the screen-shot in Figure 1, two different levels of detail of the search results are simultaneously displayed. The overview map displays the search results in a compressed format by showing the query term frequencies and an abstract representation of the document surrogate title. The detail window shows a small fraction of the search results set at a time, and includes the labels for the query term columns, the title of the documents, and access to the hidden information. A scroll box in the overview map indicates the location of the document surrogates in the detail window. These coordinated views allow the user to both investigate the document surrogates in detail, as well as gain insight into the features of the entire set of search results displayed.

The use of query term frequencies and the visual rep-

resentation of this information is similar to TileBars [8]. Both *HotMap* and TileBars represent the frequency of query terms using a tile mosaic metaphor. While TileBars indicates the locations of the query terms within the entire documents, *HotMap* indicates the total frequency of the query terms for the document surrogates only, resulting in a more compact visual representation. Since *HotMap* provides a visual representation of the document surrogates, there is no need to retrieve the contents of each document (which is not a feasible approach for an interactive web search interface).

TileBars could be adapted to handle the document surrogates provided by a web search engine, resulting in visual representation similar to *HotMap*. However, since TileBars does not provide any support for interaction or exploration, the outcome would be a static representation of the web search results. Although the visual representation of the query terms would assist in the evaluation of individual document surrogates, the results would need to be considered in a sequential fashion. In addition to providing a visual representation of the web search results based on the document surrogates, *HotMap* extends this work by providing views of the search results at two levels of detail to support visual inspection, and tools for interactively re-sorting and exploring the web search results.

3.3 Manipulation and Exploration

Interaction is an important aspect of an effective information retrieval support system [23]. Allowing the users to interact with and manipulate the search results allows the users to take an active role in the information retrieval process, rather than the passive role that is common in traditional information retrieval systems.

Although the query term frequencies for each document surrogate are represented visually, the underlying information is numeric. The sorting of the search results based on this information is achieved by clicking on the column header corresponding to a query term. Sorting in decreasing and increasing order are supported, as is nested sorting (which is selected by holding down the control key while clicking the desired column headers). The use of this nested sorting feature allows the users to easily resort the query results based on the importance they give to their query terms. The results of this sorting are reflected instantly in both the overview map and the detail window.

Since two levels of detail of the search results are presented simultaneously (i.e., the overview map and detail window), coordination between these views is necessary. The user interaction features in these views is as one would expect: clicking or dragging the scroll box in the overview map will move the detail window to be centred on the selected location; scrolling in the detail window will move the scroll box in the overview map to the corresponding loca-

tion. These views of the search results allow the user to easily identify areas of interest in the overview map and jump to that location, as well as keep track of the location of the document surrogates being viewed in detail with respect to the rest of the search results.

In order to provide a compact representation of the search results, it is necessary to hide some of the information the users may find useful in determining document relevance. This hidden information includes the snippet and the URL of the document. While it is necessary to persistently display the titles in the detail window so the users can identify the document surrogates, other supplemental information can be hidden and displayed when needed. Tooltips are used to display this information when the user hovers their mouse cursor over a document surrogate.

Of course, the final information seeking task is to view the document corresponding to a document surrogate found to be relevant. This is achieved by clicking the title of the document in the detail window. The defacto standards for web link highlighting and underlining are used to indicate to the user that this is an available option. Clicking a document title will open that document in a new browser window, as well as indicate that the link has been followed by changing the link colour from blue to purple. A video showing the *HotMap* system in use is available on the author's web site¹.

4 Evaluation Framework

A user study was conducted to compare *HotMap* to the list-based representations used by many web search engines. Ten computer science graduate students were recruited, and asked to perform web searches over two task sessions using both *HotMap* and an interface designed to look identical to Google.

In the first task session, the participants were provided with queries and textual descriptions of their information need. Based on the feedback provided in this first session, refinements were made to the *HotMap* system, and the participants were invited back for a second task session. In the first session, the participants often had difficulty deciding the relevance of a document when they were not familiar with the assigned topic. Therefore, in the second session, the participants were asked to perform a web search on a topic of their choosing in which they are knowledgeable.

A questionnaire provided during the first session indicated that all the participants were expert computer users. 70% of the participants identified themselves as having a high level of experience with web searching; 30% identified themselves as having a moderate level of experience with web searching. All the participants indicated that they used Google as their primary web search engine.

¹<http://www.cs.uregina.ca/~hoeber/HotMap/>

Table 1. The relevance scores used to rate the document surrogates considered by the participants.

Score	Description
4	This document is relevant.
3	This document is probably relevant.
2	This document is probably not relevant.
1	This document is not relevant.

During the second task session, each participant was asked to perform a search first using *HotMap*, then using a Google-like interface. As the users considered each document surrogate, they were asked to provide a relevance score (see Table 1). This relevance score was recorded, along with the document surrogate number and the time. The task concluded when ten document surrogates were assigned a score of 4 (relevant).

Since all the participants were already expert Google users, varying the order of the interface provided to the user would have had little impact on their impression or performance in the study. But since the participants conducted the same search using the two interfaces, it is possible that they were able to recognize some of the documents. However, this recognition of previously seen documents is difficult to avoid since the participants were searching for topics which were familiar to them.

5 Results

5.1 User Performance

The user performance was evaluated based on two task goals: finding five relevant documents, and finding ten relevant documents. These goals were chosen to represent moderate and high levels of fulfillment of the users' information needs. The evaluation of the performance of the users was measured based on the time it took the participants to reach the task goal, and the relevance scores assigned to the document surrogates while reaching the task goal.

Since each participant in this study used a different query, it is difficult to aggregate the performance data. However, this study can provide insight into how this system may perform in real-world situations.

5.1.1 Time Efficiency

As the participants conducted their searches using the two interfaces, the time taken to find five and ten relevant documents was recorded. For each participant, their time using

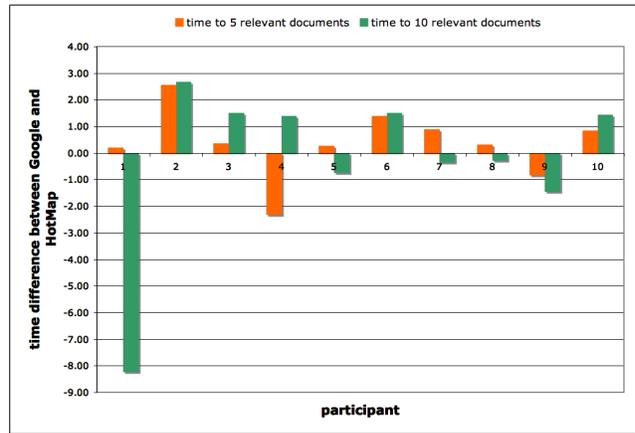


Figure 2. This graph depicts the time difference between using the Google-like interface and *HotMap* to find five and ten relevant documents. Positive values represent instances where the participant was faster using Google; negative values represent instances where the participant was faster using *HotMap*.

the Google-like interface provided a baseline, and the difference in the time taken to achieve the same goal using *HotMap* was calculated (see Figure 2).

Clearly, the difference in the time taken to complete the task goals using *HotMap* and the Google-like interface is quite varied. This variability in time can be attributed to the different queries used by the participants. For some queries, the search results were very specific; for others they were rather broad. When the results were very specific, it was easy for the participants to evaluate the search results in the Google-like interface quickly since many high quality documents often appeared on the first page of the search results. When the results were broad, it was quicker for the participants to explore the search results using *HotMap* since the high quality results were often distributed throughout the search results. In a number of cases, the participants were faster using the Google-like interface to find five relevant documents, but faster using *HotMap* to find ten relevant documents. In these cases, the remaining five relevant documents were buried deep in the search results; the exploration features of *HotMap* proved to be quicker to find these documents than the linear evaluation supported by the Google-like interface.

5.2 Relevance Efficiency

In addition to the amount of time taken to achieve the task goals, another indicator of efficiency is the number of

documents that were considered in order to find five and ten relevant documents. This relevance efficiency provides an indicator of the quality of the search results. If fewer non-relevant documents need to be considered to find a set of relevant documents, users will often consider the search results to be of higher quality than if they are required to consider a larger number of non-relevant documents.

While finding five relevant documents, some participants considered more low-relevance documents using *HotMap*; others considered more low-relevance documents using the Google-like interface (see Figure 3). While finding ten relevant documents, the benefit of using *HotMap* becomes clear: in 70% of the cases, the participants considered fewer low-relevance documents using *HotMap* than the Google-like interface.

From this data, we can see that there is a benefit to using *HotMap* to explore the search results when the goal is to achieve a high level of fulfilment of the information need. Doing so resulted in considering fewer low-relevance documents in all but three cases. In two of those cases (3 and 10), the participants did not assign any documents with low relevance scores using either interface. Therefore, only participant 6 considered more low-relevance documents using the Google-like interface when finding ten relevant documents.

5.3 Subjective Measures

After each search task, the participants were asked to evaluate the interface based on their confidence in their ability to find documents relevant to the search task, the ease of use of the interface for evaluating the search results, and their satisfaction with using the interface to evaluate the search results. These evaluations were assigned on a Likert scale, the results of which are summarized in Figure 4. Since these measures were taken after the participants found ten relevant documents using the two interfaces, they are not necessarily valid for the task goal of finding five relevant documents.

Participants showed a high level of confidence in their ability to find relevant documents using both interfaces, although *HotMap* scored marginally better in this measure. Given that all the participants used Google as their primary web search engine, we can assume that they are generally confident in the search results provided by that search engine. That they were just as confident using *HotMap* can be attributed to using the search results provided by the Google API. Further, the features of *HotMap* did not erode this confidence in the search results; for some participants, their confidence was even enhanced.

In terms of ease of use, *HotMap* scored better than the Google-like interface. Many participants reported the Google-like interface as moderately easy to use. While the

reading of the text would likely be reported as being very easy to do, the list-based representation and layout of the search results may have made the reading the search results not as easy to use as one might expect. By contrast, the responses regarding the ease of use of *HotMap* were mostly positive.

The participants reported a high degree of satisfaction with using *HotMap* to evaluate the web search results. Most participants were quite satisfied with the ability to manipulate and explore the search results to find relevant documents. With respect to the Google-like interface, the satisfaction responses were somewhat skewed towards a neutral reaction.

After all the tasks were completed, the participants were asked to rate their preferences for a web search interface, assuming that the search results were the same. 80% of the participants indicated that they would prefer to use *HotMap* over the list-based representation used by Google.

6 Conclusions

Wise et al., noting 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” [22], gave a very clear motivation for investigating other methods for presenting information retrieval results in non-textual manners. Providing the results of a search in a visual manner may allow this upper limit to be exceeded.

However, for a web search interface implemented as a meta-search, there are additional constraints beyond what would be present for a traditional information retrieval system. In addition to dealing with a collection that includes billions of documents, access to the textual contents of the individual documents is not provided with the search results, and there is an expectation from the users that the results be displayed in near real-time. For these reasons, we have found it necessary to restrict our visualizations to the information that is present in the document surrogates that are provided in the search results.

Although the process of counting the occurrences of the query terms in the document surrogates may seem simplistic, it has its basis in how some users evaluate the search results presented in the common list-based representation. That is, users will often try to find which documents make use of the terms in their query. While query term highlighting in the title and snippet do support this task, visualizing this information can be more effective since it allows the users to *see* the information rather than having to *read* the information. This has the added benefit of allowing the document surrogate to be presented in a far more compact manner. Since the users may still wish to see how their query

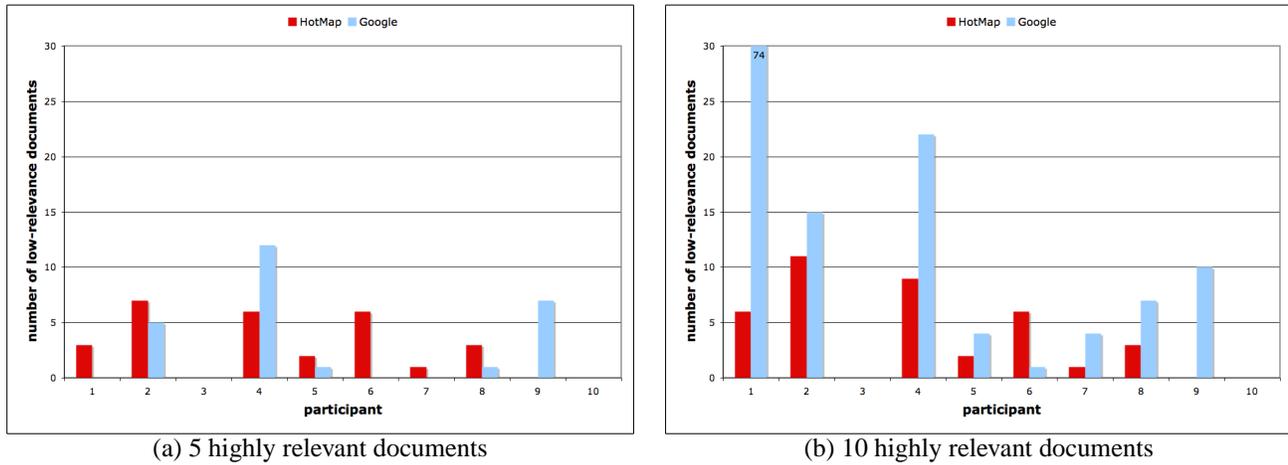


Figure 3. These graphs show the number of low-relevance documents (relevance scores of 1 or 2) viewed before finding 5 highly relevant documents (a) and 10 highly relevant documents (b). Lower values represent better performance (i.e., fewer low-relevance documents considered).

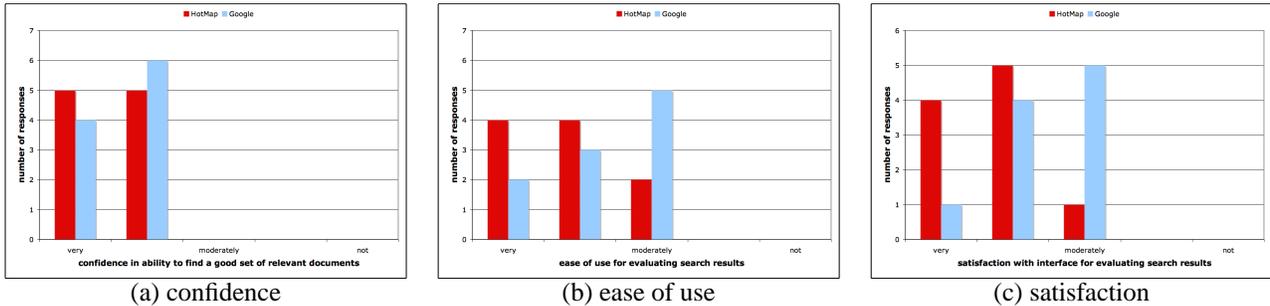


Figure 4. These histograms show the subjective reactions of the participants based on their confidence in their ability to find documents relevant to the search task (a), their ease of use with the interface for evaluating the search results (b), and their satisfaction with using the interface for evaluating the search results (c). In each of these subjective measures, *HotMap* received more positive responses than the Google-like interface.

terms are used in the document, access to the snippet is provided via a tool tip.

The ability to re-sort the search results based on query term frequencies support the users' tasks of manipulating the search results in order to more easily identify relevant documents. This, coupled with the ability to identify documents of interest in the overview map provides support for an interactive exploration of the search results. Although the user studies indicated that using *HotMap* may be slower than Google when finding relevant documents, this may not be a fair comparison since all the participants were expert Google users and novice *HotMap* users. What was clear from the user studies was that when seeking ten relevant documents, most participants considered fewer low-relevant documents using *HotMap* than using Google.

Further the subjective reactions and preferences for a web search interface were in favour of *HotMap*.

More complete user evaluations using a larger and more diverse participant pool are currently under way. Future plans include generalizing the meta-search interface so that other search engines can also be used; investigating the use of advanced search features such as phrases; and to integrate this work with our previous work on visual query expansion to create a complete visualization framework for web-based information retrieval.

References

- [1] K. Andrews, C. Gutl, J. Moser, and V. Sabol. Search result visualization with xFind. In *Proceedings of the Second In-*

ternational Workshop on User Interfaces to Data Intensive Systems, 2001.

- [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), 2000.
- [3] C. A. Brewer. www.colorbrewer.org, 2005.
- [4] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. In *Proceedings of the Seventh International World Wide Web Conference*, 1998.
- [5] S. Eick. Graphically displaying text. *Journal of Computational Graphics and Statistics*, 3(2), 1994.
- [6] Google. Google web API. www.google.com/apis/, 2005.
- [7] Grokker. Grokker search engine. <http://www.grokker.com/>.
- [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*, 1995.
- [9] M. Hearst. User interfaces and visualization. In R. Baeza-Yates and B. Ribeiro-Neto, editors, *Modern Information Retrieval*. Addison-Wesley, 1999.
- [10] T. Heimonen and N. Jhaveri. Visualizing query occurrence in search result lists. In *Proceedings of the International Conference on Information Visualization*, 2005.
- [11] Kartoo. Kartoo search engine. <http://www.kartoo.com/>.
- [12] Mooter. Mooter search engine. <http://www.mooter.com/>.
- [13] L. Nowell, R. France, D. Hix, L. Heath, and E. Fox. Visualizing search results: Some alternatives to query-document similarity. In *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval*, 1996.
- [14] T. Paek, S. Dumais, and R. Logan. Wavelens: A new view onto internet search results. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 2004.
- [15] M. Porter. An algorithm for suffix stripping. *Program*, 14(3), 1980.
- [16] D. Powers and D. Pfitzner. The magic science of visualization. In *Proceedings of the International Conference on Cognitive Science*, 2003.
- [17] E. Rennison. Galaxy of news: An approach to visualizing and understanding expansive news landscapes. In *Proceedings of the 7th Annual ACM Symposium on User Interface Software and Technology*, 1994.
- [18] C. Silverstein, M. Henzinger, H. Marais, and M. Moricz. Analysis of a very large web search engine query log. *SIGIR Forum*, 33(1), 1999.
- [19] A. Spink, D. Wolfram, B. J. Jansen, and T. Saracevic. Searching the web: the public and their queries. *Journal of the American Society for Information Science and Technology*, 52(3), 2001.
- [20] E. Tufte. *Envisioning Information*. Graphics Press, 1990.
- [21] C. Ware. *Information Visualization: Perception for Design*. Morgan Kaufmann, 2004.
- [22] J. A. Wise, J. J. Thomas, K. Pennock, D. Lantrip, M. Pottier, A. Schur, and V. Crow. Visualizing the non-visual: Spatial analysis and interaction with information from text documents. In *Proceedings of IEEE Information Visualization*, 1995.
- [23] Y. Yao. Information retrieval support systems. In *Proceedings of the 2002 IEEE World Congress on Computational Intelligence*, 2002.