COLLABORATIVE FILTERING: A NEW APPROACH TO SEARCHING DIGITAL LIBRARIES

Janet Webster, **Seikyung Jung** and **Jon Herlocker**

At Oregon State University (OSU), the School of Electrical Engineering and Computer Science (EECS) and the OSU Libraries are working together on a project to improve the effectiveness and accessibility of digital information created and collected by academic libraries. This project focuses on making digital resources more accessible through an innovative search interface that incorporates collaborative filtering. New approaches to search interfaces will help make the growing wealth of online content more accessible and useful. This paper discusses the problem, explains how collaborative filtering works, describes the System for Electronic Recommendation Filtering (SERF), and then presents initial results from an installation in the OSU Libraries. The productive collaboration at OSU between the Libraries and EECS portends the future of development of search systems; by working together, we can harness the expertise of librarians, computer scientists, and information users to develop more useful search interfaces and increase access to the libraries' resources and services.

Introduction

The continuing growth of digital resources has both advantages and disadvantages for libraries and their patrons. The wealth of accessible information stimulates thinking, enhances teaching and facilitates research. Yet, that same wealth can become an avalanche of irrelevant or inappropriate information. Digital information increases, and simultaneously, people's expectations of easy access to it heighten. Many library patrons prefer to use Google-like search interfaces rather than having to explore a library's carefully crafted web site or a subject specific portal. However, this common approach using current search engines often produces unsatisfactory or insufficient results because those search systems do not understand and integrate quality or context. At Oregon State University (OSU), the School of Electrical Engineering and Computer Science (EECS) and the OSU Libraries are working together to attempt to address this insufficiency and improve the digital library experience.

As a modern research library, the OSU Libraries have many digitally accessible resources. These include web-based interfaces for traditional journal indexes and collection catalogs to new digital special collections that incorporate highly interactive map-based interfaces. The quantity of these resources is so large that they cannot usefully be enumerated on a single web page. To provide access to these resources, the OSU Libraries expend extensive time organizing them into a navigable web site with logical links. This time and staff intensive work produces a tool with mixed use requiring constant maintenance.

A standard web search technology is also available as a search option of the OSU Libraries web site. However, its utility is poor and its usage is low. So, we recognize the need for a better search system to serve the needs of the library community.

What Is the Problem?

Existing Search Systems

The weaknesses of existing search technology suggest the need for a new approach. Existing search technology is content-based; it is based on matching keywords in a user query to keywords appearing in the full-text or content of a document. Such an approach fails to retrieve the best results because it is unaware of word context or higher level concepts, does not recognize synonyms, and cannot identify documents that have little or no text. In the university library context, the weakness of the keyword-based search is aggravated by the fact that many of the links to popular resources have little or no text associated with them.

Examples from the OSU Libraries include: the electronic journals list[1]; the databases of databases[2]; and the digitized Linus Pauling research notebooks.[3] As interactive tools, many electronic journals and most database indexes are designed (intentionally or not) so that their content cannot be indexed by web crawlers. The content contained in these resources is sometimes referred to as the 'deep web'. The Linus Pauling site, maintained by the OSU Libraries' Special Collections, contains images of the notebooks; without optical character recognition, images cannot be indexed by a search engine. However, perhaps the most notable weakness of content-based search systems is their lack of ability to differentiate the quality of search results; the simple presence of a keyword in a document does not indicate that the document is a high quality information source for the searcher's immediate need.

Researchers have proposed several possibilities to overcome some of these weaknesses of content-based search. Three approaches are used with varying success depending on the setting and the need:

- Link analysis uses back links (links from other web pages) and forward links (links from a web page to other pages) to create a type of citation analysis (Page et al. 1998). The assumption is that highly linked pages are more important. Google was the first commercially successful example of this strategy. Most commercial search engines now use some form of link analysis.
- Popularity-ranking works by tracking what links users prefer and use from the search engine's results page. It combines the number of clicks a link receives from the results page with the 'stickiness' of a link (e.g. the elapsed time between clicks from the results page.) This approach builds in user preference rather than web page creator's preferences through linking. The search engine, DirectHit (now part of Teoma, a web search facility that is part of Ask Jeeves Inc.), incorporates this approach.[4,5]

- Query categorization starts with the queries and looks for commonality among queries, and then returns search results that match similar queries. Two examples are MetaCrawler,[6] a search engine that concatenates search results from others, and FAQFinder (Burke et al. 1997), a natural language search engine of Frequently Asked Questions (FAQs).

Of these three approaches, the link analysis technique has been the most commercially successful. Many universities and research institutions including OSU have used a 'Google box' to run their institutional search engines. However, when applied to the OSU Libraries web site, we found that the Google approach does not significantly improve performance compared to traditional keyword-only searching (Jung et al. 2004a). This is because the library domain is not large enough to generate enough cross-links to effectively alter the search result rankings. Similar constraints apply to the other two approaches when used in a library setting. None of the generic web search engines are designed with the focused wealth of a library in mind and then built to utilize that wealth better.

Within the library realm, search system development focuses on our traditional finding tools such as catalogues and citation indices. Current work on federated searching ties the library's digital resources together so a patron moves seamlessly from the catalogue to a journal citation to full text (van de Sompel and Hochstenbach 1999). Linking in this way improves the library experience and increases use of library resources, yet it is limited to certain resources. Rethinking the library's search interface is challenging as it forces a rethinking of the scope of resources users need and expect. Users are no longer limited to the catalogue and the indices; they want to search across the vast information landscape. And, they want their searches to produce relevant results.

Library Patrons' Expectations

Creating an enhanced search system must address the variety of patrons and their expectations of not just the library, but the information searching experience. University library users vary in age, expertise and motivation; yet all want a satisfactory search experience. Several recently published studies on the searching behaviour and learning styles of academic library users point out the expectation of easy access, anytime and anywhere to everything needed at the time (Friedlander 2002; OCLC Online Computer Library Center, Inc. 2002). Demand is increasing for more electronic versions of print resources with obvious linkages between the citation source and the full text. Patrons also bemoan the lack of time to use library systems, so expect ease and speed.

Students are often more facile searchers than faculty and research staff. Consequently, focusing on their expectations and behaviour can inform system design. Students have eclectic needs that change semester to semester, and their interest level and investment varies with time and assignment. Abram and Luther (2004) summarize the expectations, learning behaviors and beliefs of the

NextGens (those born between 1982 and 2002) − the current wave of university students. Four of these characteristics are of particular interest in the context of designing and enhancing library search systems. First, students tend to be nomadic. Also, they are adept at multi-tasking so can juggle multiple searches and applications. Third, the students tend not to be concerned with the format of information and focus on the content; whether it is a journal article or a book matters not as long as it is accessible and informs their quest. Finally, they are collaborative as demonstrated by the high use of instant messaging and chat, and are willing to share information and opinions constantly.

Reshaping library search systems to accommodate these characteristics and expectations is challenging. Lossau (2004) notes the 'Library users have been 'empowered' by Google-like search engines to make their own choice about a search tool and to approach the world of information without any training'. This empowerment should push libraries to respond with new ways of increasing access to academic collections.

Integrating Human Judgement into a Library Search System

New library search systems must address both the technical issues of improving searching given the limitations of size and density of the library's digital resources, and the expectations of users for an easy and efficient interface. The usual internal focus on library-developed tools must expand so users can search the depth of library resources as well as the wealth of digital information outside of the library. Finally, we should recognize the role of users in helping libraries identify useful resources, and then efficiently direct users towards them.

A starting point for design is to examine why link-based search systems are successful. In part, their success relates to their ability to present relevant results fast and with an easy to maneuver interface. Setting interface design aside for the time being, the relevancy and quality of results is crucial to a successful system. The link-based systems assume that links are implicit records of human relevance judgments; a web page author links to an external page if that page is both relevant and perceived to have some value to the linking page. So, link-based analysis begins to integrate humans 'in the loop' of identifying relevant and high-quality documents just as the reference librarian traditionally does in the non-web environment. Reliance on linkages may be flawed given commercialization (e.g. paid or sponsored links reflect commercial interests rather than human judgment) or scale (e.g. specialized web pages may have a limited audience but are highly relevant to that audience). Assuming that significant human analysis can improve the search results, can we develop a search system that incorporates human analysis in a way other than tracking linkages? So, the OSU Libraries and EECS decided to apply collaborative filtering (CF) to the library search system problem. We perceive it as a technology with great potential to help us address the technical issues and user expectations.

Collaborative Filtering

CF is the process whereby a community of users with overlapping interests work together to separate interesting information from the non-interesting. In CF, each member of the community shares their evaluation of each content item they experience. Then each user can tap into the collection of past evaluations by all other members of the community, and use those evaluations to help select new information.

CF has its origins in the simple concept of informed recommendations that can be shared broadly. Gladwell's (1999) article in the *New Yorker* provides good motivation for exploring CF as he discusses the power of human input and opinion as applied to promoting small press books. He describes how a small bookstore owner with years of experience with books and customers can be compared to a sophisticated search system that compiles and synthesizes the recommendation and use patterns of people. Both can steer the user towards appropriate, relevant and perhaps surprising choices. Another example of the use of CF is an experimental system within the education sector described by Relker and Walker (2003). The authors explore how CF can enhance a collection of educational web resources by involving the users of those resources. A small group of teachers and students describe, review and rate the resources. These ratings and reviews then help the users select appropriate resources. The system attempts to leverage the opinions and expertise of the individuals to help the entire group or community of users. Finally, libraries have long used collaborative filtering to build shared catalogue record databases with OCLC's Worldcat collection being one example.[7] With OCLC, each member library contributes some amount of effort so that everybody can benefit from the resulting organization of content without a massive central expenditure.

CF, initially developed and implemented for entertainment and commerce, incorporates the results of human analysis of content on a massive scale through specialized search systems. Early implementation of CF focused on recommending items to individuals in entertainment related domains, such as music (Shardanand and Maes 1995), movies (Hill et al. 1995) and jokes (Goldberg et al. 2001). Amazon.com, perhaps the most well-known example of CF, has expanded its use of CF to cover anything sold from its catalogue. More recently, CF has begun to be applied to document searching and recommending. One of the most successful examples is the ResearchIndex system with the goal of recommending scientific literature using both citation analysis and explicit ratings given by readers (Cosley et al. 2002; McNee et al. 2002).

Of particular interest to the library search system we are developing is a research system called AntWorld, designed to help users manage their web searching better and to share their findings with other people (Menkov et al. 2000).[8] AntWorld is a web search support tool, where users describe their 'quests' before browsing or searching the web. When a user enters a new quest, that quest is compared to previously entered quests. At any point during a quest, a

user may choose to 'judge' the currently viewed web page, in essence rating its relevance to the quest. AntWorld has not been applied directly to the academic library setting nor tested for usability. It is of interest, though, because it matches similar queries rather than similar users. In the library environment, we do not compare one student to the next, but compare reference question to others.

The OSU Approach

OSU's System for Electronic Recommendation Filtering (SERF) attempts to incorporate CF while still taking advantage of link analysis searching. By so doing, the limitations of collection size and sparse descriptions may be overcome. SERF is a library web site search portal that incorporates explicit human evaluation of content on a large, de-centralized scale and tracks users' interactions with search results. By using specific information contexts rather than popular interests (e.g. movie or music preferences), SERF addresses the shared information needs among users and the consequent evaluation and recommendations of information retrieved. It can also incorporate the expertise of librarians and their judgments of relevant information.

The system grew out of collaboration between the OSU Libraries and EECS. The Libraries are committed to improving the library search experience for patrons, yet are limited by technical expertise in information retrieval systems. There is great interest in providing better access to traditional and digital collections, and integrating search systems into a single effective tool. Working with faculty and students in EECS provides a means to integrate the expertise of librarians into the development process. It also raises opportunities to explore differing perspectives on the search experience and on core values held by the Libraries. The collaboration provides a real world setting to explore the possibilities of a new approach. The Libraries' web site is not overly large, but it is very complex with content targeted at multiple and diverse users. The setting can be used to learn by observing patrons and react to their responses.

In SERF, a user issues a question or a statement of the current information need (e.g. a query). Users are encouraged, but not required, to enter grammatically correct, complete sentences. If previous users have issued similar queries, then SERF recommends documents, sites, or databases that those previous users found relevant and useful. The system determines that resources are relevant to a question by observing either an explicit user statement that the resource is valuable (e.g. a vote), or some user activity that implies that a resource is useful (e.g. emailing the page or the last page viewed). To find similar information needs, queries are compared using a keyword matching technique. An experimental prototype of SERF is currently available to all users of the OSU Libraries web site.

Traditional CF applications assume that users have fairly consistent information needs over time. People using a movie recommendation system will always be looking for a movie and their preferences will remain quite

constant. Consequently, these systems can match users with similar interests by simply comparing two users' historical profiles. This traditional CF approach does not work in the library setting where the needs of users can be different every time they come into the library environment.

SERF's goal is to match information contexts rather than users. An information context takes into account the user's immediate information need as well as past interests and even demographics. A user's query indicates the immediate information need. User profiles capture past interests and demographics if the user elects to maintain a login and profile. While each user may have different information needs every time the library is used, groups of users have similar information needs. For example, 300 students in a first year writing class would all need the same types of materials; if we can learn those needs by observing the first few students, then later students should be directed to the useful resources easily. On the other end of the spectrum, researchers in the biological sciences may have differing specialties, but would use similar resources. SERF addresses how the library can assist multiple people with similar information needs locate appropriate information. By matching information contexts, the system learns from the experience of the first person's to express the need and find resources for later users of the library.

The shared premise of the OSU Libraries and EECS is that by enabling the entire community of users to participate in organizing and recommending information, the libraries' resources will increase in value. They will be more accessible in terms of time spent searching and relevancy to a user's need. Increased access elevates the importance of the Libraries, both their collections and their services, while improving patrons' searching experiences. Ultimately, in addition to their searching function, search systems become tools for shaping resources and services collected.

Testing the Approach

The Test Bed

Since CF depends on user input or activity to learn, CF research requires real-world test-beds with active communities of users. Using the OSU Libraries web site as a test bed presents various challenges. Resolving how to integrate the system with existing library systems is currently addressed by keeping SERF separate from regular library systems. SERF currently accesses all library resources through standard web (e.g. http) interfaces. SERF essentially 'wraps' the web pages of the library by re-writing URLs, so that all clicks are first routed through SERF. This approach means that library systems staff members do not have to be involved in ongoing maintenance of another system, yet we can track every click and every link the user selects.

Recommendations are only as good as the people making them, so noisy and untrustworthy data can be problematic. So far, malicious voting is not a problem, but as use scales up, it will be monitored. Interface design is critical and

benefits from ongoing user feedback from librarians and library patrons. Finally, CF systems traditionally rely on a small number of willing and invested users to actively recommend resources. Given the sporadic use of certain library resources and the often small numbers of users, SERF developers are exploring how to infer recommendations from user behavior. One promising indicator appears to be the last page viewed in a query session (Jung et al. 2004b). This indicates the end of a query session suggesting the user found the answer.

SERF in Detail

A link to SERF appears on the OSU Libraries home page. It is described as 'an experimental search engine'. Users can also select the generic search option that uses a Google-like engine to search the web site.

Initial search page. On the initial search page, users can log into SERF anonymously (Figure 1). For returning users, the search interface page also contains a list of links to previous questions asked by the user, resources that are frequently visited by the user, and explicit bookmarks that the user has created. Users are encouraged to use a complete, natural language sentence rather than a few keywords. The longer query establishes more information context (Anick 2003; Belkin et al. 2003). An example of a query is automatically displayed in the query box as one mechanism to encourage longer statements. Once users log in and submit text queries, their activity is tracked using wrapping technology.

Search results and recommendations. After receiving a search query, SERF compares the current user's query with past information contexts, locates those that are the most similar, and recommends the resources that were valuable to those past, similar information contexts (Figure 2). Associated with each recommendation is the original question from the previous information context. The question is displayed alongside the recommendation. By examining those questions, the user then determines if the recommended past information contexts are truly related to their immediate information need.

No more than two of the most similar questions and up to three of their highest rated documents are displayed. If more than two similar questions exist, users may elect to view additional similar questions by clicking a link; similarly, if more than three documents are rated highly for a question, users can also choose to view the other documents by clicking a link next to the recommended question. Limiting the immediate display is a way of improving precision in search results.

Retrieving and viewing information resources. SERF will learn and grow its capacity to recommend as usage increases. Early users and users with new queries will have limited or no results. To overcome this, results from the Google-like search engine running on the Libraries web site are displayed. Users then can peruse both recommended resources (if any) and those returned by the Google-like search engine.
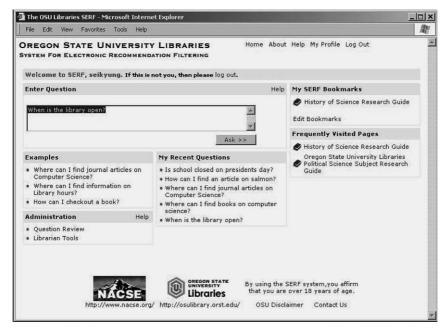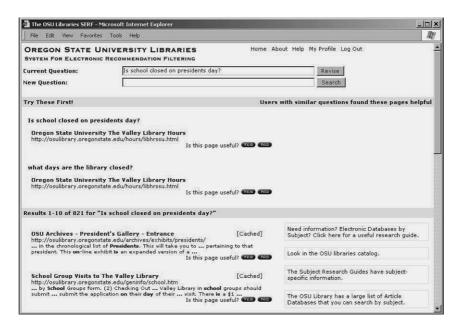
FIGURE 1
The initial search screen

FIGURE 2
The recommendations and search results screen

As users browse and read the information resources, they can rate each document's utility (Figure 3). Although users may not have viewed the document through SERF, this feature allows users who have existing knowledge of document contents to provide feedback without navigating to the document. Librarians requested this feature, because they could often tell immediately from the title and/or URL of a document if the document was relevant to their question. These ratings are used by the system to indicate if a document contributed to answering a specific information need, as defined by the query. Users can revise their queries or generate new queries once information has been found.

Initial Findings

The initial round of real time testing of SERF on the OSU Libraries web site took place between January and April 2004. The primary research question in this round was 'Can a user find the right information more efficiently and effectively when given automatically detected recommendations from past users in a library website?' During the experiment, 1,433 search transactions occurred. Each search transaction represents a single information context and the associated user behavior. Of the 1,433 search transactions, 239 were discarded because they
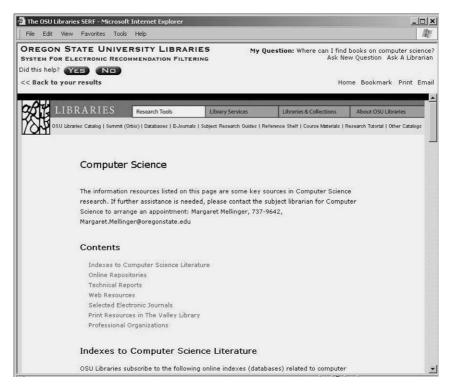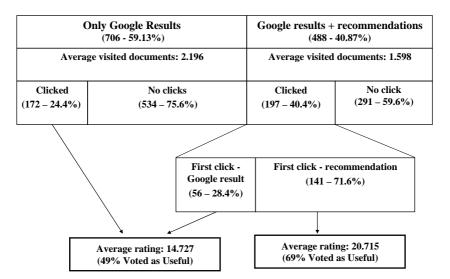

FIGURE 3
The document display with voting option screen

either represented usage by a member of the research group, or the data in the transaction was corrupt or inconsistent. These filtering steps left 1,194 search transactions.

Initial analysis of the transactions suggests that recommendations do help users find information more efficiently. A more complete analysis can be found in Jung *et al*. (2004a).

- Even with the immaturity of SERF, approximately 40 per cent of the transactions had at least one recommendation while the remaining 60 per cent of the transactions had only Google results. This suggests that the system learns quickly, yet also indicates that using the conventional search engine simultaneously is important in the early stages of the system.
- Users were more likely to click on a document from the recommendations than a document from the Google search results.
- Users were more efficient when there were recommendations (Figure 4). The number of documents visited when there are recommendations is 1.598 while the number visited when there are only Google results is 2.197. Also, users found the recommended resources more valuable; 69 per cent of the first visited documents from the recommendations versus 49 per cent from the Google results of 14.
- SERF was quite successful in getting users to articulate their information contexts. Approximately 68 per cent of the queries were natural language statements.

Only Google Results (706 - 59.13%)		Google results + recommendations (488 - 40.87%)	
Average visited documents: 2.196		Average visited documents: 1.598	
Clicked (172 – 24.4%)	**No clicks** (534 – 75.6%)	**Clicked** (197 – 40.4%)	**No click** (291 – 59.6%)

First click - Google result (56 – 28.4%)

First click - recommendation (141 – 71.6%)

Average rating: 14.727 (49% Voted as Useful)

Average rating: 20.715 (69% Voted as Useful)

FIGURE 4

Three months of SERF usage compares results with and without recommendations (vote as useful =30, vote as not useful =0)

- Approximately 30 per cent of queries submitted were directional queries (e.g. 'Where is the map room?), 30 per cent were about general information (e.g. How do I connect my laptop computer to the network?), and the remainder were about resources. This type of breakdown can help the OSU Libraries revise the web site to provide more obvious answers to directional and general questions through new links or better FAQs.
- Thirty per cent of the users of SERF rated at least one document after selecting at least one result. This is a high number given the lack of incentive to spend valuable time to give feedback.

Our results suggest that combining CF with link analysis searching has the potential to significantly improve the academic library search experience. Initial use of SERF indicates that library users are willing to participate in rating resources. We intend to examine what factors compel users to provide ratings as more and better ratings will allow the system to 'learn' faster and be more responsive. One approach is experimenting with system design so users perceive that improvement in their search performance is tied to rating more items. Allowing users to easily return to items with high ratings may help as well. Another factor may be tied to public recognition of those who contribute constructive ratings; recognition may encourage ratings.

SERF users responded positively to the interface and usually entered complete sentences, a requirement to build the best information contexts and enhance the results for future users. More research will investigate whether the longer queries actually improve search performance for the current user. This new approach appears to decrease the number of documents that users viewed to find useful information. SERF users were more likely to select a recommendation on the search result page. Moreover, documents recommended by SERF were rated higher than search results from Google.

The Future of Collaborative Filtering and Libraries

The initial testing of SERF is very encouraging and suggests that CF could transform library search systems. We face technical and philosophical challenges, and these motivate us to continue the collaboration between the OSU librarians and EECS. By working with computer scientists, librarians gain more technical expertise with search interface design, the importance of query structure, and how to do usability testing. By working with librarians, computer scientists have access to real world test beds complete with users and information resources. The collaboration also forces discussion of philosophical issues where there may be differences of opinion, or level of importance.

Two examples include the privacy of users, and the authority and evaluation of information. Maintaining users' privacy while developing personalized services and involving users in system development is difficult if not impossible. People appear to be more willing to collaborate and sacrifice privacy

for more efficient access to information, yet still have concerns (Kobsa and Teltzrow 2005). Discussion of the responsibility of librarians vis-a-vis patron privacy is shifting from 'how to protect users privacy' to 'how to inform users about privacy concerns' (Balas 2001; Pace 2001; Shuler 2004). Working with system developers provides an excellent opportunity for librarians to integrate information about privacy into library personalization systems. Collaboration reinforces the importance of educating system developers and users on the benefits and costs of personalization. Involving users in evaluating and recommending resources changes the role of the library from guarantor of the authority of information to one of mediator. We can learn to trust our users in new ways, and hone our skills in teaching them how to evaluate information.

Librarians and computer scientists need to find better ways to provide search and discovery technologies for our valuable library collections. As our digital collections, resources and services increase, so should our ability to provide a coherent and efficient way to search across them. A successful search system may incorporate different features and employ a variety of approaches; however, it should be easy to use, fast, reliable, and able to 'learn'. We believe that incorporating users' expertise and opinions can help our users find and use the wealth of information in our institutions. Our project is one example of collaboration shaping the direction of library search systems by coupling the expertise of librarians with that of computer scientists to better address how our users expect to find information.

Acknowledgements

Notes

1. http://osulibrary.oregonstate.edu/research.php/ejournal.php?arg $=$a
2. http://osulibrary.oregonstate.edu/research.php/db.php?arg $=$a
3. http://osulibrary.oregonstate.edu/special collections/rnb/
4. http://sp.teoma.com/docs/teoma/about/searchwithauthority.html
5. http://www.searchengines.com/directhit.html
6. http://www.metacrawler.com/info.metac/search/help/aboutmetasearch.htm

7. OCLC, Online Computer Library Center. http://www.oclc.org/
8. http://aplab.rutgers.edu/ant/

References

ABRAM, S. and LUTHER, J. 2004. Born with a chip. *Library Journal*, 129(8): 34−7.

ANICK, P. 203. Using terminological feedback for web search refinement: A log-based study. In *Proceedings of the 26th annual international ACM SIGIR on research and development in information retrieval*, edited by C. Clarke, G. Cormack, J. Callan, D. Hawking and A. Smeaton, New York, NY: ACM Press, 88−95.

BALAS, J.L. 2001. How should privacy be protected in the electronic library? *Computers in Libraries*, 21(6): 53−5.

BELKIN, N.J., COOL, C., KELLY, D., KIM, G., KIM, J-Y., LEE, H-J., MURESAN, G., TANG, M-C. and YUAN, X-J. 2003. Query length in interactive information retrieval. In *Proceedings of the 26th annual international ACM SIGIR on research and development in information retrieval*, edited by C. Clarke, G. Cormack, J. Callan, D. Hawking and A. Smeaton, New York, NY: ACM Press, 205−12.

BURKE, R., HAMMOND, K., KULYUKIN, V., LYTINEN, S., TOMURO, N. and SCHOENBERG, S. 1997. Natural language processing in the FAQ Finder System: results and prospects. In Working Notes, *AAAI spring symposium on NLP on the WWW*, Menlo Park, CA: AAAI Press, 17−26.

COSLEY, D. LAWRENCE, S. and PENNOCK, D.M. 2002. REFEREE: An open framework for practical testing of recommender systems using ResearchIndex. In *Proceedings of the 28th international conference on very large databases*, San Francisco, CA: Morgan Kaufman, 35−46.

FRIEDLANDER, A. 2002. *Dimension and use of the scholarly information environment: introduction to a data set assembled by the Digital Library Federation and Outsell, Inc*. Washington, DC: Digital Library Federation, Council on Library and Information Resources.

GLADWELL, M. 1999. Annals of marketing: The silence of the sleeper: how the information age could blow away the blockbuster. *New Yorker*, 75(29): 48−55.

GOLDBERG, K., ROEDER, T., GUPTRA, D. and PERKINS, C. 2001. Eigentaste: A constant-time collaborative filtering algorithm. *Information Retrieval*, 4(2): 133−51.

HILL, W., STEAD, L., ROSENSTEIN, M. and FURNAS, G. 1995. Recommending and evaluating choices in a virtual community of use. In *Proceedings of the SIGCHI conference on human factors in computing systems*, edited by I.R. Katz, R. Mack, L., Marks, M.B. Rosson and J. Nielsen, New York, NY: ACM Press, 194−201.

JUNG, S., HARRIS, K., WEBSTER, J. and HERLOCKER, J.L. 2004a. SERF: integrating human recommendations with search. In *Proceedings of the 13th ACM conference on information and knowledge management*, New York, NY: ACM Press, 571−80.

JUNG, S., KIM, J. and HERLOCKER, J. 2004b. Applying collaborative filtering for efficient document search. In *The Proceedings of the IEEE/WIC/ACM international conference on web intelligence*, edited by N. Zhong, H. Tirri, Y. Yao, L. Zhou, J. Liu and N. Cercone, Los Alamitos, CA: IEEE Computer Society, 640−43.

KOBSA, A. and TELTZROW, M. 2005. Impacts of contextualized communication of privacy practices and personalization benefits on purchase behavior and perceived quality of recommendation. In *Beyond Personalization 2005, International Conference on Intelligent User Interfaces*, http://www.grouplens.org/beyond2005/full/kobsa.pdf

LOSSAU, N. 2004. Search engine technology and digital libraries: libraries need to discover the academic internet. *D-Lib Magazine*, 10(6), DOI:10.1045/june2004-lossau.

MCNEE, S.M., ALBERT, I, COSLEY, D., GOPALKRISHNAN, P., LAM, S.K., RASHID, A.M., KONSTAN, J.A. and RIEDL, J. 2002. On the recommending of citations for research papers. In *Proceedings of the 2002 ACM conference on computer supported cooperative work*, edited by E.F. Churchill, J. McCarthy, C. Neuwirth and T. Rodden, New York, NY: ACM Press, 116−25.

MENKOV, V., NEU, D.J. and SHI, Q. 2000. AntWorld: a collaborative web search tool. In *Proceedings of the 2000 workshop on distributed communications on the web*, edited by P. Kropf, G. Babin, J. Plaice and H. Unger, Berlin: Springer-Verlag, 13−22.

OCLC Online Computer Library Center, Inc. 2002. OCLC white paper on the information habits of college students: How academic librarians can influence students' web-based information choices, *http://www5.oclc.org/downloads/community/informationhabits.pdf*.

PACE, A.K. 2001. It's a matter of privacy. *Computers in Libraries*, 21(6): 50−2.

PAGE, L., BRIN, S., MONTWANI, R. and WINOGAD, T. 1998. *The PageRank citation ranking: Bringing order to the web*. Technical report, Stanford University Database Group.

RECKER, M.M. and WALKER, A. 2003. Supporting 'word-of-mouth' social networks through collaborative information filtering. *Journal of Interactive Learning Research*, 14(1): 79−98.

SHARDANAND, U. and MAES, P. 1995. Social information filtering: algorithms for automating 'word of mouth'. In *Proceedings of the SIGCHI conference on human factors in computing systems*, edited by I.R. Katz, R. Mack, L., Marks, M.B. Rosson and J. Nielsen, New York, NY: ACM Press, 210−17.

SHULER, J. 2004. Privacy and academic libraries: widening the frame of discussion. *The Journal of Academic Librarianship*, 30(2): 157−9.

VAN DE SOMPEL, H. and HOCHSTENBACH, P. 1999. Reference linking in a hybrid library environment: part 1: Frameworks for linking. *D-Lib Magazine*, 5(4), DOI: 10.1045/april99-van_de_sompel-pt1.

Janet Webster, Guin Library, Oregon State University, 2030 SE Marine Science Drive, Newport, OR 97365, USA. E-mail: Janet.webster@oregonstate.edu

Seikyung Jung, Electrical Engineering and Computer Science Department, Oregon State University, Corvallis, OR 97330, USA. E-mail: jung@cs.orst.edu

Jon Herlocker, Electrical Engineering and Computer Science Department, Oregon State University, Corvallis, OR 97330, USA. E-mail: herlock@cs.orst.edu