PHOAKS:

A System for Sharing Recommendations

Loren Terveen, Will Hill, Brian Amento, David McDonald, and Josh Creter

A collaborative filtering system that recognizes and reuses recommendations.

FINDING RELEVANT, HIGH-QUALITY INFORMATION ON THE WORLD-WIDE Web is a difficult problem. PHOAKS (People Helping One Another Know Stuff) is an experimental system that addresses this problem through a collaborative filtering approach. PHOAKS works by automatically recognizing, tallying, and redistributing recommendations of Web resources mined from Usenet news messages.

The feasibility of automatic recognition of recommendations is supported by empirical results. First, Usenet messages are a significant source of recommendations of Web resources: 23% of Usenet messages mention Web resources, and 30% of these mentions are recommendations. Second, recommendation instances can be machine-recognized with nearly 90% accuracy. Third, some resources are recommended by more than one person. These multi-confirmed recommendations appear to be significant resources for the relevant community. Finally, the number of distinct recommenders of a resource is a plausible measure of resource quality. A comparison of recommended resources with resources in FAQs (lists of Frequently Asked Questions maintained by human topic experts) indicates the more distinct recommenders a resource has, the more likely it is to appear in the FAQs.

PHOAKS is distinguished from other recommender systems by two major design principles: *role specialization* and *reuse*. Many recommender systems, particularly ratings-based systems [1, 3, 4], are built on the assumption of role uniformity. They expect all users to do the same types of work in return for

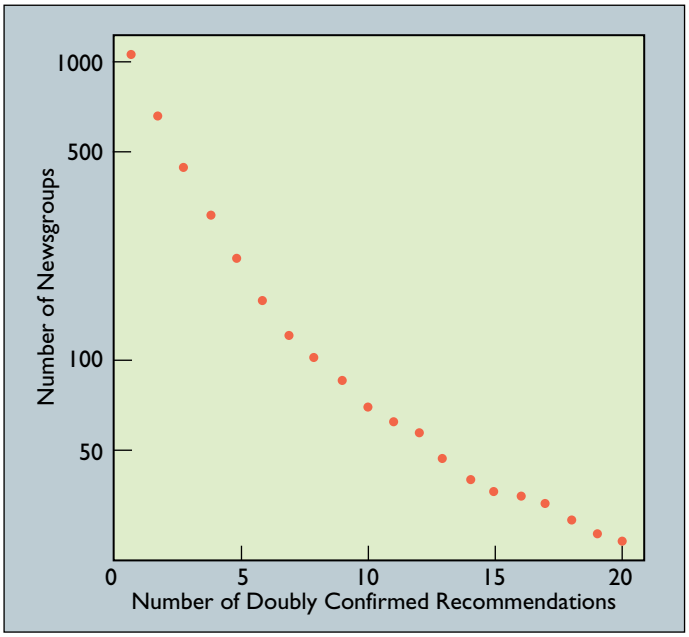

Figure 1. The distribution of doubly confirmed recommendations

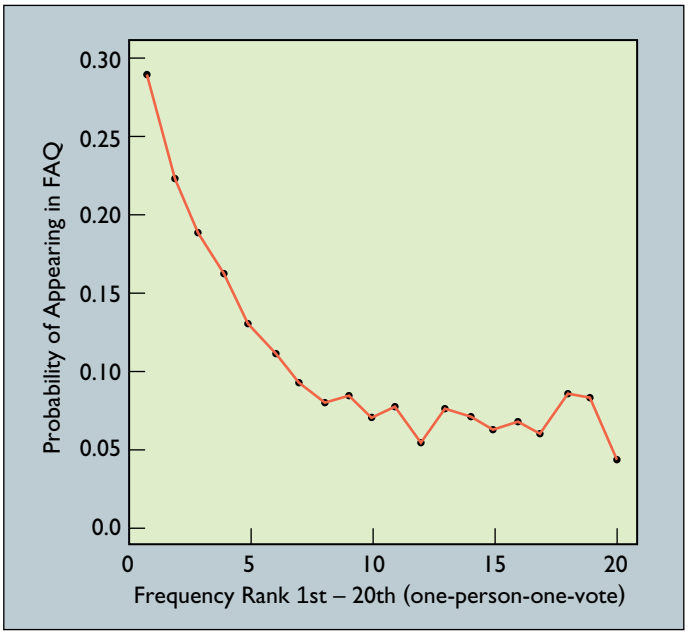

Figure 2. Comparing recommended resources to FAQ resources

the same types of benefits. In the case of ratings-based systems, for example, everyone rates objects of interest. Yet there is evidence that people naturally prefer to play distinct producer/consumer roles in the information ecology [2]; in particular, only a minority of people expend the effort of judging information and volunteering their opinions to others. Independently, we have observed such role specialization in Netnews; authors volunteer long lists of recommended Web resources at a stable, but low, rate. PHOAKS assumes the roles of recommendation provider and recommendation recipient are specialized and different. PHOAKS reuses recommendations from existing online conversations. This reuse requires no extra work from providers and no judgments of information quality from PHOAKS users, another difference with ratings-based systems.

The PHOAKS system contains six months or more of recommendations and associated data for about 1,500 newsgroups. Thousands of new opinions about Web resources are added weekly.[1]

What Counts as a Recommendation?

The basic idea of collaborative filtering is people recommending items to one another. Readers of Usenet news know this is a normal practice in newsgroups. Posters often volunteer their impressions and opinions about all sorts of items, including Web pages. They may state what a page is useful for and how useful it is. PHOAKS searches messages for mentions of Web pages (URLs) and counts a mention as a recommendation if it passes a number of tests. First, the message must not be cross-posted to too many newsgroups. Messages posted to a large number of groups are so general they are not likely to be thematically close to any of the groups. Second, if the URL is part of a poster's signature or signature file, it is not counted as a recommendation. Third, if the URL occurs in a quoted section of a previous message, it is ruled out. Fourth, if the textual context surrounding the URL contains word markers that indicate it is being recommended and does not contain makers that indicate it is being advertised or announced, then it is categorized as a recommendation. We have developed rather complicated categorization rules that implement this basic strategy to distinguish the different purposes for which Web resources are mentioned.

In a representative sample of 1.3 million messages processed between February and August of 1996, 23% of the messages mention Web resources, with computer- and science-related groups having a slightly higher percentage and recreational groups a slightly lower percentage.

AFTER MUCH ANALYSIS, TESTING, AND ITERATION of our categorization rules, we have developed a fairly accurate rule set. There are two aspects of accuracy: *precision* (the percentage of resources the rules classify into a certain category that actually

[1] PHOAKS is available at http://www.phoaks.com/phoaks/. As of December 1996, more than 3,000 visitors access recommendations each day.

PHOAKS consists of a general architecture for filtering information from electronic messages and a set of techniques for generating and managing dynamic Web-based interfaces.

belong to the category) and *recall* (the percentage of resources that belong to a category that the rules actually classify into that category). A validation study of more than 600 URL mentions shows that our rules for recognizing recommendations have 88% precision and 87% recall.

HOW SHOULD WE RANK RECOMMENDED resources within a newsgroup? In other words, how can we automatically compute an approximate measure of resource quality? We selected the number of distinct recommenders of a resource as a measure. This metric values independent opinions in estimating the worth of a resource. We have done an analysis that focuses on resources with at least three recommenders—"doubly confirmed" recommendations. Figure 1 shows, for a set of 1,042 newsgroups with at least 20 recommended resources, the number of newsgroups that have from 1 to 20 doubly confirmed recommendations. For example, 429 newsgroups had at least three doubly confirmed recommendations, 217 had at least five, and 68 had at least 10.

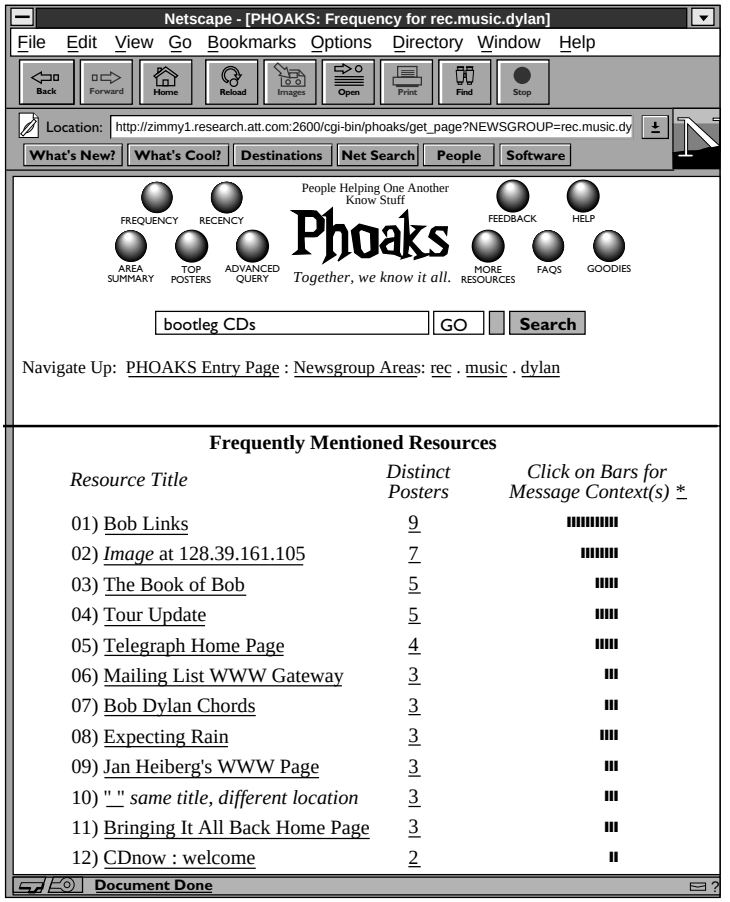

Figure 3. PHOAKS resource summary page

How can we tell whether the number-of-recommenders metric is a good one? To try to answer this question, we analyzed the intersection between resources recommended on Usenet (that were not in FAQ messages) and resources in newsgroup FAQs. We obtained FAQs by tailoring the basic PHOAKS message-filtering architecture to identify Usenet messages that posted FAQs. Since FAQs contain the kind of information a human topic expert thinks is of appropriate quality and relevance, we consider FAQs an appropriate baseline for judging the quality of resources our rules classify as recommendations.

Figure 2 shows the intersection between recommended resources and FAQ resources for 313 newsgroups (groups for which we had both FAQs and recommendation data). The X axis shows resources as ranked from 1 to 20 by the number-of-recommenders measure, and the Y axis shows the percentage of resources from each rank present in the related FAQ(s). For example, 29% of the 313 top-ranked resources, 22% of the 313 second-ranked resources, and 19% of the 313 third-ranked resources occur in the relevant FAQ(s). The graph shows the more distinct recommenders a resource has, the more likely it is to appear in the FAQ. Thus, the number-of-recommenders measure appears to be consistent with human judgments of quality.

The PHOAKS System

PHOAKs consists of a general architecture for filtering information from electronic messages and a set of techniques for generating and managing dynamic Web-based interfaces. So far, we have used the PHOAKS architecture to recognize and process Web resource recommendations and FAQ messages. The architecture consists of three main processes:

- *Search*—search messages for a specified pattern (such as "http://") and extract contextual information surrounding each instance of the pattern
- *Categorization*—apply rules that classify each instance of the pattern (e.g., URLs used as recommendations vs. personal home pages)
- *Disposition*—process the categorized information in some way (e.g., store it in a database or fetch the content of a URL)

We have a continuously updated dynamic Web interface to our database of Web-resource recommendations. Theses interfaces presents information about recommended resources, recommenders, and recommendation context for about 1,500 newsgroups. The system has been publicly available since early February 1996, and, as of the end of December 1996 had been accessed by more than 300,000 visitors. Figure 3 shows a page that summarizes recommended resources for the newsgroup rec.music.dylan.

Future Work

HERE ARE A FEW OF THE INTERESTING DIRECtions we are pursuing. First, we are continuing to analyze the relationship between resources recommended in Usenet messages and resources that appear in FAQs. We are especially interested in the temporal dimension. So, for example, we will determine to what extent Usenet messages are a leading indicator of FAQ content. Second, we will use FAQs to enhance our interface to recommendation data. For example, one will be able to go from a resource to references to that resource in FAQs. Thus, we intend to combine the best of automatically mined recommendations (e.g., timeliness) with human-determined recommendations (e.g., long-term relevance and quality). Third, we will apply our generic filtering architecture to extract other types of information from electronic messages. We are particularly interested in intranet applications and education applications. Fourth, we are exploring the issues of how to compute credibility of recommenders and affinity between those who offer and seek recommendations for a particular topic. This is a much more difficult problem than it is for ratings-based system, but solving it would help us increase the quality and relevance of the recommendations we offer. Finally, we are exploring a combination of keyword search and collaborative filtering we call "community-sorted search." The basic idea is to run a query using a normal keyword search engine, then filter the results through the databases maintained by PHOAKS. Results are clustered by the newsgroups that mention them (thus, disambiguating the meaning of the query), then ordered by the frequency of mention. C

Acknowledgments

Paul Resnick, Lynn Cherny, Jon Helfman, Julia Hirschberg, Erik Ostrom, Steve Crandall, Don Hindle, Fernando Peirera, and Steve Whittaker made many valuable comments on this work. System support from Dave Korman and Shankar Ishwar helped to get the PHOAKS Web site going. Finally, we wish to thank all PHOAKS users who continue to give us helpful feedback.

References

1. Hill, W.C., Stead, L., Rosenstein, M., and Furnas, G. Recommending and evaluating choices in a virtual community of use. In *Proceedings of the ACM Conference on Human Factors in Computing Systems, CHI'95* (Denver, May 7–11). ACM Press, New York, N.Y., pp. 194–201.
2. Maltz, D., and Ehrlich, K. Pointing the way: Active collaborative filtering. In *Proceedings of the ACM Conference on Human Factors in Computing Systems, CHI'95* (Denver, May 7–11). ACM Press, New York., N.Y., pp. 202–209.
3. Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., and Riedl, J. GroupLens: An open architecture for collaborative filtering of netnews. In *Proceeding of the ACM 1994 Conference on Computer Supported Cooperative Work*. pp. 175–186.
4. Shardanand, U., and Maes, P. Social information filtering: Algorithms for automating "word of mouth." In *Proceedings of the ACM Conference on Human Factors in Computing Systems, CHI'95* (Denver, May 7–11). ACM Press, New York. N.Y., pp. 210–217.

Loren Terveen (terveen@research.att.com) is senior member of the technical staff, Human-Computer Interface Research, at AT&T Labs—Research, in Florham Park, NJ.
Will Hill (willhill@research.att.com) is a research scientist, Human-Computer Interface Research, at AT&T Labs—Research, in Florham Park, NJ.
Brian Amento (brian@cs.vt.edu) is a graduate student researcher at Virginia Tech's computer science department.
David McDonald (dmcdonal@ics.uci.edu) is a graduate student researcher, information and computer science, at the University of California, Irvine.
Josh Creter (jcreter@pomona.edu) is a student at Pomona College, Claremont, Calif.