A Live-User Evaluation of Collaborative Web Search[*]

Barry Smyth, Evelyn Balfe, Oisin Boydell, Keith Bradley, Peter Briggs, Maurice Coyle, Jill Freyne
Smart Media Institute, Department of Computer Science,
University College Dublin, Ireland
{firstname.surname}@ucd.ie

Abstract

Collaborative Web search exploits repetition and regularity within the query-space of a community of like-minded individuals in order to improve the quality of search results. In short, search results that have been judged to be relevant for past queries are promoted in response to similar queries that occur in the future. In this paper we present the results of a large-scale evaluation of this approach, in a corporate Web search scenario, which shows that significant benefits are available to its users.

1 Introduction

Collection size, document diversity, and limited searcher expertise all combine to make the Web a very challenging information retrieval environment. In 2000 the entire World-Wide Web consisted of just 21 terabytes of information; now it grows by 3 times this figure every single day [Roush, 2004; Lyman and Varian, 2003]. Moreover, the average search query contains only about 2 query terms [Lawrence and Giles, 1998] and the terms used are often poorly chosen [Bollmann-Sdorra and Raghavan, 1993; Furnas *et al.*, 1987]. These problems have led to rapid developments in the term-based matching approaches at the heart of modern search engines. For the most part this has meant looking for new sources of knowledge with which to guide search. For example, Brin & Page [Brin and Page, 1998] and Kleinberg [Kleinberg, 1999] have argued for the need to consider factors such as link-connectivity information, while others have sought to exploit context as a way to disambiguate vague queries (see [Lawrence, 2000]). Still others have begun to consider the structure of the query-space as a new source of search knowledge. For example, [Fitzpatrick and Dent, 1997; Glance, 2001; Raghavan and Sever, 1995; Wen, 2002] have all demonstrated how query logs can be mined to identify useful past queries that may help the current searcher.

In [Freyne *et al.*, 2004; Smyth *et al.*, 2003; In Press], a novel approach to Web search—*collaborative Web search*—was introduced. It combined techniques for exploiting knowledge of the query-space with ideas from social networking to develop a Web search platform capable of adapting to the needs of (ad-hoc) communities of users. In brief, the queries submitted and the results selected by a community of users are recorded and reused in order to influence the results of future searches for similar queries. Results that have been reliably selected for similar queries in the past are promoted. For example, users of an AI-related Web site might have a tendency to select case-based reasoning results in response to vague queries such as *'CBR'*, while largely ignoring alternatives such as Google's higher-ranking *'Central Bank of Russia'* or *'Comic Book Resources'* results. In this instance collaborative search will gradually adapt its result-lists to emphasise case-based reasoning results, for searches that originate from such a site, perhaps through a search-box on the site.

While intuitively appealing, the collaborative Web search approach has never been fully evaluated under realistic conditions. Previous evaluations have been limited to the use of artificial users [Freyne *et al.*, 2004] or closed-world search scenarios [Smyth *et al.*, 2003; In Press]. In our work we have implemented the collaborative Web search technique as a robust and scalable meta search engine architecture and the central contribution of this paper is to evaluate its deployment in a realistic, real-world Web search setting involving the employees of a local software company over an extended period of time. While the results indicate that there is indeed a significant benefit accruing from collaborative Web search, they also serve to highlight certain issues, in relation to the manner in which promotions are made, that are likely to lead to critical problems over time. We conclude by discussing how these problems have been overcome in our implementation.

2 Regularity & Repetition in Web Search

Collaborative Web search is motivated by regularity and repetition that is assumed to be inherent in Web search, especially among the searches of communities of like-minded individuals. It proposes to exploit these regularities when responding to new queries by reusing the result selections from similar past queries. But how commonplace is community-based search? And how regular and repetitive is its query-space?

2.1 The Case for Community-Based Web Search

While most searches are conducted through generic search engines, servicing the needs of individuals, many are never-

[*] The Support of the Enterprise Ireland Informatics Initiative and Science Foundation Ireland is gratefully acknowledged.

theless examples of community-based searches. For instance, the use of a Google search box on a specialised Web site (e.g. a motoring enthusiast's site) means that its searches are likely to be initiated by users with some common (motoring) interest. Alternatively, searches originating from a computer laboratory assigned to 2^{nd} year students are likely to share certain characteristics related to their studies (courses, projects etc.) and social lives (college societies, local gigs etc.)

Of course, more formalised examples of community-based search are also possible. The advent of blogging services and social networking services such as Friendster and Orkut pave the way for a growing number of community-based search applications. While the precise nature of a community's shared interests may not be easy to characterise, they are nevertheless likely to be encoded within the search patterns (queries and result selections) of the community's members.

2.2 How Much Repetition?

If many searches can be traced back to ad-hoc communities of searchers, what degree of regularity can be observed? We can begin to answer this question by profiling the degree of term overlap between queries from different communities of searchers. One way to measure query similarity is by the degree of overlap between query terms as in Equation 1; for example, *Sim('jaguar pictures','jaguar photos')=0.33*.

$$Sim(q, q') = \frac{|q \cap q'|}{|q \cup q'|} \qquad (1)$$

Previous analyses of a variety of search engine logs have shown that query repetition is prevalent in specialised search scenarios that are likely to attract communities of like-minded searchers. For example [Smyth *et al.*, In Press] report how it is common to find that up to 70% of search queries may share at least 50% of their query terms with other queries'; this drops to 30% for more general search scenarios. Later in this paper we describe a major evaluation of collaborative Web search involving the employees of a local software company. Prior to this evaluation, we performed a similar query analysis over 9 weeks worth of search sessions extracted from the company's Internet access logs. Our working hypothesis at the time was that these employees would behave as a community of like-minded searchers and that their search queries would exhibit a high degree of similarity, thus motivating collaborative Web search.

The results are presented in Figure 1 as the percentage of queries at set similarity thresholds and the average number of similar queries for these different thresholds. The results show that the group of searchers do appear to behave as a community of like-minded users as high degrees of repetition are noted for many similarity thresholds. For example, we see that nearly 60% of queries share at least 50% of their query terms with other queries and that on average each of these queries shares 50% of its terms with about 6 other queries.

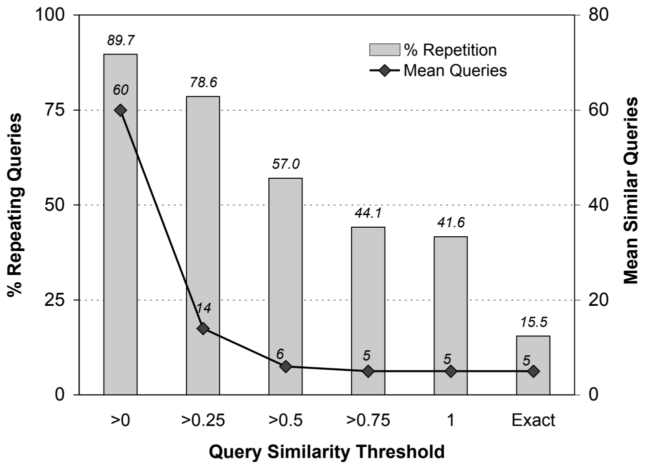


Figure 1: Query repetition in the search logs of the community used as part of the evaluation in Section 4.

3 A Review of Collaborative Web Search

The collaborative Web search technique is conceived of as a form of meta-search; see Figure 2 for the summary architecture and refer to [Freyne *et al.*, 2004; Smyth *et al.*, In Press] for a more comprehensive technical description. Briefly, each new user query, q_T, is submitted to a set of underlying search engines and their results are combined to form a meta-search result-list, R_M. The novelty of collaborative Web search stems from the way that this result-list is processed to produce a new result-list, R_T, that reflects the learned preferences of a community of like-minded searchers. It achieves this by recording the selections of searchers. In other words, collaborative search records the fact that a result s^i has been selected for query q_T, and then reuses this information for similar queries in the future, by promoting results that were reliably selected in the past.

3.1 Profiling Community Preferences

The *hit-matrix*, *H*, is a key data structure for collaborative Web search. It is a record of the results selected in past search sessions by a specific community of users, and multiple hit matrices can be readily maintained to reflect the separate preferences of many different communities. Each time a searcher (from a specific community) selects a result page, p_j, that was retrieved for query, q_T, the value of H_{Tj} is incremented. Thus, H_{Tj} represents the number of times that p_j has been selected as a result for q_T. The row of H that corresponds to q_T provides a complete account of the number of all page selections for this query over all search sessions that have used this query. Note that no record is maintained of which user selected which result, so in effect the hit matrix serves as an anonymous account of community preferences.

3.2 Reusing Similar Queries

The similarity between a new query, q_T, and a search record (row) in a hit-matrix can be estimated by the term overlap between the new query and the query of the past search record (Equation 1); see [Balfe and Smyth, 2005] for a number of alternative query similarity models. Collaborative Web search selects those rows from the hit matrix whose corresponding query has a similarity to q_T that is above some specified threshold (typically 0.5). The pages associated with these

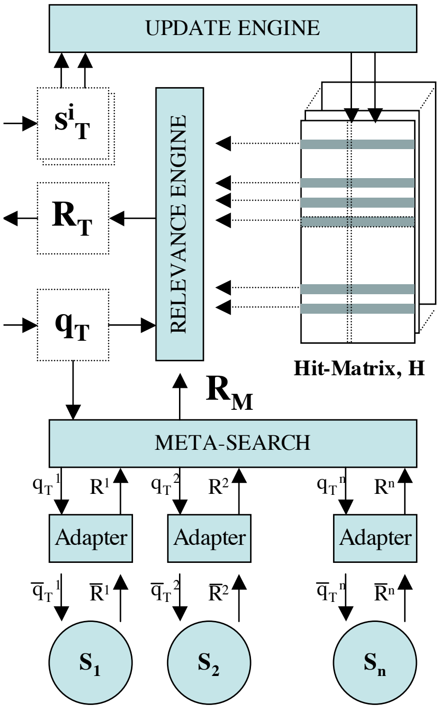


Figure 2: The collaborative web search architecture.

records (rows) are called *promotion candidates* and they are assumed to be potentially relevant to the new target query because they have been relevant for similar queries, and for the same community of searchers, in the past.

3.3 Result Relevancy & Ranking

Consider a page, p_j, that is associated with query, q_i. The relevance of p_j to q_i is estimated by the relative number of times that p_j has been selected for q_i; see Equation 2. And the relevance of p_j to q_T is a combination of $Relevance(p_j, q_i)$ for all q_i's $(q_1, ..., q_n)$ deemed similar to q_T, as shown in Equation 3. Each $Relevance(p_j, q_i)$ is weighted by $Sim(q_i, q_T)$ to discount the relevance of results from less similar queries; $Exists(p_j, q_i) = 1$ if $H_{ij} \neq 0$ and 0 otherwise.

$$Relevance(p_j, q_i) = \frac{H_{ij}}{\sum_{\forall j} H_{ij}} \tag{2}$$

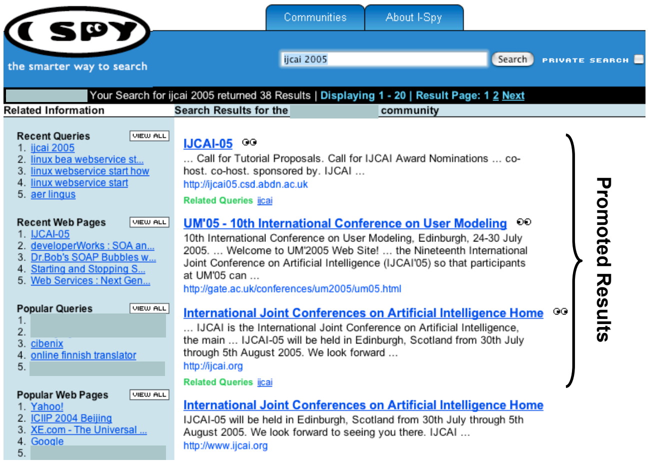
$$WRel(p_j, q_T, q_1, ..., q_n) = \tag{3}$$

$$\frac{\sum_{i=1...n} Relevance(p_j, q_i) \bullet Sim(q_T, q_i)}{\sum_{i=1...n} Exists(p_j, q_i) \bullet Sim(q_T, q_i)}$$

This weighted relevance metric is used to rank-order the promotion candidates. These ranked pages are then listed ahead of the remaining meta-search results, which are themselves ranked (according to a standard meta-search scoring metric), to give R_T. Of course, alternative promotion models can also be envisaged but are omitted here for space reasons.

3.4 Communities and Collaboration

Obviously this approach assumes that the contents of a given hit-matrix reflect some relatively uniform domain of interests, for a community of like-minded searchers. Collaborative Web search contemplates the creation of multiple hit-matrices to enable different communities of users to access a search service that is adapted for their query-space and their preferred pages. For example, a large Web portal might create different hit-matrices for different portal sections (e.g. News, Sports, Entertainment, Business sections) on the grounds that searchers are more likely to submit queries that are related to the content that is found within this portal section.



ijcai 2005

[redacted] Your Search for ijcai 2005 returned 38 Results | Displaying 1 - 20 | Result Page: 1 2 Next

Related Information

Search Results for the [redacted] community

Recent Queries
1. ijcai 2005
2. linux bea webservice st...
3. linux webservice start how
4. linux webservice start
5. aer lingus

Recent Web Pages
1. IJCAI-05
2. developerWorks : SOA an...
3. Dr.Bob's SOAP Bubbles w...
4. Starting and Stopping S...
5. Web Services : Next Gen...

Popular Queries
1. [redacted]
2.
3. cibenix
4. online finnish translator
5. [redacted]

Popular Web Pages
1. Yahoo!
2. ICIIP 2004 Beijing
3. XE.com - The Universal ...
4. Google
5. [redacted]

IJCAI-05
... Call for Tutorial Proposals. Call for IJCAI Award Nominations ... co-host. co-host. sponsored by. IJCAI ...
http://ijcai05.csd.abdn.ac.uk
Related Queries ijcai

UM'05 - 10th International Conference on User Modeling
10th International Conference on User Modeling, Edinburgh, 24-30 July 2005. ... Welcome to UM'2005 Web Site! ... the Nineteenth International Joint Conference on Artificial Intelligence (IJCAI'05) so that participants at UM'05 can ...
http://gate.ac.uk/conferences/um2005/um05.html

International Joint Conferences on Artificial Intelligence Home
... IJCAI is the International Joint Conference on Artificial Intelligence, the main ... IJCAI-05 will be held in Edinburgh, Scotland from 30th July through 5th August 2005. We look forward ...
http://ijcai.org
Related Queries ijcai

International Joint Conferences on Artificial Intelligence Home
IJCAI-05 will be held in Edinburgh, Scotland from 30th July through 5th August 2005. We look forward to seeing you there. IJCAI ...
http://www.ijcai.org

Figure 3: The I-SPY result page for the *'ijcai 2005'* query.

3.5 An Example Session

Collaborative Web search has been implemented in the form of the I-SPY search engine (http://ispy.ucd.ie). I-SPY can be configured to use a range of different search engines as its base-level search engines, including Google, Teoma, HotBot etc., and it allows users to use existing search communities or to create new ones via a simple form-based interface.

Figure 3 shows the results of a typical search for the query *'ijcai 2005'* by a particular I-SPY community. The result-list is presented in the main panel, flanked by recent and popular queries and web pages lists; certain sensitive information items have been blanked out in the figure. In this case the top 4 results are shown and the first 3 of these are result promotions; indicated by the *'I-SPY eyes'* icon next to the promoted result titles. This means that these results have been previously selected for this query or for similar queries. In fact we can see from the *'related queries'* lists after the first and third results that these have been previously selected for the similar query *'ijcai'*. The results shown are obviously relevant to the target query. The top result is for the main *IJCAI 2005* home page and the third result corresponds to the main *IJCAI Conferences* page, for example. However it is also worth noting that the second result is for the forthcoming user modeling conference, *UM 2005*. This page has been promoted because it has been selected in the past, for the current query, by members of the current community—these community members have a specific business interest in user modeling technology—but ordinarily this result would not be expected to appear so high in the result list for *'ijcai 2005'*.

This result is, however, relevant to this query given the community context, especially since *UM 2005* takes place directly before *IJCAI 2005* and in the same city. This type of promotion speaks to the potential power of I-SPY to promote results that are uniquely relevant to the specific needs of a community of like-minded searchers results that would ordinarily be lost among the competing results of traditional, generic search engines.

4 Live-User Evaluation

Past evaluations of collaborative Web search have included a mixture of artificial-user and live-user studies [Smyth *et al.*, 2003; Freyne *et al.*, 2004; Smyth *et al.*, In Press]. However these studies have been limited; for example, the live-user evaluation studied a narrowly focused, single-shot, question-answering search task which did not allow for more realistic open-ended search scenarios over an extended period of time. In this section we describe the results of a more realistic trial, which took place over a 4-week period among the 50 staff members of a local software company.

4.1 Preliminaries

The trial began on Monday, November 8, 2004 and the results presented in this paper account for the 4 working-weeks (Monday to Friday) up to and including December 3. During this time employees were asked to use I-SPY as their primary search engine; prior to the trial 90% of search sessions used Google. I-SPY was configured to draw on Google and HotBot as a source of search results and a new community was created for participants with a hit-matrix trained from search log data for the 9 weeks prior to the start of the trial. I-SPY's query-similarity threshold was set at 50%, so that only those past sessions that shared more than 50% of their query terms with the current target query would be considered to be similar for the purposes of result promotion (see Section 3). Participants were introduced to I-SPY via a short explanatory email and encouraged to use it as they would a normal search engine. Over the 4 weeks more than 1500 queries were submitted and more than 1800 result URLs were selected.

Figure 4 presents a histogram of the number of search sessions with different numbers of promotions. It shows that 46% of search sessions contained at least 1 promoted result; on average these sessions contained 3.7 promotions. This speaks to the potential for I-SPY's result-promotion technique to usefully influence a significant percentage of searches. The results are in broad agreement with the pre-trial query-overlap analysis described in Section 2, which suggested that we could expect up to 57% of new queries to have 50%-similar queries from the past to draw on as a source of promotions (see Figure 1).

4.2 Successful Sessions

While the above figures indicate that I-SPY is making promotions in roughly half of the search sessions, the real test is whether these promotions turn out to be relevant for the searcher, and whether they are *more* relevant than non-promoted results. Evaluating the relevance of search results in a trial such as this is difficult to do, at least in a direct fashion. Standard search interfaces do not provide a facility to allow users to indicate how well their information needs have been answered by search results, and while it would be possible to add such a facility to I-SPY for the purpose of measuring relevance in this trial, many users indicated that they would find this to be a nuisance. For this reason we examine a less direct measure of relevance.

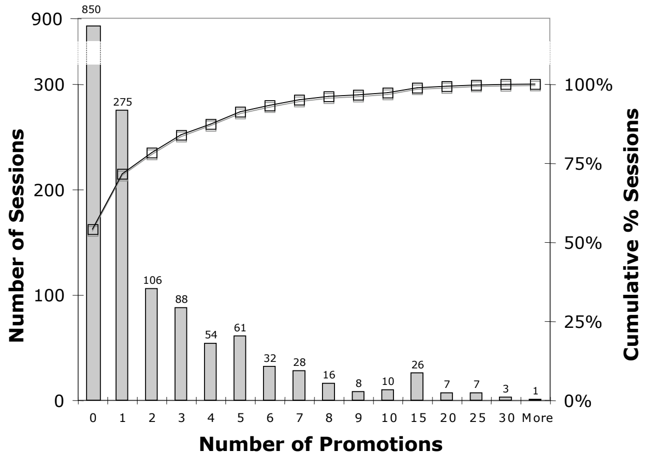


Figure 4: A histogram of the number of search sessions with different numbers of result promotions.

We propose that the selection of at least one result in a given search session acts as a crude, but nevertheless useful indicator of result-list relevance. We refer to a search session, where at least one result has been selected, as a *successful session*. If no results are selected (a *failed session*) then we can be relatively confident that the search engine has not retrieved a result that is *obviously* relevant to the searcher. Note that we do not distinguish here between sessions with different numbers of selected results, mainly because it is not possible to conclude much from the frequency of result selections. For example, one might be tempted to conclude that users selecting more results is a sign of increasing result relevance, except that a similar argument can be made in support of decreasing result relevance, on the basis that the initial selections must not have satisfied the users.

To analyse the ability of collaborative search to deliver successful sessions, we split the search sessions into those that contained promotions (*promoted sessions*) and those that did not (*standard sessions*). The former correspond to sessions where collaborative search has the potential to influence relevance, whereas the latter serve as a pure meta-search benchmark against which to judge this influence. Incidentally, there appears to be no difference between the queries for the promoted sessions when compared to those for standard sessions and both sets of queries have almost identical distributions; for example, an average of 2.4 terms per query for the promoted sessions compared to 2.5 for the standard sessions was measured. Indeed, given enough time it is likely that many of the standard queries would eventually be paired with new similar queries and so participate in future promoted sessions.

Figure 5(a) presents the average percentage of successful sessions among the promoted and standard sessions and

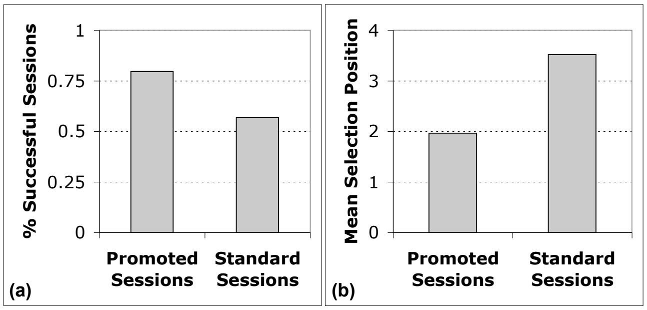


Figure 5: Promoted vs. Standard Sessions: (a) the percentage of successful sessions; and (b) the mean position of selected results among successful sessions.

demonstrates a clear advantage for the promoted sessions. On average, 80% of the promoted sessions were successful, compared to 56% for the standard sessions, a difference that is significant at the 99% confidence level. In other words, the collaborative search result-promotion mechanism leads to a 40% relative improvement in the chances that a given search will translate into a successful search session.

4.3 Selection Positions

As a complementary measure of result-relevance, it is also interesting to compare the promoted and standard sessions in terms of the average position of selected results within successful sessions; that is, those sessions in which selections have been made. We would like to see relevant results appearing higher up in result-lists. Moreover, assuming that users are likely to select results that at least appear to be more relevant than those that do not, then we would like to minimise the mean position of a selected result.

Figure 5(b) presents the mean position of the selected results among the successful sessions of the promoted and standard sessions. This once again shows a clear advantage for the former. On average, the mean position of a selected result among the successful promoted sessions is 1.96, compared to 3.51 for the successful standard sessions. This difference is statistically significant at the 99% confidence level and corresponds to a 44% reduction in the position of relevant results for promoted sessions compared to standard sessions.

It is worth commenting on the importance of this observed difference in the selection positions. While there is an advantage due to the promoted sessions, one might ask whether the observed reduction of one or two places is likely to be important. We believe that it is, for a number of reasons, not the least of which is that results should be ordered by their expected relevance as a matter of course. In addition, users have a tendency to focus their attention on the top-ranked results. The fact that promoted sessions have a higher success rate than the standard sessions is likely due to this difference in the position of apparently relevant results, because for the most part I-SPY promotes results from lower-down in the standard result-lists (returned by Google and HotBot) to higher positions.

This observed difference may become even more important in other search scenarios, such as mobile Web search, where screen-space is so restricted as to severely limit the number of results that may be presented on a single screen. For instance, on many mobile devices (eg. WAP phones), screen-space is so restricted that only 3 results can be presented per screen. The positional advantage enjoyed by I-SPY results suggests that it has the potential to ensure that relevant results will appear on the first page of such mobile-search results. In fact 99% of the result selections that occur in the promoted sessions are for results in the top 3 of a result-list, compared to only 79% of the standard session selections. Moreover, 93% of promoted session selections are for the top result, compared to only 63% of standard session selections.

5 Discussion

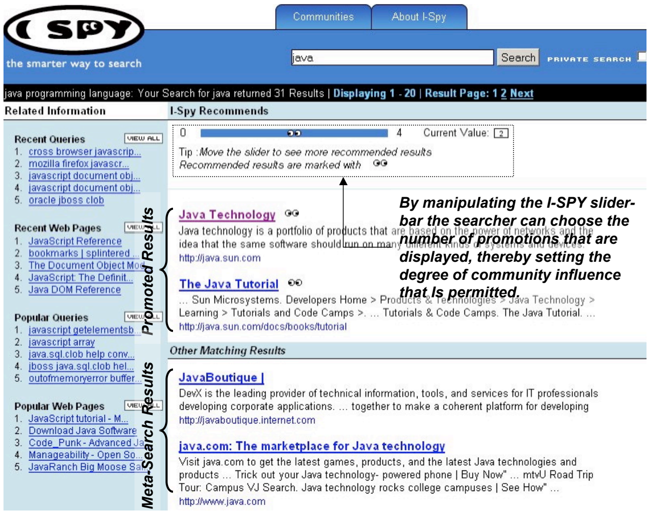
The results so far indicate that I-SPY's collaborative Web search has the potential to significantly improve search performance. Result promotions are made frequently, and when they are they translate into more successful search sessions and a better ranking for relevant results. In this section we briefly consider the number of promotions made during a session and the likely success of this session. In our analysis to date we have noticed that promoted sessions can contain up to 10 or more promoted results; this may be a problem because too many promotions may *swamp* result-lists to the detriment of search performance. In addition, sessions with many promoted results are likely to be caused by the reuse of large numbers of past search sessions, some of which may be the result of less reliable query overlaps, which in turn are more likely to contribute results of limited relevance to the target query.

One solution that we have adopted recently is to provide the searcher with a facility to adjust the level of community personalization that is offered, by manipulating a slider-bar to increase the number of promoted results that are displayed (see Figure 6). We are also considering different result-integration strategies to allow for a more flexible combination of I-SPY relevance and meta-search result scores.

We are concerned about issues relating to fairness, reliability and security. For example, it should be clear that as it stands, older results will tend to be preferred over newer results; the former will have had a greater opportunity to attract selections. This may cause problems when it comes to the promotion of very recent results. We are currently looking at ways to address this issue, for example by using a suitable decay function to gradually erode the selections of older results. We are also investigating ways to detect false selections by unreliable searchers as a way to defend against the fraudulent activities of self-interested parties; see also [O'Mahony *et al.*, 2002; Smyth *et al.*, In Press].

6 Conclusions

Collaborative Web search is an approach to Web search that exploits the natural regularity that exists within the search behaviours of ad-hoc communities of users. It espouses the reuse of search sessions for past queries that are similar to the current target query, resulting in the active promotion of those results that have been preferred by the community in


I SPY
the smarter way to search
Communities
About I-Spy
java
Search
PRIVATE SEARCH
Related Information
I-Spy Recommends
Recent Queries
VIEW ALL
1. cross browser javascrip...
2. mozilla firefox javascr...
3. javascript document obj...
4. javascript document obj...
5. oracle jboss clob
Recent Web Pages
1. JavaScript Reference
2. bookmarks | splintered ...
3. The Document Object Mod
4. JavaScript: The Definit...
5. Java DOM Reference
Popular Queries
1. javascript getelementsb...
2. javascript array
3. java.sql.clob help conv...
4. jboss java.sql.clob hel...
5. outofmemoryerror buffer...
Popular Web Pages
1. JavaScript tutorial - M...
2. Download Java Software
3. Code_Punk - Advanced Ja
4. Manageability - Open So...
5. JavaRanch Big Moose Sa...
0
4
Current Value: 2
Promoted Results
Other Matching Results
Meta-Search Results

java programming language: Your Search for java returned 31 Results | Displaying 1 - 20 | Result Page: 1 2 Next

Tip : *Move the slider to see more recommended results*
Recommended results are marked with ●●

Java Technology ●●
Java technology is a portfolio of products that are based on the power of networks and the idea that the same software should run on many different kinds of systems and devices.
http://java.sun.com

The Java Tutorial ●●
... Sun Microsystems. Developers Home > Products & Technologies > Java Technology > Learning > Tutorials and Code Camps >. ... Tutorials & Code Camps. The Java Tutorial. ...
http://java.sun.com/docs/books/tutorial

JavaBoutique |
DevX is the leading provider of technical information, tools, and services for IT professionals developing corporate applications. ... together to make a coherent platform for developing
http://javaboutique.internet.com

java.com: The marketplace for Java technology
Visit java.com to get the latest games, products, and the latest Java technologies and products ... Trick out your Java technology- powered phone | Buy Now" ... mtvU Road Trip Tour: Campus VJ Search. Java technology rocks college campuses | See How" ...
http://www.java.com

By manipulating the I-SPY slider-bar the searcher can choose the number of promotions that are displayed, thereby setting the degree of community influence that Is permitted.

Figure 6: I-SPY's promotion slider-bar allows the searcher to manipulate the degree of community personalization.

the past. In effect, this approach offers a form of anonymous personalization, thus protecting the privacy of individuals.

In this paper we have presented the results of a significant evaluation of the collaborative Web search approach in a corporate search environment. The results show the clear benefits that are available: the promotion of results translates into more successful search sessions and relevant results are ranked more highly. The evaluation has also helped to clarify the potential pitfalls of swamping result-lists with too many promotions. We have found that the best benefits arise from between 5-8 result promotions indicating the need for a mechanism to limit the number of promotions on a session by session basis. In response, we have described how I-SPY's interface has been adapted to include a means for searchers to interactively control the number of promotions at search time.

References

[Balfe and Smyth, 2005] Evelyn Balfe and Barry Smyth. An Analysis of Query Similarity in Collaborative Web Search. In *Proceedings of the European Conference on Information Retrieval*, pages 330–344. Springer-Verlag, 2005.

[Bollmann-Sdorra and Raghavan, 1993] Peter Bollmann-Sdorra and Vijay V. Raghavan. On the Delusiveness of Adopting a Common Space for Modeling IR objects: Are Queries Documents? *Journal of the American Society for Information Science*, 44(10):579–587, 1993.

[Brin and Page, 1998] Sergey Brin and Lawrence Page. The Anatomy of a Large-Scale Hypertextual Web Search Engine. *Computer Networks and ISDN Systems*, 30(1–7):107–117, 1998.

[Fitzpatrick and Dent, 1997] Larry Fitzpatrick and Mei Dent. Automatic Feedback using Past Queries: Social Searching? In *Proceedings of the 20th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 306–313. ACM Press, 1997.

[Freyne *et al.*, 2004] Jill Freyne, Barry Smyth, Maurice Coyle, Evelyn Balfe, and Peter Briggs. Further Experiments on Collaborative Ranking in Community-Based Web Search. *Artificial Intelligence Review*, 21(3–4):229–252, 2004.

[Furnas *et al.*, 1987] George W. Furnas, Thomas K. Landauer, Louis M. Gomez, and Susan T. Dumais. The Vocabulary Problem in Human-System Communication. *Communications of the ACM*, 30(11):964–971, 1987.

[Glance, 2001] Natalie S. Glance. Community Search Assistant. In *Proceedings of the International Conference on Intelligent User Interfaces*, pages 91–96. ACM Press, 2001.

[Kleinberg, 1999] Jon M. Kleinberg. Authoritative Sources in a Hyperlinked Environment. *Journal of the ACM*, 46(5):604–632, 1999.

[Lawrence and Giles, 1998] Steve Lawrence and C. Lee Giles. Context and Page Analysis for Improved Web Search. *IEEE Internet Computing*, July-August:38–46, 1998.

[Lawrence, 2000] Steve Lawrence. Context in Web Search. *IEEE Data Engineering Bulletin*, 23(3):25–32, 2000.

[Lyman and Varian, 2003] Peter Lyman and Hal R. Varian. How Much Information. *Retrieved from http://www.sims.berkeley.edu/how-much-info-2003 on January 14th, 2005*, 2003.

[O'Mahony *et al.*, 2002] Michael O'Mahony, Neil Hurley, and Guenole C. M. Silvestre. An Attack on Collaborative Filtering. In *Proceedings of the 13th International Conference on Database and Expert Systems Applications*, pages 494–503. Springer-Verlag, 2002.

[Raghavan and Sever, 1995] Vijay V. Raghavan and Hayri Sever. On the Reuse of Past Optimal Queries. In *Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 344–350. ACM Press, 1995.

[Roush, 2004] Wade Roush. Search Beyond Google. *MIT Technology Review*, pages 34–45, 2004.

[Smyth *et al.*, 2003] Barry Smyth, Evelyn Balfe, Peter Briggs, Maurice Coyle, and Jill Freyne. Collaborative Web Search. In *Proceedings of the 18th International Joint Conference on Artificial Intelligence, IJCAI-03*, pages 1417–1419. Morgan Kaufmann, 2003. Acapulco, Mexico.

[Smyth *et al.*, In Press] Barry Smyth, Evelyn Balfe, Jill Freyne, Peter Briggs, Maurice Coyle, and Oisin Boydell. Exploiting Query Repetition & Regularity in an Adaptive Community-based Web Search Engine. *User Modeling and User-Adapted Interaction: The Journal of Personalization Research*, (In Press).

[Wen, 2002] Query clustering using user logs. *ACM Transactions on Information Systems*, 20(1):59–81, 2002.