

Personalizing Web Search: Communities and Collaboration

Barry Smyth

Adaptive Information Cluster
School of Computer Science and Informatics
University College Dublin
Ireland
barry.smyth@ucd.ie

Introduction

Conservative recent estimates of the Web's current size refer to its 10 billion documents and a growth rate that tops 60 terabytes of new information per day (Roush 2004). In 2000 the entire World-Wide Web consisted of just 21 terabytes of information, now it grows by 3 times this every single day (Lyman & Varian 2003). This growth frames the information overload problem that is threatening to stall the information revolution going forward. In short, users are finding it increasingly difficult to locate the right information at the right time in the right way. Search engine technologies are struggling to cope with the sheer quantity of information that is available, a problem that is greatly exacerbated by the apparent inability of Web users to formulate effective search queries that accurately reflect their current information needs.

This talk will focus on how so-called personalization techniques are being used in response to the information overload problem (Billsus, Pazzani, & Chen 2000; Perkowski & Etzioni 2000; Reiken 2000; Smyth & Cotter 2002b; 2002a), and the experiences gained and lessons learned when it comes to the deployment of these techniques. In particular, we will focus on the personalization of Web search, taking special care to consider the important privacy issues (see also (Kobsa 2002) that such personalization brings to the fore. These issues motivate a unique approach to personalized Web search—Collaborative Web Search (CWS)—which focuses on the delivery of personalization at the level of a community of like-minded searchers.

Collaborative Web Search

The CWS approach is a form of meta-search, post-processing the results from some underlying search engine in response to the learned preferences of a community of searchers (Smyth *et al.* 2004; 2005). CWS exploits the high degree of query repetition and selection regularity that naturally exists in specialised search tasks (Smyth *et al.* 2004). This facilitates the promotion of results that are most likely to be relevant for the target query according to the preferences of the community in question. Search communities can be defined in a variety of ways. For example, one

common type of ad-hoc search community is formed by the searchers that tend to use a search-box as part of some specialised Web site. For instance, consider a motoring Web site with a Yahoo! search-box to make it easy for its visitors to initiate Web searches directly from the site. We believe that it is probable that users of such a search box are likely to initiate searches on a motoring theme. Thus, queries such as 'jaguar' are more likely to indicate an interest in cars rather than cats or operating systems, and the selection's of searchers should back this up. A different type of community might be made up of searchers within a particular organisation—for example, the students of a certain class or the employees of a certain company—and once again we might expect a degree of relatedness among the searches of community members.

The central novelty of CWS stems from its representation, recording, and reuse of community-based search histories. In short, when a community member submits a new search query q , this query is directed to the underlying search engine, but in addition, CWS also interrogates the search history of the particular community. Any results that have been consistently selected in the past by the community for this or similar queries are considered to be relevant and are used as the basis for promotion. This effectively means re-ranking the result-list that is returned by the underlying search engine so that such community-relevant results are boosted in the final result-list.

An Example Session

CWS has been implemented in 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, Yahoo, Teoma, HotBot etc., and it allows users to use existing search communities or to create new ones via a simple form-based interface.

Figure 1 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. 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 result titles.

(SPY)
the smarter way to search

Communities About I-Spy

ijcai 2005 Search PRIVATE SEARCH

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

Related Information Search Results for the community

Recent Queries VIEW ALL

1. [ijcai 2005](#)
2. [linux bea webservice st...](#)
3. [linux webservice start how](#)
4. [linux webservice start](#)
5. [aer lingus](#)

Recent Web Pages VIEW ALL

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 VIEW ALL

1. [REDACTED]
2. [REDACTED]
3. [cibenix](#)
4. [online finnish translator](#)
5. [REDACTED]

Popular Web Pages VIEW ALL

1. [Yahoo!](#)
2. [ICIIP 2004 Beijing](#)
3. [XE.com - The Universal ...](#)
4. [Gooble](#)
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>

Promoted Results

Figure 1: The I-SPY result page for the 'ijcai 2005' query.

These promoted 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 recent 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* took 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.

Deployment Experiences

The results of recent deployment trials of the CWS idea will be presented along with the lessons that have been learned from these deployments (Smyth *et al.* 2004; 2005; Boydell *et al.* 2005). These trial deployments point to a benefit for CWS in terms of result relevance, especially in the face of the vague search queries that are commonplace in Web search. These deployments have also helped to highlight opportunities for improving the basic CWS technique. For example, CWS contemplates the availability of many different search communities, each community reflecting the different search preferences and habits of a different collection of users. This perspective begs the question of how groups of communities might cooperate on specific search tasks: recent work that has explored this issue, by evaluating the potential for cross-community collaboration, will also be presented (Freyne & Smyth 2005).

References

- Billsus, D.; Pazzani, M.; and Chen, J. 2000. A learning agent for wireless news access. In *Proceedings of Conference on Intelligent User Interfaces*, 33–36.
- Boydell, O.; Gurrin, C.; Smeaton, A. F.; and Smyth, B. 2005. A Study of Selection Noise in Collaborative Web Search. In Kaelbling, L. P., and Saffiotti, A., eds., *Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence*, 1595–1597.
- Freyne, J., and Smyth, B. 2005. Communities, Collaboration and Cooperation in Personalized Web Search. In *Proceedings of the 3rd Workshop on Intelligent Techniques for Web personalization (ITWP’05) in conjunction with the 19th International Joint Conference on Artificial Intelligence (IJCAI ’05)*, 73–80.

- Kobsa, A. 2002. Personalized Hypermedia and International Privacy. *Commun. ACM* 45(5):64–67.
- Lyman, P., and Varian, H. R. 2003. How Much Information? Retrieved from <http://www.sims.berkeley.edu/how-much-info-2003/>.
- Perkowitz, M., and Etzioni, O. 2000. Towards adaptive web sites: Conceptual framework and case study. *Journal of Artificial Intelligence* 18(1-2):245–275.
- Reiken, D. 2000. Special issue on personalization. *Communications of the ACM* 43(8).
- Roush, W. 2004. Search Beyond Google. *MIT Technology Review* 34–45.
- Smyth, B., and Cotter, P. 2002a. Personalized Adaptive Navigation for Mobile Portals. In *Proceedings of the 15th European Conference on Artificial Intelligence - Prestigious Applications of Artificial Intelligence*. IOS Press.
- Smyth, B., and Cotter, P. 2002b. The Plight of the Navigator: Solving the Navigation Problem for Wireless Portals. In *Proceedings of the 2nd International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (AH’02)*, 328–337. Springer-Verlag.
- Smyth, B.; Balfe, E.; Freyne, J.; Briggs, P.; Coyle, M.; and Boydell, O. 2004. Exploiting query repetition and regularity in an adaptive community-based web search engine. *User Modeling and User-Adapted Interaction: The Journal of Personalization Research* 14(5):383–423.
- Smyth, B.; Balfe, E.; Boydell, O.; Bradley, K.; Briggs, P.; Coyle, M.; and Freyne, J. 2005. A Live-user Evaluation of Collaborative Web Search. In *Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI ’05)*, 1419–1424. Morgan Kaufmann. Edinburgh, Scotland.