A Recommender System Approach to Enhance Web Search and Query Formulation

Michael O'Mahony, Maeve O'Brien, Oisín Boydell, and Barry Smyth

School of Computer Science and Informatics University College Dublin, Ireland {michael.p.omahony, maeve.m.obrien, oisin.boydell, barry.smyth}@ucd.ie

Abstract. While search engines are the primary means by which information is located online, significant issues remain when trying to satisfy the needs of searchers, especially in the face of the type of vague queries that dominate Web search. In this paper, we tackle this problem by applying a recommender system approach to Web search which allows users to dynamically interact with the result-space that is of interest to them. Our proposed recommendation interface also facilitates query expansion through a context-sensitive tag cloud, helping searchers to efficiently assimilate potential expansion terms that are mined from results of interest. We present findings from a live user trial of our approach which indicate, for example, that it facilitates users to locate relevant information more quickly when compared to using standard search engine result lists.

1 Introduction

Search engines continue to enjoy a privileged position in the work of the Web and are the primary means by which millions of users locate information online. However, despite the considerable advances that have been made in Web search engine technology, significant problems remain when it comes to helping people locate relevant information. For example, even leading search engines continue to struggle with the vague queries that are commonplace [10]. Even when users do specify more detailed queries, they often rely on niche terms that may not be well represented in the information items they seek [2]. And, of course, different users will always have different tastes and preferences, which are rarely expressed as part of a query and thus remain beyond the scope of conventional Web search engines. Together, these so-called vague query, vocabulary gap, and one-size-fitsall problems conspire to limit the effectiveness of Web search.

While the most significant search engine developments have focused on the core indexing and ranking components, the traditional search interface has remained relatively static. Generally there is little support provided to searchers when their queries fail to deliver relevant results. At best, some search engines offer limited query suggestion services, but typically users are left to reformulate queries themselves. We believe that there is significant potential to add value to existing search interfaces by harnessing recommendation techniques to offer users a more interactive search platform, particularly with respect to search discovery tasks [6], where users are searching for unfamiliar information. In this

context, we see recommender systems as playing a *supporting* role by providing an *overlay interface* as a complement to an existing search interface.

In this paper we describe a recommender systems approach to improving existing Web search interfaces. The system is designed specifically to support users in search discovery tasks by helping them to more efficiently navigate within a result-space returned by a search engine such as Google. We do this by harnessing recommendation techniques to support a form of *faceted search* [5, 13]. For example, searchers can choose to *expand* the result-space in the region of a particular result by clicking an icon that is added to each result in the resultlist. When a result is expanded, recommended results are promoted from deeper within the result-list, based on their relevance to the expanded result. In this way searchers can explore the result-space by incrementally expanding results that are relevant to their needs. In addition, we maintain a query cloud of the key terms that are mined from the titles and snippets of expanded results. This query cloud is an attempt to capture the essence of what appears to be relevant to the searcher, based on their interactions to date. The query cloud is designed to provide users with query expansion support by highlighting useful terms; we also describe a simple interaction tool which allows users to manipulate their queries directly by interacting with the query cloud.

In the next section we briefly outline recent research that is related to our work. In Section 3 we describe our core approach and how it has been integrated with a traditional search engine (Yahoo). Finally, in Sections 4 and 5 we describe the results of a live-user trial that highlight the potential benefits of this approach. We show, for example, that the proposed approach helps users to locate relevant information more quickly and we highlight how people tend to interact with our search recommendations, as well as describing the results of a comprehensive usage questionnaire.

2 Background Context

Our approach to enhancing Web search is to facilitate efficient navigation through complex result-spaces by allowing users to focus in on certain regions of the result-space, either by promoting results that are similar to those already selected by users or by suggesting query terms to refine current queries. These ideas are central to the *faceted search* paradigm [5, 13]. Faceted search is a combination of direct query-based search, as popularised by standard Web search engines, and navigational search in which information is located by browsing taxonomies which became popular through Web directories such as Yahoo Directory¹ and the Open Directory Project². In [4], a faceted search interface is proposed whereby a traditional query-based interface is augmented by a secondary view of the results which can be filtered by keywords extracted from the result list. Our proposed interface provides a form of faceted search, although the

¹ http://dir.yahoo.com/

² http://www.dmoz.org/

use of content-based recommendation techniques offers a novel navigation alternative to the use of strict taxonomies. Further, our use of a dynamic tag cloud as a form of interactive query refinement provides searchers with an effective means for them to better express their search needs within a session.

Of course recommender systems have been used previously to assist users when surfing and browsing the Web. For example, [3] present a related approach where relevant pages are suggested to Web surfers based on their past navigation histories. In addition, collaborative Web search (CWS) [11] is a community-based personalized Web search system where results are recommended to community members based on the past search history of the community as a whole. Similarly, CubeSVD [12] uses a modified collaborative filtering algorithm to identify similar searchers from search logs in order to re-rank and suggest relevant results.

The above systems base their recommendations on *persistent profiles* that capture the long-term navigation or search histories of users. In contrast our approach relies on short-term user feedback to drive its recommendations. A similar perspective is adopted by UCAIR [9], which also harnesses short-term search behaviour to dynamically re-rank Web search results. However, unlike UCAIR our system keeps the searcher informed during result-space exploration by summarizing the different information facets of their search through a dynamic tag cloud. This tag cloud also assists the searcher in expanding their original query by suggesting terms from the user's recent result-set interactions.

Query expansion techniques have been studied extensively within information retrieval [1] as a way to improve result relevance. Automatic query expansion techniques typically select expansion terms based on the top-ranking retrieved results [8]. These approaches make the assumption that all these top results are relevant to the searcher, and so an alternative is to use *implicit relevance feedback* to select expansion terms. For example, the results so far selected by the searcher might be used as a source of expansion terms [8]. We adopt a similar approach in this paper although expansion terms are presented in the form of an evolving tag cloud that is designed to support users when expanding or refining their queries, rather than suggesting explicit expansions or refinements.

From a search interface perspective, SurfCanyon³ is an excellent example of a next-generation approach, and allows users to interact with results returned by existing search engines; in short, users can select a result and receive recommendations drawn from related results that appear further down a result list. SurfCanyon launched in the latter stages of our own research and presents an interface similar to that proposed here. However, unlike our approach, SurfCanyon does not provide query refinement functionality directly to the user and as such fails to offer users with an alternative mode of result-space navigation.

3 Recommendation for Web Search

The system architecture utilises a client-server model allowing for the client-side component to be directly integrated into the user's browser as a plugin. Cur-

³ http://www.surfcanyon.com/



Fig. 1. A subset of the top-10 results retrieved for the keyword *jaguar*, with a set of recommended results displayed for result #2. The tag cloud shows recommended query terms, which are mined from expanded search results. Point-and-click query expansion of highlighted terms is facilitated via the pop-up menu.

rently, the plugin has been configured to work with Yahoo. When a new query is submitted, the plugin displays the top-10 results which are retrieved from Yahoo. In parallel, the plugin resends the same query to Yahoo (via an AJAX request to the recommendation server) to obtain further results; in total the top-200 results are retrieved. These results are clustered by the recommendation server and the partition obtained, together with the result-result similarity matrix, are sent to the client browser via AJAX (Section 3.2). These data are then available to the browser plugin to support user requests for recommendations (Section 3.3).

3.1 Example Search Session

Consider, for example, a user searching for information using the vague query *jaguar*. The results returned for this query are shown in Figure 1 and clearly reflect a variety of different interpretations of this query. Once the client plugin receives the result cluster information from the recommendation server, the default Yahoo search page is augmented with *result expansion* icons (Figure 1).

In this example, the user has expanded the second result (*Jaguar - Wikipedia*, pertaining to wild-cats). By default, the 3 most similar results from the cluster of the expanded result are suggested (also pertaining to wild-cats). These recommendations typically come from much lower down the default search listing; in this example, from pages 5, 14, and 19. Each recommended result can be further expanded and so on; in this way users can efficiently navigate through complex result-spaces using a form of 'random access' rather than 'sequential lookup'.

As users navigate through the result-space, the interface dynamically builds an interactive tag cloud by mining the title and snippet texts of expanded results. In this way, the tag cloud provides the searcher with an interactive summary of key terms, based on results that they have found to be interesting. By selecting a term the user can add it to their current query using a simple pop-up menu which provides access to a variety of search operators; in the example the user has selected the term *panthera* and is adding this as a quoted term to the query.

3.2 Clustering Search Results

Using the vector space model, each result r_i is represented by a vector in the term-space, which is defined by the title and snippet terms of the top-200 retrieved results. Following the removal of stop words and the application of word stemming, terms are weighted according to the standard TF-IDF scheme.

Using the clustering algorithm and the minimum total distance criterion described in [7], we produce a partition P by grouping the search results into a set of clusters $C = \{C_1, \ldots, C_n\}$. We also compute result-result similarity values (calculated as the cosine of the angle between result vectors) and return these data, along with P, to the client for use in the result recommendation algorithm.

3.3 Recommending Search Results

When a user selects a result r_i for expansion, recommendations are generated by selecting the n = 3 most similar results for r_i in cluster C_j , where $r_i \in C_j$. If the number of results in cluster C_j is less than n, then additional results are selected on the basis of similarity to r_i beyond the cluster C_j . The advantage of the clustering approach is that it discovers additional structure within a result set, over and above that achieved using result-similarities alone, thus leading to an improved likelihood of good quality recommendations being made.

3.4 Generating Tag Cloud Summaries

The tag cloud is generated from title and snippet terms of expanded results as follows. Suppose that result r_i is the k^{th} expanded result. Ignoring stop words, the weight of each term t is calculated as $w_t(k) = w_t(k-1) + n_t \ln(k+c)$, where c is a constant (set to 1) and n_t is the number of times that term t appears in the title and snippet of result r_i . $w_t(0) = 0$ and for k > 1, $w_t(k-1) = 0$ when term t has not appeared in the tag cloud previously. Accordingly, the cumulative weight (i.e. relevance) of each term is captured as the user interacts with the result-set. Terms with higher weights are displayed in larger font sizes in the tag cloud, thereby providing a visual and intuitive indication of term significance.

4 User Trial Methodology

To evaluate our approach to recommendation we conducted a live-user trial using 23 participants from University College Dublin, logging the search behaviour and obtaining their explicit feedback in the form of a post-trial questionnaire.

During the trial each participant completed a variety of search tasks by interacting with 3 different search interfaces:

- 1. *Interface 1* was the familiar Yahoo search interface to serve as a benchmark against which to judge performance compared to our enhanced interfaces.
- 2. *Interface* 2 incorporated the ability for users to expand results to generate cluster-based recommendations. It did not include the tag cloud feature.
- 3. *Interface 3* added the interactive tag cloud to Interface 2, thus providing users with a full range of search support.

In total 12 search tasks were designed to test search performance under conditions where users are known to encounter difficulty. For example, tasks were focused on topics with multiple meanings, or required users to search using acronym terms. Tasks were a mixture of fact- and homepage-finding searches, covering a range of typical search topics (travel, entertainment etc). Tasks were randomly ordered for each participant to control for any task effects. No time limits were placed on tasks and all tasks were to be completed in one session.

Before the trial users were presented with a brief tutorial on the relevant enhancements of Interfaces 2 and 3. During the trial participants were asked to complete 4 of the randomly ordered tasks for each interface condition. Participants completed each task by submitting an answer form after each interface condition. After the trial, participants completed a survey questionnaire to rate different aspects of the interfaces, overall user experience and satisfaction levels.

5 Results

Our evaluation was primarily concerned with three principal usability factors. First, we were interested in the overall *efficiency* of the interfaces in terms of their ability to help users to complete tasks with as few result 'clickthrus' as possible. Second, we considered the *effectiveness* of the interfaces in terms of their ability to present results that are relevant to the users needs. Finally, we were interested in the *satisfaction* level of users with the overall search experience. Our results indicate that, on the whole, users are more effective in their search using the result recommender and tag cloud features. Users reported positive subjective reactions to these features, and completed assigned search tasks more efficiently. These results are described in detail in the following sections.

5.1 Search Efficiency

Figure 2 (left) presents the average number of clickthrus per task per person. Compared to the Interface 1 benchmark, users of Interfaces 2 and 3 completed search tasks with fewer clickthrus. In each case, users of Interfaces 2 and 3 completed tasks with an average of 1.7 result clickthrus per search task, compared to 2.2 result clickthrus per search task for the standard Yahoo interface.

Figure 2 (middle) presents the average number of queries per task per person. Users of the standard Yahoo search interface required an average of 3.7 queries



Fig. 2. The average number of x lickthroughs per task per person (left), the average number of queries submitted per task per person (middle), the average number of query terms per-query per-person (right).

per task, whereas users of Interfaces 2 and 3 required 3.3 and 4 queries per task respectively. On the face of to interface 2 helped users to complete their search tasks with approximately 12% fewer queries than the Yahoo interface, whereas Interface 3 led to an increase (8%) relative to the Yahoo interface. This suggests that the result expansions of Interface 2 are helping users to discover results that would otherwise be buried deep in result-rankings, bringing to the surface results that would ordinarily require additional queries. However the increase associated with Interface 3 suggests that the tag clouds are actually promoting increased query submission. And while users of Interfaces 2 and 3 are completing their tasks with a similar number of clickthrus, these results suggest that Interface 3 users are often trading new query submissions for result expansions.

Figure 2 (right) plots the average number of query terms per query per person and echoes the trends in querying behaviour observed for average number of queries submitted per task per person. Users enter fewer terms per query in Interface 2 (mean=2.9), compared to 3.5 and 3.2 terms per query for Interfaces 1 and 3, respectively. This trend is to be expected given that users generally tend to expand queries on refinement, leading to an increase in query lengths.

5.2 Recommendation Effectiveness

Within the conditions of the trial where exploring recommended results was an option (Interfaces 2 and 3) users expanded on at least one recommended result 40% of the time; that is, in 40% of searches (i.e. individual query submissions resulting in the return of a single result-list) users selected at least one result for expansion. A key question is whether the resulting recommendations were found to be relevant. In fact, overall, across Interfaces 2 and 3, at least one of the 3 recommended results was selected approximately 25% of the time, a very significant indicator of relevance in a search context. Approximately 66% of these clickthrus were associated with the top recommendation, with 17% of clickthrus associated with both recommendations 2 and 3, respectively. This indicates a strong correlation between recommendation ranking and perceived relevance, although there is likely to be a strong order-bias in these clickthru rates.

Interestingly the percentage of users selecting recommended results is significantly higher in Interface 2 than in Interface 3. For example, for Interface 2, recommendations were selected 39% of the time compared with only 16% of the time for Interface 3. This is certainly an extremely strong indicator of recommendation relevance for Interface 2, especially since, on average, the default result list position (that is the Yahoo ranking) for recommended results is 82. Clearly these recommendations are frequently relevant and, given that they are for results which are so deeply hidden within the default result list, it is unlikely that they would have been located by searchers themselves.

But why should the clickthru rate be so much lower for Interface 3? After all, both interfaces use the same recommendation technique to generate the recommended results. We suggest that the reason for this clickthru disparity can be found in the tag clouds that are available in Interface 3. And while Interface 3 users only select recommendations 16% of the time, they interact with the tag cloud 56% of the time. To put this another way, the combination of result recommendations and term recommendations (via the tag cloud), promotes a positive user interaction in 62% of Interface 3 searches (some users both expanded results and interacted with the pop-up menu).

These results obviously raise interesting questions concerning the tradeoff that exists in search between result selections and query reformulation. The difference in recommendation clickthru rates between Interfaces 2 and 3 suggest that when faced with the option of either (a) 'gambling' on a result or (b) improving their query in the hope of finding a better result, users tend to opt for the latter. Indeed this tendency is reflected in overall performance where over 51% of queries resulted in no clickthrus; recall that the study was specifically designed to induce vague queries, in turn producing imprecise result lists.

Interestingly, for this study, a user's *preferred* behaviour was not necessarily the most *efficient* behaviour. Overall search efficiency was highest in Interface 2 where search tasks were completed with fewer result clickthrus *and* fewer query refinements. In Interface 3, the tag cloud helps reduce the cognitive load of query reformulation, and this seems to make users more likely to re-query in favour of result selection. And while there was a reduction in overall search efficiency in Interface 3 (in terms of the average number of queries required to complete a task) this is not to rule out the tag cloud as a powerful tool for users, and the results certainly highlight the fact that users found the tag cloud to be extremely useful in query formulation. Further, the results suggest that this feature encourages the use of more advanced queries. Users rarely use advanced search operators in regular search, and our results confirm this finding across both Interface 1 and 2 conditions. In Interface 3, however, there is a significant increase in the use of advanced search, with, for example, the use of the minus search operator increasing from 0% in Interfaces 1 and 2 to 9% in Interface 3.

5.3 Subjective Satisfaction

After completing all 12 search tasks each user completed a detailed questionnaire to rate their experience, on a scale of 1-7 on a variety of measures associated with

usability. Specifically, these measures were (a) the speed of finding what they were looking for, (b) the quality of results displayed, (c) the ease-of-use of the interfaces, and (d) the perceived value of the interfaces to the user. Although subjective ratings are statistically considered 'soft' measures, they do grasp a very critical issue in a user interface – even the most efficient user interface is useless if people do not *like* it. Scores of greater than 4 indicate a relatively positive response, whilst scores of 4 and below indicate possible dissatisfaction.

From the results, the users' evaluation of Interfaces 2 and 3 were generally positively skewed relative to the standard, which tended to more evenly distributed about the average. Due to limitations of space, we only present averages of the total scores across the 4 satisfaction measures here, which saw Interface 3 receiving the highest average rating of 4.7 followed by an average of 4.5 for Interface 2 and an average of 4.3 for the default Yahoo interface (Interface 1).

5.4 Discussion

It is worth highlighting that due to the limited number of trial participants, the observed differences reported in Section 5.1 are not statistically significant at the 0.05 level. Standard statistics rely strongly on population variance to compute significance, and the inherent variability in the user search population makes extracting reliable differences from limited data extremely difficult. In addition, search interfaces such as Yahoo and Google are perceived by most to work remarkably well. Users like, and are very familiar with, such interfaces and as such they are an extremely tough benchmark to meet, let-alone surpass. Taking into account the associated cost in adjusting strongly engrained search-routines, the observed level of user interactivity and satisfaction with respect to the proposed interfaces is, we believe, noteworthy and worthy of further investigation.

Users interacting with standard search interfaces tend to rarely search past the first page of results, they use vague queries, and rarely do they use advanced search operators. The recommendation features introduced here appear to help users in each of these areas, bringing their attention to relevant results that would otherwise be buried deep within the search rankings, and facilitating more advanced and specific query formulation. This clearly benefits user search performance, leading to a 29% decrease in the number of result selections per search task, and an overall increased level of user satisfaction.

6 Conclusion

In this paper, we have described a recommender system approach to Web search. The approach utilises the clustering of search results to inform a recommendation process, which enables users to explore the result-space in the region of selected results. Query expansion terms are also mined from expanded results and presented to searchers in the form of an interactive tag cloud. The results of a live-user trial indicate that this recommendation approach has the potential to improve search efficiency and leads to an improved user experience for searchers.

The trial results also raise interesting questions about the user's perceived cost of query formulation via text entry versus term selection, and further work is needed to more fully understand the conditions under which users gamble clicking on a result, over improving their query specificity. The results also point to the need for certain usability fixes, especially in relation to the manner in which users interact with the tag cloud. Future work will also test the search interface under a broader set of recommendation algorithms and experimental conditions, a larger pool of trial subjects, and over an extended period of time.

7 Acknowledgments

This work was carried out with the support of SFI Grant No. 03/IN.3/I361.

References

- R. A. Baeza-Yates and B. Ribeiro-Neto, editors. *Modern Information Retrieval*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1999.
- O. Boydell and B. Smyth. Enhancing case-based, collaborative web search. volume 4626/2007, pages 329–343. Springer, 2007.
- X. Fu, J. Budzik, and K. J. Hammond. Mining navigation history for recommendation. In Proceedings of the 5th international conference on Intelligent User Interfaces, pages 106–112, 2000.
- H. Joho and J. M. Jose. Slicing and dicing the information space using local contexts. In Proceedings of the First International Symposium on Information Interaction in Context (IIiX'06), 2006.
- 5. G. Marchionini and B. Brunk. Towards a general relation browser: A GUI for information architects. *Journal of Digital Information*, 4(1), 2003.
- V. L. O'Day and R. Jeffries. Orienteering in an information landscape: how information seekers get from here to there. In CHI '93: Proceedings of the INTERACT '93 and CHI '93 conference on Human factors in computing systems, pages 438–445, New York, NY, USA, 1993. ACM.
- B. Raskutti and C. Leckie. An evaluation of criteria for measuring the quality of clusters. In *Proceedings of the Sixteenth International Joint Converence on Artificial Intelligence (IJCAI99)*, pages 905–910, 31 July - 6 August, Stockholm, Sweden, 1999. Morgan Kaufmann.
- G. Salton and C. Buckley. Improving retrieval performance by relevance feedback. Journal of the American Society for Information Science, 41(4):288–297, 1990.
- X. Shen, B. Tan, and C. Zhai. Implicit User Modeling for Personalized Search. In Proceedings of the Fourteenth ACM Conference on Information and Knowledge Management (CIKM 05), 2005.
- C. Silverstein, H. Marais, M. Henzinger, and M. Moricz. Analysis of a very large web search engine query log. SIGIR Forum, 33(1):6–12, 1999.
- B. Smyth, E. Balfe, J. Freyne, P. Briggs, M. Coyle, and O. Boydell. Exploiting query repetition and regularity in an adaptive community-based web search engine. User Modeling and User-Adapted Interaction, 14(5):383–423, 2004.
- J. Sun, H. Zeng, H. Liu, Y. Lu, and Z. Chen. Cubesvd: A novel approach to personalized web search. In *In Proceedings of the 14th international conference on World Wide Web*, pages 652–662, 2005.
- K.-P. Yee, K. Swearingen, K. Li, and M. Hearst. Faceted metadata for image search and browsing. In CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems, pages 401–408, New York, NY, USA, 2003. ACM.