Automatic Topic Learning for Personalized Re-Ordering of Web Search Results

Orland Hoeber and Chris Massie

Abstract The fundamental idea behind personalization is to first learn something about the users of a system, and then use this information to support their future activities. When effective algorithms can be developed to learn user preferences, and when the methods for supporting future actions are achievable, personalization can be very effective. However, personalization is difficult in domains where tracking users, learning their preferences, and affecting their future actions is not obvious. In this paper, we introduce a novel method for providing personalized re-ordering of Web search results, based on allowing the searcher to maintain distinct search topics. Search results viewed during the search process are monitored, allowing the system to automatically learn about the users' current interests. The results of an evaluation study show improvements in the precision of the top 10 and 20 documents in the personalized search results after selecting as few as two relevant documents.

Key words: Machine learning, Web search, Personalization

1 Introduction

One potential problem with current Web search technologies is that the results of a search often do not consider the current interests, needs, and preferences of the searcher. The searcher's opportunity to affect the outcome of a search occurs only as they craft the query. The results for the same query submitted by two different people are the same, regardless of the differences between these people and what

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

Chris Massie
Department of Computer Science, Memorial University, St. John's, NL, A1B 3X5, Canada
e-mail: massiec@cs.mun.ca

they were actually seeking. This paper describes a method for automatically capturing information about the current interests of individual searchers, using this information to generate a personalized re-ordering of the search results. This solution is implemented in a prototype system called *miSearch*.

When modern information retrieval systems fail, in most cases it is due to difficulties with the system understanding an aspect of the topic being searched [2]. Clearly, the short queries that are common in searching the Web [7, 16] provide very little information upon which the search engine can base its results. The solution that has been employed by the major Web search engines is to return a large set of search results and let the users decide what is relevant and what is not. Our goal in this research is to capture additional information about what users think is relevant to their active search goals, and subsequently use this to re-order the search results. This work is inspired by the traditional information retrieval approach to relevance feedback [15], as well as the concept of "information scent" [13].

Personalization within the context of this research is defined as "the task of making Web-based information systems adaptive to the needs and interests of individual users" [12]. This definition highlights the two fundamental difficulties in personalization: how do we capture the interests of users in a non-obtrusive manner; and how do we adapt the system such that these interests are promoted and supported. With respect to *miSearch*, the first of these difficulties is addressed through *automatic topic learning*; the second is addressed through the *personalized re-ordering of Web search results*. A novel aspect of this work is the support it provides for users to create and maintain *multiple search topics*, such that the interests the searcher shows in one topic does not adversely affect their interests in other topics.

2 Related Work

Others have explored methods for personalization within the domain of Web search, including work from the top search providers, as well as in the academic literature. The Google search engine currently includes a personalization component that automatically learns searcher preferences through their search activities. The outcome is that searchers who have logged into the system are provided with a combination of personalized search results and recommendations [8]. Researchers at Yahoo! have investigated the use of data mining techniques on both the query and click data stored in their search engine logs [18]. The primary purpose in their work was to assess the potential for personalization. Although they found that it took a few hundred queries for distinct topics to become apparent, repeated site clicks were shown to be useful in identifying special interest topics.

Ahn et al. [1] developed a system directed at the exploratory search activities of expert searchers. Users can create and maintain notes about their search activities, from which a vector-based task model is automatically generated. The searcher may choose to view the search results sorted by relevance to the query, relevance to the task model, or relevance to both the query and the task model. Other features include the representation of the task model as a tag cloud, the creation of personalized snippets, and the highlighting of important terms in the snippets (using an approach similar to that in [5]). The utility of the proposed method was demonstrated via an in-depth user study.

Ma et al. [10] developed a method that maps user interests (from documents such as resumes) to categories in the Open Directory Project (ODP) [11]. These categories are then used to generate text classifiers, which are employed as part of the search process. When a user conducts a Web search, the full textual contents of the documents are retrieved and classified with respect to the categories in which the users have shown interest. The authors found the system to work well when seeking a small set of documents.

Sugiyama et al. [17] captured both long-term and short-term preferences based on the user's Web browsing activities. Gaps in the interest profiles are automatically filled based on matches to similar users. Clusters of all the user profiles on the system are generated; when conducting a search, the results are re-sorted based on their similarity to the clusters most similar to the searcher's profile. The authors found the system to be quite effective once sufficient information was gained to train the preference models.

A common theme among these Web search personalization methods is the use of complex techniques to capture the searcher's interests, and subsequently personalize the search results. In many cases, this move towards more complexity is necessitated by the single personalized profile maintained for each user. However, since searchers will commonly seek information on numerous topics that may have little relationship to one another, we suggest that a single profile is not appropriate. The method employed in our research (and implemented in *miSearch*) allows the searchers to maintain multiple topics of interest, choosing the appropriate one based on their current search activities. As a result, we are able to employ much simpler methods for capturing, inferring, and storing user interest in these topics, along with personalizing the order of the search results. The details of our approach are provided in the following sections.

3 Multiple Search Topics

Since people who search the Web have the potential to be seeking information on many different topics (sometimes simultaneously), creating a personalized model of their interests as a single collection of information may not be very effective. In some cases, a searcher may show particular interest in documents that contain a certain term; whereas in other cases, the same searcher may find all the documents that use this term irrelevant.

While it may be possible to deduce when the searcher has changed their search interests from one topic to another, a more accurate method is to have the user implicitly indicate their current topic of interest as an initial step in the search process. Such topics will form high-level concepts that provide a basis for collecting information about the searcher's preferences (as described in Section 4), and guide the subsequent personalized re-ordering of the search results (as described in Section 5).

When using *miSearch*, at any time during the search process a new topic can be created by the user. Similarly, the user may choose to switch to a previously created topic whenever they like. Since this process is not normally performed as part of a Web search, the goal is to make it as unobtrusive as possible. As such, we collect minimal information when creating new topics of interest, and allow the searcher to switch topics with just a simple selection from the topic list.

4 Automatic Topic Learning

When presented with a list of potentially relevant documents (e.g., a list of Web search results), searchers use many different methods for choosing which documents to view. Some scan the titles of the documents; others carefully read and consider the title and snippet; still others consider the source URL of the document. Regardless of what information searchers use, when they choose specific documents to view, there must be "something" in the information they considered that gave them a cue that the document might be relevant. The goal of the automatic topic learning process is to capture this "information scent" [13].

As users of *miSearch* select documents to view from the search results list, the system automatically monitors this activity, learning the preferences of each user with respect to their currently selected topic. Rather than sending users directly to the target documents when links in the search results lists are clicked, the system temporarily re-directs users to an intermediate URL which performs the automatic topic learning based on the details of the search result that was clicked. The system then re-directs the Web browsers to the target documents. This process occurs quickly enough so as to not introduce any noticeable delay between when a search result is clicked and when the target document begins to load.

The automatic topic learning algorithm uses a vector-based representation of the topic, with each dimension in the vector representing a unique term that appeared in the title, snippet, or URL of the search result clicked by the searcher. Selecting to view documents provides positive evidence of the potential relevance of the terms used to describe those documents; the topic profile is incrementally updated based on this evidence of relevance.

The algorithm takes as input the title, snippet, and URL of the clicked search result, as well as the searcher's currently selected topic of interest. The outcome of the algorithm is an update to the topic profile vector stored in the database. The steps of the algorithm are as follows:

1. Load the topic profile vector from the database.
2. Combine the title, snippet, and URL together into a document descriptor string.
3. Split the document descriptor string into individual terms based on non-word characters.
4. Remove all terms that appear in the stop-words list and words that are shorter than three characters.
5. Stem the terms using Porter's stemming algorithm [14].
6. Generate a document vector that represents the frequency of occurrence of each unique stem.
7. Add the document vector to the topic profile vector using vector addition.
8. Save the updated topic profile vector to the database.

5 Personalized Re-Ordering of Web Search Results

Once a topic profile vector has been generated, it is possible to use this information to re-order the Web search results. The goal of this re-ordering is to move those documents from the current search results list that are most similar to the topic profile to the top of the list. The premise is that the title, snippet, and URL of relevant search results will be similar to previously selected documents (as modeled in the topic profile vector).

The algorithm for re-ordering the search results receives as input the title, snippet, and URL of each document in the search results list, along with the current search topic selected by the searcher. The steps of the algorithm are as follows:

1. Load the topic profile vector from the database.
2. For each document in the search results list:
   a. Combine the title, snippet, and URL together into a document descriptor string.
   b. Split the document descriptor string into individual terms based on non-word characters.
   c. Remove all terms that appear in the stop-words list and words that are shorter than three characters.
   d. Stem the terms using Porter's stemming algorithm [14].
   e. Generate a document vector that represents the frequency of occurrence of each unique stem.
   f. Calculate the similarity between the document vector and the topic profile vector using Pearson's product-moment correlation coefficient [4].
   g. Save the value of the similarity measure with the document.
3. Re-sort the search results list in descending order based on the similarity measure.

While it would be possible to re-apply the personalized re-sorting technique as each document is viewed (and the topic profile is updated), it has been shown that such instant update strategies are not well-received by users, even when they provide more accurate results [3]. Clearly, usability issues arise when the search results are re-ordered interactively as a user selects to view a document and directs their

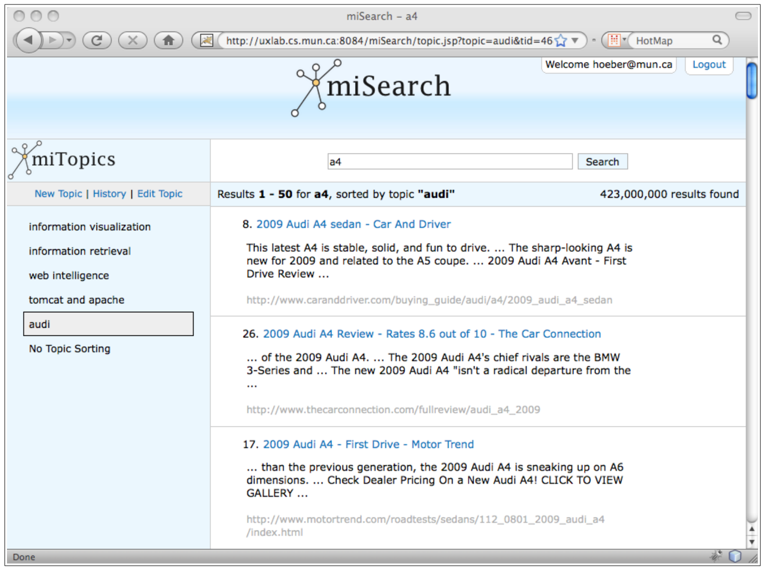


Fig. 1 A screenshot of the *miSearch* system. Note the personalized order of the search results based on previous selection of relevant documents.

attention away from the search results list. Instead, *miSearch* performs the personalized re-ordering of the search results only as each page of search results is loaded, or when users select new topics or re-select the current topic.

6 User's Model of Search

The user's model of search when using *miSearch* is altered slightly from the normal Web search procedures. In particular, users must first login, and subsequently select (or create) a topic prior to initiating a search. The login feature allows the system to keep track of multiple simultaneous users; the topic selection supports the personalization based on multiple topic profiles. The remaining process of evaluating the search results list and selecting potentially relevant documents to view remains unchanged.

The features described in the paper have been implemented in *miSearch*. The system currently uses the search results provided by the Yahoo! API [19], displaying fifty search results per page in order to provide a reasonable number of search results to personalize. Figure 1 shows a screenshot of the system. A public beta-version is currently available [1]; readers of this paper are invited to create accounts and use the system for their Web search needs.

[1] `http://uxlab.cs.mun.ca/miSearch/`

7 Evaluation

In order to measure the effectiveness of the Web search personalization methods described in this paper, twelve queries were selected from the TREC 2005 Hard Track[2] as the basis for the evaluation. In general, the queries in this collection represent topics that are somewhat ambiguous, resulting in search results that contain a mix of relevant and non-relevant documents. Queries were chosen to provide a range of ambiguity. The selected queries and a brief description of the information need are listed in Table 1.

For each of the queries, the top 50 search results provided by the Yahoo! API were retrieved and cached. The two authors of this paper, along with a third colleague, independently assigned relevance scores on a four-point relevance scale to each search result. Only the information provided by the search engine (title, snippet, and URL) was considered when assigning relevance scores. The possibility that a relevant document may not appear relevant in the search results list, or vice versa, is beyond the scope of this research. Discussions and consensus among the three evaluators resulted in *ground truth* relevance scores for each of the 50 search results produced for the twelve test queries.

Table 1 Queries selected from the TREC 2005 Hard Track for the evaluation of *miSearch*.

ID	Query	Description
310	"radio waves and brain cancer"	Evidence that radio waves from radio towers or car phones affect brain cancer occurrence.
322	"international art crime"	Isolate instances of fraud or embezzlement in the international art trade.
325	"cult lifestyles"	Describe a cult by name and identify the cult members' activities in their everyday life.
354	"journalist risks"	Identify instances where a journalist has been put at risk (e.g., killed, arrested or taken hostage) in the performance of his work.
363	"transportation tunnel disasters"	What disasters have occurred in tunnels used for transportation?
367	"piracy"	What modern instances have there been of old fashioned piracy, the boarding or taking control of boats?
372	"native american casino"	Identify documents that discuss the growth of Native American casino gambling.
378	"euro opposition"	Identify documents that discuss opposition to the introduction of the euro, the european currency.
397	"automobile recalls"	Identify documents that discuss the reasons for automobile recalls.
408	"tropical storms"	What tropical storms (hurricanes and typhoons) have caused significant property damage and loss of life?
625	"arrests bombing wtc"	Identify documents that provide information on the arrest and/or conviction of the bombers of the World Trade Center (WTC) in February 1993.
639	"consumer on-line shopping"	What factors contributed to the growth of consumer on-line shopping?

[2] `http://trec.nist.gov/data/t14_hard.html`

In order to determine the quality of a particular ordering of the search results, the precision metric was used. Precision is defined as the ratio of relevant documents retrieved to the total number of documents retrieved. For the purposes of this study, we considered any document assigned a score of 3 or 4 on the 4-point relevance scale as "relevant". Precision was measured at two different intervals within the search results set: P_{10} which measures the precision among the first 10 documents, and P_{20} which measures the precision among the first 20 documents. While it would be possible to measure the precision over larger sets of documents, the opportunity for improvements diminishes as we approach the size of search results set used in this evaluation. Note that while it is common in information retrieval research to also use the recall metric (ratio of relevant documents retrieved to the total relevant documents in the collection), the calculation of this metric with respect to Web search is not feasible due to the immense size of the collection (billions of documents) [9].

7.1 Hypotheses

Within this evaluation method, we use the precision achieved by the original order of the search results (as retrieved using the Yahoo! API) as the baseline performance measure. The two experimental conditions represent the performance of the system after selecting the first two relevant documents, and after selecting the first four relevant documents. Using the two levels of precision measurement discussed in the previous section (P_{10} and P_{20}), we arrive at four hypotheses:

H1: After selecting the first 2 relevant documents, there will be an increase in the precision among the first 10 documents in the re-orderd search results list.
H2: After selecting the first 2 relevant documents, there will be an increase in the precision among the first 20 documents in the re-orderd search results list.
H3: After selecting the first 4 relevant documents, there will be an increase in the precision among the first 10 documents in the re-orderd search results list.
H4: After selecting the first 4 relevant documents, there will be an increase in the precision among the first 20 documents in the re-orderd search results list.

7.2 Results

In order to determine whether the measurements from this experiment support or refute the hypotheses, we calculated the percent improvement (or deterioration) from the baseline measurements to the measurements after selecting two and four relevant documents. For all four cases under consideration, a statistically significant improvement was measured, as reported in Table 2. Significance was determined using ANOVA tests at a significance level of $\alpha = 0.05$.

Based on this statistical analysis, we conclude that H1, H2, H3, and H4 are all valid. As expected, the measurements also improve between selecting two and four

Table 2 Average percent improvement over baseline precision measurements. Statistical significance is verified with ANOVA tests.

Precision	2 Relevant Documents Selected	4 Relevant Documents Selected
P_{10}	**H1**: 89% ($F(1,23) = 16.36, p < 0.01$)	**H3**: 128% ($F(1,23) = 15.20, p < 0.01$)
P_{20}	**H2**: 40% ($F(1,23) = 9.64, p < 0.01$)	**H4**: 52% ($F(1,23) = 16.35, p < 0.01$)

relevant documents (H1 to H3, and H2 to H4). The decrease in precision between P_{10} and P_{20} is also to be expected, since as we consider a larger set of documents for relevance, the chance of non-relevant documents being included increases due to the limited number of documents available (e.g., 50 in these experiments).

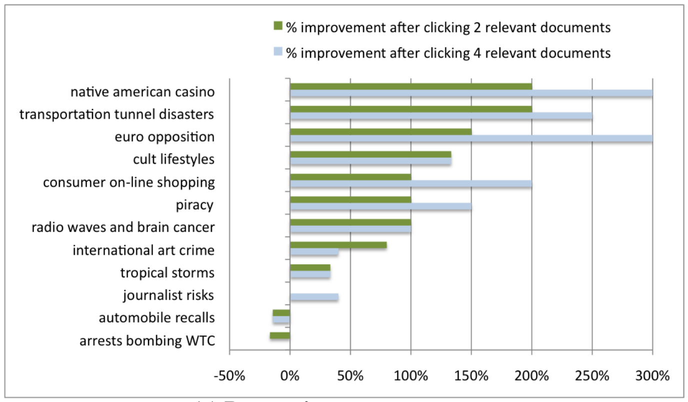
Our selection of test queries was intentionally chosen to provide a range of ambiguous queries. Since positive improvement was not discovered in all cases, it is worthwhile to consider the success of the technique with respect to each individual query. Figure 2 depicts the percent improvements over the baseline performance at both precision levels. In most cases, a significant increase in performance was found. However, in a few cases, the precision decreased as a result of the personalization.

Upon further analysis, we discovered that in all cases where there was a decrease in the precision scores with respect to the baseline ("automobile recalls" and "arrests bombing wtc" at the P_{10} level, and "journalist risks" at the P_{20} level), the baseline precision (from the original order of the search results) was already high (i.e., 0.6 or higher). The measured precision scores are provided in Figure 3. Clearly, in the cases where the precision measurements are already high, the ability to make improvements via personalization is limited. A logical conclusion from this is that personalization is of more value when the performance of the underlying search engine is poor, and of less value when the underlying search engine can properly match the user's query to the relevant documents.

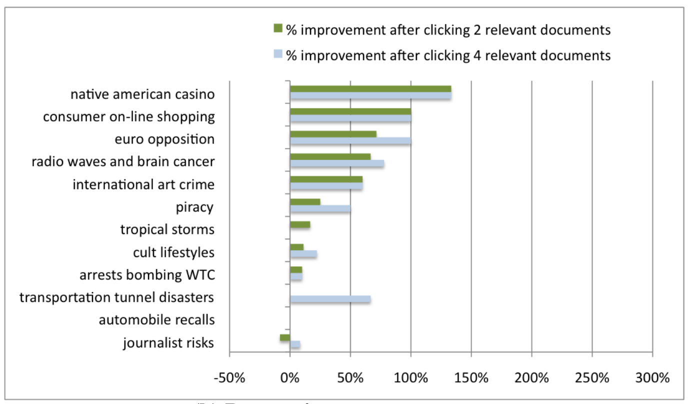
8 Conclusions & Future Work

This paper describes the key features of *miSearch*, a novel Web search personalization system based on automatically learning searchers' interests in explicitly identified search topics. A vector-based model is used for the automatic learning of the topic profiles, supporting the calculation of similarity measures between the topic profiles and the documents in the search results set. These similarity measures are used to provide a personalized re-ordering of the search results set.

An evaluation using a set of difficult queries showed that a substantial improvement over the original order of the search results can be obtained, even after choosing to view as few as two relevant documents. We attribute this success to the methods for allowing searchers to maintain multiple distinct search topics upon which to base the personalized re-ordering. This results in less noise during the automatic topic learning, producing a cleaner modeling of the searcher's interests in the topics.



(a) Percent improvement at P_{10}.

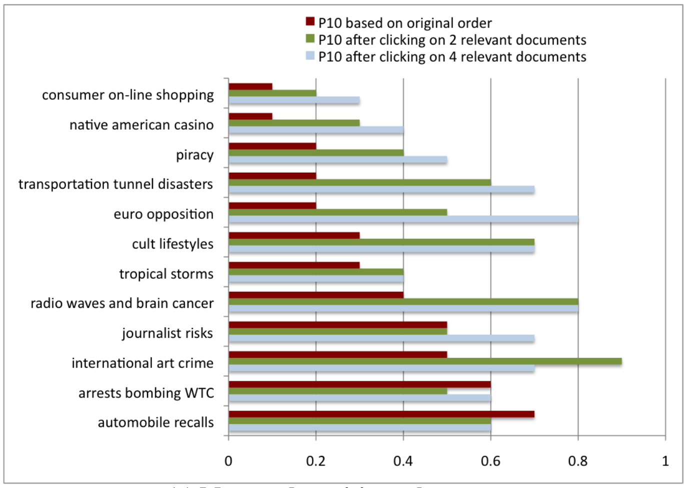


(b) Percent improvement at P_{20}.

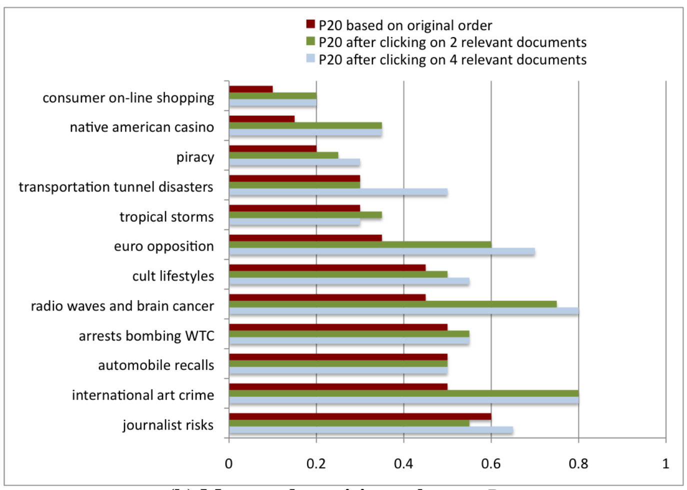
Fig. 2 The precent improvement over the baseline precision for each of the test queries, sorted by the degree of improvement after selecting two relevant documents.

Although the results reported in this paper have shown the methods used in *miSearch* to be very effective, we believe there is room for further improvement. We are currently investigating methods for re-weighting the contributions to the topic profile vectors during their construction, resulting in a dampening effect and the ability for the topics to model a user's changing understanding of their information need (i.e., topic drift).

Analysis of the techniques over a much larger collection of difficult search tasks, and under conditions where the searchers might incorrectly select non-relevant document to view, is needed to determine the robustness of the methods used in *miSearch*. In addition, user evaluations are in the planning stages, which will allow us to determine the willingness of searchers to pre-select topics during their search process. A longitudinal study will allow us to evaluate the value of the personalization methods in real-world search settings [6].



(a) Measured precision values at P_{10}.



(b) Measured precision values at P_{20}.

Fig. 3 The measured precision values for each of the test queries, sorted by the baseline precision (i.e., the original search results order).

Acknowledgements This research has been made possible the first author's Start-Up Grant provided by Faculty of Science at the Memorial University, as well as the first author's Discovery Grant provided by the Natural Science and Engineering Research Council of Canada (NSERC). The authors would like to thank Aaron Hewlett for assisting with the software development, and Matthew Follett for assisting with the relevance score judgements.

References

1. Ahn, J., Brusiloviksy, P., He, D., Grady, J., Li, Q.: Personalized web exploration with task models. In: Proceedings of the World Wide Web Conference, pp. 1–10 (2008)
2. Buckley, C.: Why current IR engines fail. In: Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 584–585 (2004)
3. He, D., Brusiloviksy, P., Grady, J., Li, Q., Ahn, J.: How up-to-date should it be? the value of instant profiling and adaptation in information filtering. In: Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence, pp. 699–705 (2007)
4. Hinkle, D.E., Wiersma, W., Jurs, S.G.: Applied Statistics for the Behavioural Sciences. Houghton Mifflin Company (1994)
5. Hoeber, O.: Exploring Web search results by visually specifying utility functions. In: Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence, pp. 650–654 (2007)
6. Hoeber, O.: User evaluation methods for visual Web search interfaces. In: Proceedings of the International Conference on Information Visualization (2009)
7. Jansen, B.J., Pooch, U.: 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)
8. Kamvar, S., Mayer, M.: Personally speaking. http://googleblog.blogspot.com/2007/02/personally-speaking.html (2007)
9. Kobayashi, M., Takeda, K.: Information retrieval on the Web. ACM Computing Surveys **32**(2), 114–173 (2000)
10. Ma, Z., Pant, G., Sheng, O.R.L.: Interest-based personalized search. ACM Transactions on Information Systems **25**(1) (2007)
11. Netscape: Open directory project. http://www.dmoz.org/ (2008)
12. Pierrakos, D., Paliouras, G., Papatheodorou, C., Spyropoulos, C.: Web usage mining as a tool for personalization: A survey. User Modeling and User-Adapted Interaction **13**(4), 311–372 (2003)
13. Pirolli, P., Card, S.: Information foraging. Psychological Review **106**(4), 643–675 (1999)
14. Porter, M.: An algorithm for suffix stripping. Program **14**(3), 130–137 (1980)
15. van Rijsbergen, C.J.: Information Retrieval. Butterworths (1979)
16. Spink, A., Wolfram, D., Jansen, B.J., Saracevic, T.: Searching the Web: The public and their queries. Journal of the American Society for Information Science and Technology **52**(3), 226–234 (2001)
17. Sugiyama, K., Hatano, K., Yoshikawa, M.: Adaptive Web search based on user profile construction without any effort from users. In: Proceedings of the World Wide Web Conference, pp. 675–684 (2004)
18. Wedig, S., Madani, O.: A large-scale analysis of query logs for assessing personalization opportunities. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 742–747 (2006)
19. Yahoo: Yahoo! developer network: Yahoo! search Web services. http://developer.yahoo.com/search (2008)