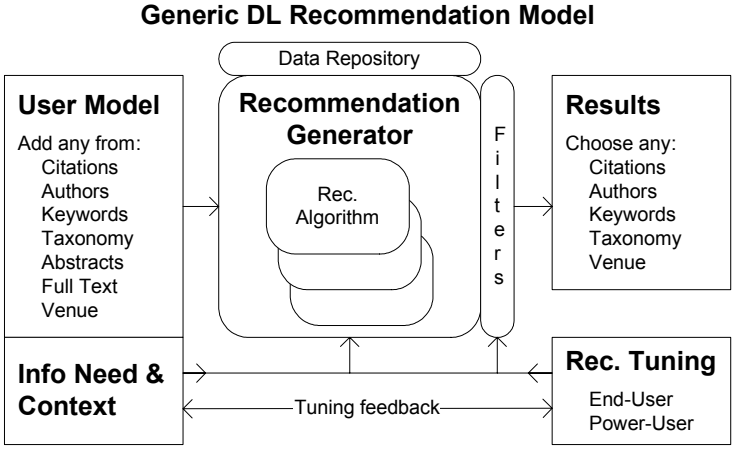
TechLens: Exploring the Use of Recommenders to Support Users of Digital Libraries

Joseph A. Konstan[*], Nishikant Kapoor[*], Sean M. McNee[*], and John T. Butler[†]
[*]Department of Computer Science and Engineering; [†]University Libraries
University of Minnesota
{konstan,nkapoor,mcnee}@cs.umn.edu; j-butl@umn.edu

INTRODUCTION AND OVERVIEW

The immense collection of valuable information in digital libraries is changing the way students and scholars access information. Indeed, just as library users are accessing libraries and research librarians remotely, the potential is increasing for value-added services that promise to help patrons use digital libraries in ever more powerful ways, moving beyond basic search to a new collection of awareness, field summary, and people-finding services.



For the past four years we've been exploring means by which recommender systems technology—the technology used today by e-commerce vendors to help customers find products—can be adapted to serve the needs of students and scholars exploring scientific literature. The model shown here illustrates the types of data that can be used to fulfill a user's need. We have already demonstrated the success of some of the basic approaches—using citations, keywords, and abstracts to find works a researcher is unfamiliar with. Much work remains, however, as we explore ways to meet a more diverse set of information needs.

A PRIMER ON RECOMMENDER SYSTEMS

Recommender systems have evolved since the early 1990's as a response to increasing information overload. These systems help users identify a subset of items within a large information space. Unlike ordinary keyword search systems, recommenders attempt to find items that match user tastes and the user's sense of quality, as well as syntactic matches on topic or keyword. For example, a music recommender will use an individual's prior taste in music to identify additional songs or albums that may be of interest.

One of the most successful technologies for recommender systems, co-invented at the University of Minnesota, is *collaborative filtering*. Collaborative filtering systems use the opinions of a community to recommend items to individuals. In the music example, a collaborative filtering recommender would identify other people who share your music tastes, and would then recommend to you music that those "neighbors" liked but that you hadn't yet heard.

To generate recommendations for research papers, we have applied collaborative filtering to the graph of citations between papers. Thus, each research paper 'votes' for its citations. By using this approach, our recommenders capture the opinions of paper authors on how all papers are related to each other--something that content-based searching alone can miss. We have also found powerful recommendations techniques from combining both collaborative- and content-based filtering together.

INITIAL RESULTS

The following papers demonstrate that recommender systems technology can be used effectively to recommend research papers, specifically by analyzing relationships between citations. We also find that different recommender algorithms return recommendations with different biases—more authoritative work, or more novel work, for example.

S.M. McNee et al. "On the Recommending of Citations for Research Papers". *Proceedings of ACM 2002 Conference on Computer Supported Cooperative Work* (CSCW 2002), November 2002, pp. 116-125.

R. Torres et al. "Enhancing Digital Libraries with TechLens+". *Proceedings of The Fourth ACM/IEEE Joint Conference on Digital Libraries* (JCDL 2004), June 2004, pp. 228-237.

Other related work of particular interest includes.

C-N Ziegler et al. "Improving Recommendation Lists Through Topic Diversification". *Proceedings of the Fourteenth International World Wide Web Conference* (WWW 2005), May 2005, pp. 22-32.

J. Herlocker et al. "An Algorithmic Framework for Performing Collaborative Filtering". *Proceedings of the 1999 ACM Conference on Research and Development in Information Retrieval*. (SIGIR 1999) Aug. 1999, pp. 230.237.

Two TechLens Demonstrations

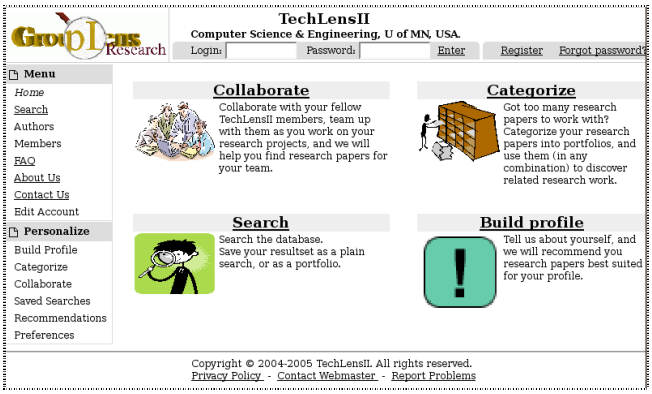
Recommending Additional Citations for a Paper

Our initial experiments were carried out in 2001/2002 using the CiteSeer citation database. We've since updated this demonstration to use the ACM Digital Library as its core database. The demonstration is designed to illustrate the power of recommendation and to allow users to explore the diversity of algorithms available for recommending papers. Below, we have a list of recommended papers for Jon Herlocker's 1999 paper, *An Algorithmic Framework for Performing Collaborative Filtering.*

Three Recommendations for Herlocker's 1999 paper. The TechLens-I Demo is at: http://frost.cs.umn.edu/techlens/

Bringing Recommendations to a Digital Library

Our newest demonstration is an attempt to envision a researcher's desktop built into a digital library. Rather than duplicate the features of existing DL interfaces, we are experimenting with specific interfaces (such as tools to gather personal bibliographic data and lists of collaborators) that can support recommendations for individuals and teams.



The TechLens-II Demo Interface. TechLens-II is available at http://frost.cs.umn.edu/techlens2/

FUTURE DIRECTIONS AND CHALLENGES

Near-Term Directions

Our current efforts are focused on two key short-term goals. First, we are trying to understand the recommendation algorithms needed to support a variety of different user-level tasks and user types. Among these are: finding additional citations for a draft paper or proposal, recommending papers to maintain awareness in a field, recommending reviewers suitable for a submitted paper, and supporting a librarian in the task of collection management. Second, we are exploring interfaces that will be effective in supporting users to fulfill these and other goals.

In addition, we recognize significant privacy concerns, and are engaged in exploring both user attitudes and technological solutions that can help address them. Recommender systems in general depend on large collections of opinion data, and few researchers are likely to contribute assessments of their colleagues' work to a repository. While our current systems use explicitly public data from published papers, we can envision more powerful services that may only be possible if people voluntarily contribute information to a recommender system.

Longer-Term Directions

Our key longer-term goal is the design of a system that enables trained and skilled users to assemble the tools to meet unanticipated information needs. Such a system would build upon the model at the top of this handout to allow a skilled searcher, or more likely a searcher aided by a skilled reference librarian, to assemble the sets of algorithms and filters needed to address specific questions. We can imagine, for example, a program chair for a large conference working with a librarian to use a digital library filled with publication and citation data to make preliminary assignments of papers to a pool of reviewers. In such a case, the recommendations would be used to match people with papers that have strong topical overlap with their interests (e.g., similar keywords or common citations). At the same time, the filters could remove people with well-identified conflicts of interest (e.g., recent collaboration).

Similar tools can be adapted to help identify the most influential papers in a new field, or to identify new fields with significant topical coherence but whose papers are diffused in the conferences and journals of other fields (perhaps as a tool for book or journal publishers).

In the long-term, we also recognize the importance of the debate over ownership rights for content, including the extent of indexing and abstract-processing rights, as a key factor inhibiting or enabling the type of digital library services we envision.

ACKNOWLEDGEMENTS

This work has been supported in part by the National Science Foundation under grants 95-54517, 97-34442, and 01-02229, as well as by a research assistantship from the University of Minnesota Libraries. We'd also like to thank our colleagues in the GroupLens Research and in the University Libraries for their insights and assistance, and our industrial partners, RefWorks and the ACM for their cooperation.