Collaborative Web Search*

Barry Smyth, Evelyn Balfe, Peter Briggs, Maurice Coyle, Jill Freyne
Smart Media Institute, University College Dublin, Ireland
{firstname.lastname} @ucd.ie

Abstract

Web search engines struggle to satisfy the needs of Web users. Users are notoriously poor at representing their needs in the form of a query, and search engines are poor at responding to vague queries. However progress has been made by introducing context into the search process. In this paper we describe and evaluate a novel approach to using context in Web search that adapts a generic search engine for the needs of a specialist community of users. This collaborative search method enjoys significant performance benefits and avoids the privacy and security concerns that are commonly associated with related personalization research.

1 Introduction

It is a tragic irony of the information age that users are finding it increasingly difficult to benefit from access to unprecedented amounts of information. Even with the help of modern search engines we regularly fail to locate relevant information in a timely manner. Many factors contribute to this *information overload* problem. Certainly, the sheer quantity of information, and its growth rate, tax even the most advanced search engines. For example, various estimates indicate that even the largest search engines often cover only a fraction of the available information space [Lawrence and Giles, 1999].

However, search engine coverage is just the tip of the iceberg, and can be greatly enhanced by using meta-search methods [Dreilinger and Howe, 1997; Selberg and Etzioni, 1997]. A more pressing problem is the limited degree to which those pages that are covered can be accurately ranked with respect to a given query. Part of this problem lies with the users. Search engines work very well for properly formulated queries, but they come up short when presented with an average Web user's query, which typically contains about 2 query terms [Lawrence and Giles, 1998]. The inevitable outcome is long lists of apparently relevant results, with genuinely useful results (for the target user) few and far between. Moreover, these problems are exacerbated by the new generation of mobile computing devices (eg. WAP phones and PDAs). Their restricted input capabilities and limited screen real-estate mean that mobile users are even less likely to provide well formulated queries or tolerate long result lists.

For the most part, recent search engine advances have focused on better ways to handle vague queries by improving existing page analysis, indexing and ranking methods. However a critical shortcoming still remains: a query might include terms that identify the primary information target, but rarely includes terms that usefully describe the search *context*. For example, a simple query for *"jaguar"* docs not indicate whether the user is interested in cars or cats, and queries for *"Michael Jordan"* do not distinguish between the basketball star and the Berkeley professor. Consequently, researchers have recently focused on ways to exploit context during search, either by explicitly establishing context up-front or by implicitly inferring it. For example, the Inquires 2 meta-search engine [Glover *et al.,* 2001] supplements keyword-based queries with a context category; users explicitly select from a set of categories such as "research paper", "homepage" etc. Alternatively, implicit context can be automatically inferred. For example, systems such as Watson [Budzik and Hammond, 2000] take advantage of user activity prior to search to judge context; Watson monitors a user's word processing activity and uses document text as the basis for query terms. In contrast, relevance feedback techniques attempt to use actual search results to inform context. For example, [Mitra *et al.,* 1998] extract correlated terms from top-ranking search results to focus context on the most relevant search results as opposed to the entire set.

In this paper we describe a novel technique for exploiting context during search: *collaborative search* acts as a front-end for existing search engines and re-ranks results based on the learned preferences of a community of users. We describe and evaluate its implementation in the I-SPY system and highlight how I-SPY achieves personalization in an anonymous fashion, without storing individual user profiles.

2 Collaborative Search

Collaborative search is motivated by two key ideas. First, specialised search engines attract communities of like-minded users and naturally limited context variations. For example, a motoring search engine is likely to attract queries with a motoring theme; here *"jaguar"* queries are more likely to relate to cars than cats. Second, by capturing the selections of

The support of the Informatics Research Initiative of Enterprise Ireland is gratefully acknowledged.

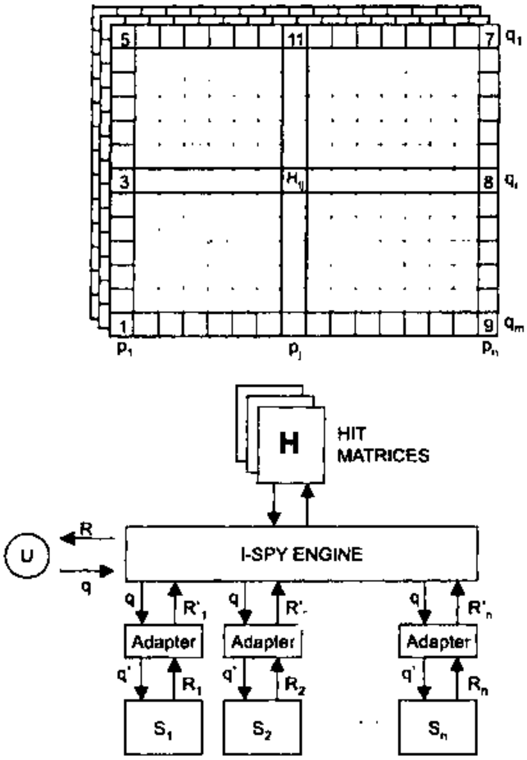

Figure 1: The I-SPY meta-search engine.

a community of users it is possible to estimate query-page relevance as the probability that a page p_j will be selected for query q_t. Collaborative search combines both ideas in a meta-search engine called I-SPY (Figure 1).

The unique feature of collaborative search is its ability to personalize search results for a community of users, but without relying on traditional context-analysis (eg. [Lawrence and Giles, 1998J) or personalization (eg. [Bradley *et al,* 2000]) techniques. I-SPY achieves this by storing the usage patterns of users as a *hit matrix, H.* Each element, $H_{i,j,}$ equals the number of times that page p_j was selected for query q_t. This matrix is a powerful source of relevancy information; after all its data reflect query-page relevancy judgments by users. The relevancy of p_3 to query q_z is estimated by the probability that P_j will be selected for query $q_{i;}$ sec Equation 1.

$$Relevance(p_j, q_i) = \frac{H_{ij}}{\sum_{\forall i} H_{ij}} \quad (1)$$

This relevancy metric is tuned to the preferences of a community of users, and the queries and pages that they tend to prefer. Deploy I-SPY on a motoring web site and its hit matrix will become populated with queries and pages that are relevant to car fans. For example, over time queries for *"jaguar"* will tend to result in the promotion of car sites because users submitting this query term will tend to select Jaguar car sites, ignoring the wild cat pages. The wild cat pages may still be returned but will be relegated to the bottom of the result list.

In fact I-SPY can deploy multiple I-SPY search agents, each with its own separate hit table. Thus the central I-SPY engine can be used to service many different search services across a range of portals, for example, each one adapted for the needs of a particular user group through its associated hit matrix. Alternatively, different hit matrices could be associated with different regions of the same site to bias search with respect to different topics. For instance, the work of [Havel i-wala, 2002] biases PageRank with respect to different topic groups in an Internet directory by generating category-biased PageRank vectors from the URLs contained in top-level directory categories. A similar strategy can be supported by 1-SPY. Placing a search box on the *Programming Languages* directory page will naturally capture queries from this domain. And the behaviour of the users providing these queries, will gradually adjust I-SPY's relevancy metric and ranking function in favour of Programming Languages pages.

3 Evaluation

For our evaluation we focus on a specific user community and search domain: computer science students and programming languages. A set of 60 queries is produced from the programming languages listed in Yahoo. I-SPY is configured to query two underlying search engines, Yahoo (which uses Google) and Splat!, and each of the 60 queries is submitted to obtain up to 30 results based on a standard meta-search ranking function. A group of 20 computer science students are asked to identify relevant results, based on the summary result descriptions returned by 1-SPY.

A *leave-one-out* evaluation methodology is employed so that each user is designated to be a *test* user with the remaining 19 serving as *training* users. The relevancy results of the training users are used to populate I-SPY's hit matrix and the results for each query are re-ranked using I-SPY's relevancy metric. Next, we count the number of these results listed as relevant by the test user for various result-list sizes (k = 5..30). Finally, we make the equivalent relevancy measurements by analysing the results produced by the untrained version of I-SPY *(Standard),* which serves as a benchmark.

Figure 2 presents the results for I-SPY and the benchmark search engine as a graph of precision versus recall for each result-list size; these are really *bounded* versions of the standard precision and recall metrics and the measures for each engine converge once a complete result-list is returned. The results indicate a significant and consistent benefit for I-SPY over the standard meta-search benchmark. For example, for result-lists of 5 items, I-SPY achieves a precision of just over 96% compared to the standard meta-search precision of only 63%. Similarly, at the same result-list size, we find an average recall for 1-SPY of 64% compared to just under 40% for the standard method. Indeed we see that I-SPY achieves 100% recall at just over 20 items whilst it takes the benchmark 30 items to achieve the same level of recall. The fact that larger relative benefits are available at smaller result-list sizes is important. Users rarely sift through large result-lists and so, the more relevant items that can be presented earlier on, the better. This means that I-SPY is likely to be especially valuable in situations where large result lists must be truncated for other reasons, such as the small screen sizes o f mobile devices.

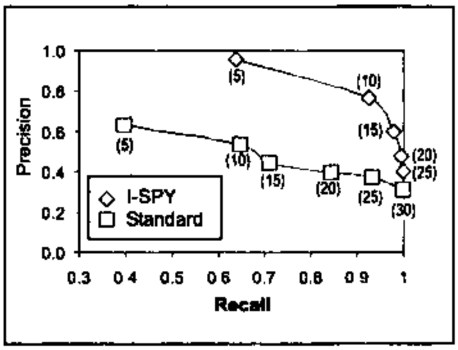

Figure 2: Precision vs. Recall Results.

4 Discussion

There are number of problems with collaborative search that need to be addressed to guarantee its applicability across a broad range of search tasks. Perhaps the most important problem is the so-called *cold-start problem.* This refers to the fact that newly indexed Web pages find it difficult to attract user attention since they will have a low relevancy score using 1-SPY's metric and thus appear far down in result-lists, limiting their ability to attract the hits they deserve for a given query. Essentially there is an in-built bias towards older pages.

There are a number of ways that this problem might be dealt with. One is to look at ways to normalize the relevancy of pages with respect to their *age.* For example, we might measure the age of a page by the time (or number of queries) since its first hit and amplify the relevancy of young pages relative to older pages.

Indeed there is another side to this problem. Just as new pages find it difficult to attract hits, so too older pages may find it easy to attract hits. In the worst case scenario this could even bias I-SPY's result-lists towards pages that are likely to be out of date and thus less relevant to current users than they were to past users. Once again, biasing relevance towards new pages should help to cope with this problem.

Of course in general there are many factors that can, and probably should, be taken into account when ranking search results. We have focused primarily on I-SPY's relevancy factor, but other factors such as the age of a page and its meta-search ranking are also appropriate. As part of our future work we will explore how best to combine these factors to produce optimal result rankings. This may or may not involve a direct combination of the rankings. For example, one option is to present search results not as a single list of results, as is normally the case, but perhaps as two or more lists of results in order to emphasise the different qualities of the returned pages. For instance, in general only a subset of search results are likely to have non-zero I-SPY relevance scores; that is, a subset of results will have been selected in the past for the current query. Therefore, it is practical to present the I-SPY results with relevancy scores as special recommendations (ranked by their relevancy). The remaining results can be presented separately, ranked by their meta-score. In turn a third list of *new* pages, ranked by meta-search score or relevancy, can also be separately presented.

5 Conclusions

Improving the accuracy of Web search engines by introducing context into the search process is an important and challenging research problem. We have described a generic search engine that can be adapted or personalized to fit the context and needs of a community of users by using the collaborative search technique. The benefits include superior precision and recall characteristics when compared to a benchmark search engine. In addition, this level of personalization is achieved without the need to store individual user profiles, leading to superior security and privacy benefits when compared to alternative approaches.

In closing it is worth highlighting that collaborative search makes no strong assumptions about the form of the underlying search engines and is generally applicable across a range of content types including Web pages, graphics and photos, audio and video. Finally, its proposed ranking metric is computationally efficient *(O(k)* in the number of search results) and requires no additional parsing of result pages.

References

[Bradley *et al,* 2000] K. Bradley, R. Rafter, and B. Smyth. Case-based User Profiling for Content Personalization. In O. Stock P. Brusilovsky and C. Strapparava, editors, *Proceedings of the International Conference on Adaptive Hypermedia and Adaptive Web-based Systems,* pages 62-72. Springer-Verlag, 2000.

[Budzik and Hammond, 2000] J. Budzik and K. Hammond. User Interactions with Everyday Applications as Context for Just-In-Time Information Access. In *Proceedings International Conference on Intelligent User Interfaces.,* pages 44-51. ACM Press, 2000.

[Dreilinger and Howe, 1997] D. Dreilinger and A. Howe. Experiences with Selecting Search Engines Using Meta Search. *ACM Transactions on Information Systems,* 15(3): 195-222, 1997.

[Glover *et al,* 2001] E. Glover, S. Lawrence, M. D. Gordon, W. P. Birmingham, and C. Lee Giles. Web Search - Your Way. *Communications of the ACM,* 44(12): 97-102, 2001.

[Haveliwala, 2002] T. H. Haveliwala. Topic-Sensitive PageRank. In *Proceedings of the World-Wide Web Conference.* ACM Press, 2002.

[Lawrence and Giles, 1998] S. Lawrence and C. Lee Giles. Context and Page Analysis for Improved Web Search. *IEEE Internet Computing,* July-August:38-^6, 1998.

[Lawrence and Giles, 1999] S. Lawrence and C. Lee Giles. Accessibility of Information on the Web. *Nature,* 400(6740): 107-109, 1999.

[Mitral*al,* 1998] M. Mitra, A. Singhal, and C. Buckley. Improving Automatic Query Expansion. In *Proceedings of ACMSIGIR,* pages 206-214. ACM Press, 1998.

[Selberg and Etzioni, 1997] E. Selberg and O. Etzioni. The Meta-Crawler Architecture for Resource Aggregation on the Web. *IEEE Expert,* Jan-Feb:l 1-14, 1997.