

A Model for Interactive Web Information Retrieval

Orland Hoeber and Xue Dong Yang

University of Regina, Regina, SK S4S 0A2, Canada
{hoeber, yang}@uregina.ca

Abstract. The interaction model supported by web search engines has changed very little since the early days of web search. Users are required to formulate their queries with very little support from the system, and are provided with a list-based representation of the web search results that promotes a sequential evaluation of the document surrogates. The short queries used by web searchers, and the few pages viewed as a result of a web search are indications of the inadequate support provided for the users' information retrieval tasks. We propose a model for web information retrieval that uses visualization and interactive visual manipulation to support the users as they take an active role in satisfying their information needs.

1 Introduction

There have been significant advances in recent years in both the sizes of the indexes used by web search engines, and the algorithms used to match users' queries to these indexes [2, 3]. Modern web search engines can perform well when users enter very specific queries; many highly relevant documents are often provided in the first few pages of the search results. However, if users are unable to craft adequate queries, or when their information needs are inherently ambiguous, web search engines do not perform well.

Most web search engine interfaces support a model of interaction based on traditional information retrieval, consisting of a cycle of query formulation, examining the search results, and either stopping or reformulating the query [5]. The problem with this model is that there is little support for the users to perform the two fundamental tasks: query formulation and search results evaluation. Evidence of this inadequate support is provided in two independent studies which have shown that web searchers often provide only one or two terms in their queries, and seldom view more than three pages of search results [10, 9].

We propose a model for interactive web information retrieval that uses information visualization and interactive visual manipulation to support query formulation and search results exploration. Since the ability to read and assess textual information is a limiting factor [11], visual representations of the users queries as well as the search results can allow the users to more effectively interpret and make sense of the information provided. Interaction techniques allow

the users to take a more active role in the web information retrieval process, rather than the passive role supported by the traditional model.

In the remainder of this paper, we present a model for interactive web information retrieval, and provide a brief overview of the systems we have implemented based on this model: VisiQ [8] for interactive query refinement; HotMap [6] and Concept Highlighter [7] for interactive search results exploration. The paper concludes with a discussion and future directions.

2 Interactive Web Information Retrieval Model

Most web search engines use a traditional model of information retrieval [5]. This model consists of two primary stages: query formulation and search results evaluation. At some point in the search results evaluation stage, the users make a decision as to whether they are satisfied with the search results or not. If they are satisfied, the web search is considered a success. Otherwise, the users must decide whether they wish to reformulate their query. If they choose to not reformulate their query, the web search is considered a failure.

The two primary drawbacks of the traditional model are that there is little support for the users in formulating queries, and little support for users in evaluating the search results. Applying the principles of information visualization and interactive visual manipulation in order to support the users as they perform these complex tasks, we propose a model that allows the users to take an active role in the process. As illustrated in Figure 1, this model adds two new cycles in addition to the traditional query reformulation cycle: interactive query refinement and interactive search results exploration.

Interactive Query Refinement

With the traditional model, the query reformulation cycle can not occur until the users have first formulated an initial query and viewed the results of this search. If they make poor choices for their query terms that result in many non-relevant documents, there is little information provided to allow the user improve their query.

The interactive query refinement loop must provide *specific supplemental information* to the users that may be relevant as they craft a query. One of the challenges of query refinement is finding a source for this supplemental information that is relevant to the users' information needs. External, topic-specific knowledge bases may be necessary to resolve this challenge.

Any tool that provides support for interactive query refinement should allow the user to *focus on their query formulation task*. Therefore, any information provided should be within the context of the query. This is in contrast to the traditional model, where the query reformulation cycle flips the user between query formulation and search results evaluation.

Information visualization techniques can promote an easier interpretation and synthesis of the information provided regarding the users' queries. Further, the visual representations of complex data can result in much quicker comprehension than the equivalent textual representations. As such, *visual representa-*

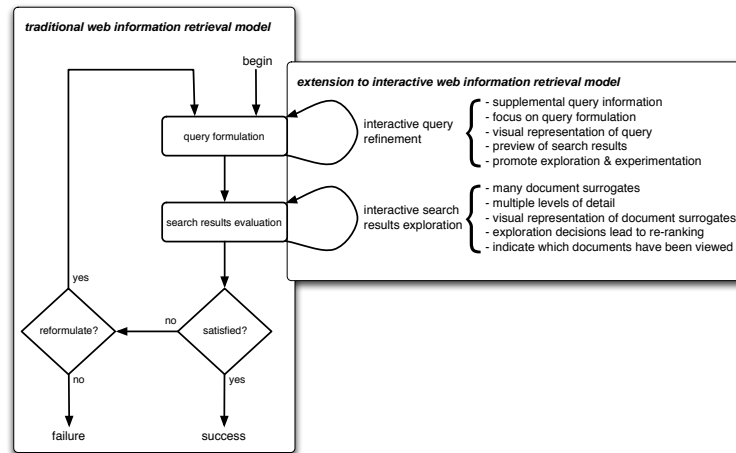


Fig. 1. The proposed model for interactive web information retrieval.

tions of the query information for an interactive query refinement tool is highly recommended.

As the users formulate their queries within this interactive query refinement loop, it is important to allow them to see how they are progressing. As such, *a preview of the search results* should be provided. This will help to promote an understanding of whether the users' queries are an accurate reflection of their information needs.

The interactive query refinement loop should *promote exploration and experimentation* with the query. As the users make refinement decisions, the results of these decisions should instantly be reflected in the visual representation of the query as well as in the preview of the search results. If minimal effort is required to return the users' queries to previous states, they will be able to easily experiment with "what if" scenarios, and explore the query space.

Interactive Search Results Exploration

In the traditional information retrieval model, the search results evaluation exists as a static stage in which the users evaluate the document surrogates one-by-one, and in the order provided by the web search engine. While this linear evaluation may be effective when there are a large portion of relevant documents in the first few pages of the search results, it is not very effective when the search results are ambiguous. It is in these cases that an interactive exploration of the search results is beneficial.

In order to promote an effective exploration of the web search results, *many document surrogates* must be retrieved and made available to the users. Providing a large number of document surrogates can allow the user to discover interesting features or patterns in the search results, and can direct the users to sets of documents that are more likely to be relevant.

Since a much larger number of document surrogates will be present in an interactive search results exploration tool, the user may wish to view both the details of specific document surrogates, and the features of the entire search results collection. Providing *coordinated views at different levels of details* can allow the users to take both a micro- and a macro-view of the web search results.

Just as information visualization techniques can allow users to more effectively make sense of their queries, generating *visual representations of the document surrogates* can allow the users to more effectively understand the features of the search results. Although converting the textual features of document surrogates into visual representations can be challenging, the benefits include the ability to provide a compact representation, and allow the users to “see” the information rather than “read” it.

As the users explore the web search results, their decisions should *result in a re-ranking* and subsequent re-sorting of the search results. This has the effect of bringing relevant documents to a more prominent location in the search results based on the features of the search results space. Further, the exploration decisions the users make must be reflected instantly in the visual representation of the search results. This will allow the users to more easily determine the outcome of their choices, and will allow them to decide whether their decisions are bringing them closer to fulfilling their information needs.

In any system that represents web search results, be it a static list or an exploratory system, it should be very easy for the user to access the specific documents. Further, as the users select documents to view, it should be *made clear that these documents have been viewed* when the user returns to the search results. This is especially important for a system that allows the users to re-rank the search results.

3 Implementation

Interactive Query Refinement

In order to explore the interactive query refinement cycle within this model, we have developed the VisiQ system to allow users to formulate and refine their queries within a visual interface [8]. This system makes use of a concept knowledge base devised for the topics in computer science that are stored in the ACM Computing Classification System [1].

The VisiQ system uses the concept knowledge base to visually represent the concepts related to the users query, and to suggest additional terms that may be relevant to their information need. This information is depicted in a graphical manner, allowing the users to easily identify their query terms, the concepts they are related to, and additional terms that have also been used to describe those concepts. The system allows the users to interactively refine their query by adding or removing terms via simple mouse click actions. A preview of the first five documents in the search results returned by the Google API [4] is provided as the users refine their query, allowing them to judge whether they are improving

the accuracy of their query to their information need. More details on this work can be obtained from [8].

Interactive Search Results Exploration

We have developed two interfaces to help us understand the complexities of web search results exploration. These systems use a similar search results interface framework that is based on a grid represented at two levels of detail: an overview map and a detailed view. Both systems send the users' queries to the Google API [4], and retrieve the top 100 document surrogates. The overview map provides a compact representation of all 100 document surrogates in a single view. The detailed view shows the specific information about each document surrogate, and provides a link to the document. As the users view the documents, the link colours change from blue to purple (as per the defacto standard for web links), allowing the users to easily see which documents they have already viewed. Both of these systems support techniques for re-ordering the search results; they differ in the supplemental information provided and how this information is used to explore the search results.

HotMap [6] calculates the frequencies of the users' query terms within the document surrogates and visually represents this information via colour coding. Each query term is provided as a column in the grid-based layout; the colour coding allows the users to easily identify "hot" documents which make frequent use of the users' query terms. The search results can be visually explored using the overview map, and a nested sorting feature allows the users to re-rank the search results based on the importance they place on the query terms.

Concept Highlighter [7] matches the users queries to the concept knowledge base in a technique similar to that used by VisiQ. These matched concepts are used as the cluster centroids by a single-pass fuzzy clustering algorithm which classifies each of the document surrogates based on their similarity to the concepts. The system allows the users to select one or more concepts that are related to their information needs. Doing so instantly re-ranks the search results based on the fuzzy membership scores, and visually depicts these scores both in an overview map and a detail window.

These systems were developed to address different situations that may arise when users are evaluating and exploring the search results. HotMap can be very effective when users have a moderate degree of knowledge about their information needs, but use one or more ambiguous terms in their queries. By re-ranking the search results based on the importance they place on the query terms, more relevant documents may be moved to the top of the search results. Concept Highlighter can be very effective when the results of a search are ambiguous. In these cases, documents that are similar to the concepts of interest can be given greater prominence.

4 Conclusions & Future Work

In this paper, we have proposed a model for web information retrieval that includes cycles of interactive query refinement and interactive search results

exploration. This is an advancement over the traditional information retrieval model, since it promotes the users in taking an active role in formulating and refining their query, and in evaluating and exploring their search results. This model is a step towards the development of a web information retrieval support system [12] that assists users as they seek the answers to their information needs.

Work is currently underway to generate a unified interactive web information retrieval system based on the various prototypes mentioned in this paper. The result will be a complete implementation that will fulfill the features of the model described in this paper. In addition to supporting interactive query refinement and interactive search results exploration, this system will provide a smooth and intuitive transition between the query formulation and the search results evaluation stages. We believe this unified system will result in a vast improvement over the existing web search engines, especially when the users' information needs are ambiguous, or when the users have difficulty in crafting specific queries.

References

1. ACM. ACM computing classification system. <http://www.acm.org/class/>.
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*, 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*, 2003.
4. Google. Google web API. www.google.com/apis/, 2005.
5. Marti Hearst. User interfaces and visualization. In Ricardo Baeza-Yates and Berthier Ribeiro-Neto, editors, *Modern Information Retrieval*. Addison-Wesley, 1999.
6. Orland Hoerber and Xue-Dong Yang. The visual exploration of web search results using HotMap. In *Proceedings of the International Conference on Information Visualization*, 2006.
7. Orland Hoerber and Xue-Dong Yang. Visually exploring concept-based fuzzy clusters in web search results. In *Proceedings of the Atlantic Web Intelligence Conference*, 2006.
8. Orland Hoerber, Xue-Dong Yang, and Yiyu Yao. Visualization support for interactive query refinement. In *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence*, 2005.
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), 2001.
10. Craig Silverstein, Monika Henzinger, Hannes Marais, and Michael Moricz. Analysis of a very large web search engine query log. *SIGIR Forum*, 33(1), 1999.
11. 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.
12. Yiyu Yao. Information retrieval support systems. In *Proceedings of the 2002 IEEE World Congress on Computational Intelligence*, 2002.