

Using Frequency-of-mention in Public Conversations for Social Filtering

Will Hill and Loren Terveen

AT&T Research, 600 Mountain Avenue, Murray Hill, NJ 07974-0636
willhill@research.att.com, terveen@research.att.com
908-582-6432, 2608

ABSTRACT

We report on an investigation of using Usenet newsgroups for social filtering of Web resources. Our main empirical results are: (1) for the period of May '96 to Jul '96, about 23% of Usenet news messages mention Web resources, (2) 19% of resource mentions are recommendations (as opposed, e.g., to home pages), (3) we can automatically recognize recommendations with at least 90% accuracy, and (4) in some newsgroups, certain resources are mentioned significantly more frequently than others and thus appear to play a central role for that community. We have created a Web site that summarizes the most frequently and recently mentioned Web resources for 1400 newsgroups.

Keywords: Human-computer interaction, human interface, computer-supported cooperative work, organizational computing, social filtering, collaborative filtering, browsing, resource discovery, World Wide Web, Usenet, netnews.

INTRODUCTION

One of today's urgent information access problems is finding high quality information on the World Wide Web. In this paper, we report on an investigation into the usefulness of Usenet newsgroups and other public conversations for social filtering of Web resources. Specifically, we inquired whether recommendations of Web resources could be extracted from netnews messages. We found that hundreds of newsgroups contain enough recommendations to be potentially useful.

Our interest in using public conversation archives for social filtering stems from our bias toward systems that require little or no extra data entry from their users. If public conversation archives do in fact contain data useful for filtering web resources, then we won't have to resort to the common social filtering practice of asking users to perform the additional task of rating or otherwise evaluating information in order to get useful recommendations. Of course, whether or not public conversation archives contain useful social filtering data is not the only issue significant for system building. We also must

Permission to make digital/hard copies of all or part of this material for personal or classroom use is granted without fee provided that the copies are not made or distributed for profit or commercial advantage, the copyright notice, the title of the publication and its date appear, and notice is given that copyright is by permission of the ACM, Inc. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires specific permission and/or fee.

Computer Supported Cooperative Work '96, Cambridge MA USA
© 1996 ACM 0-89791-765-0/96/11 ..\$3.50

design computations to extract the data effectively and efficiently, create useful presentations of the data, and strive to ensure that the original conversational participants are happy with the reuse of their conversational records. This paper addresses the first two questions in depth and then touches briefly on the third and fourth questions. We will address the third and fourth questions fully in a subsequent paper.

Here in summary form are our basic questions and answers concerning the technical feasibility of using Usenet messages for social filtering of Web resources.

- *Do posters of Usenet news mention urls (universal resource locators) in the messages they post?*

Yes, roughly 23% of the messages for the time period covered. We looked at 1.3M messages from 1438 active newsgroups over a period of at least 30 days and found that 305K of them contained urls in the message body.

- *For what purposes do posters mention these urls?*

We observed three frequent and many other less frequent purposes. Our machine classification algorithm classified 19% of the url mentions as recommendations or announcements of web resources, 46% as sharing contact information in a "signature", and 8% of the url mentions as part of quotes from previous messages. The remaining 26% were classified as other types including use as examples, descriptions, and citations.

- *Can url mentions that recommend or announce web resources be machine-recognized?*

Yes, with 90% precision.

- *On a "one person, one vote" basis, are some urls mentioned more frequently than others?*

By "one person, one vote" basis we mean that we tally only one vote per url per person even if a person mentions the url more than once. The idea is to defeat individual attempts to flood a conversation with pointers to a single web resource.

The answer to the question is yes, for almost all newsgroups. However, during a second data collection study covering 1344 active groups, only 484 groups showed significant variation. Our working hypothesis is that when participants in a group conversation do mention

some resources significantly more often than others, these are more central to the group's conversation, and thus are important resources for the group topic.

These data suggest the possibility of creating a useful system for social filtering in the large - for as many topics as there are active newsgroups. In our most recent study, 2 months of automatic processing of 1438 active newsgroups has yielded 93K url recommendations. Experimenting with appropriate interfaces to the large number of recommendations that pass through Usenet is a promising area. For newsgroup topics that range from downhill skiing to the X-Files to Java, from Japanese culture to beer brewing to emacs, we can provide ways for users to access the resources recommended in the relevant newsgroup.

The remainder of the paper examines our basic questions and their answers in more detail and gives an overview of a prototype system.

RELATED RESEARCH

Malone et al. [10] propose three types of information filtering activities: cognitive, economic and social. This paper concentrates on computer-mediated social filtering. The prototypical case of social filtering is people communicating judgements of quality to other people that they know personally. However, a basic thesis of this work and one developed elsewhere [7] is that personal relationships are not necessary to social filtering. In fact, social filtering and personal relationships can be teased apart and put back together in interesting new ways. For instance, the communication of quality judgments can occur through less personal, or even impersonal relationships, with Usenet news being an example. Obviously, people want a satisfying mix of both personal and impersonal relationships. Various kinds of social filtering systems have been proposed and implemented, each illustrating a different mix of the personal and impersonal, of privacy and connectedness, of cost and benefit.

Goldberg's Tapestry system [5] is an email system that encourages the entry of free text annotations with which users can later filter messages. The two tasks of writing annotations (entering filtering data) and specifying queries (using filtering data) require significant user effort but may offer great reward. The classic problem [6] is that the benefit is either in the future or comes to different people than those who put in the work. Therefore there are few domains where such invested efforts pay off readily to the right participants.

Aiming at the support of work groups, Maltz and Ehrlich [11] describe collaborative filtering systems that enable users to explicitly create recommendations (via pointers and informational digests) and direct their recommendations to specific colleagues. These systems are informed by observed work group practice, including the significant finding that only a minority of participants adopt the role of information recommender within a group.

The thread of work on history-enriched digital objects (HEDO) [8][9][17] attempts to explore a region of the interface design space that minimizes additional user tasks. Usage

information is automatically recorded and is graphically presented on the relevant interface objects; thus, no additional input is required. HEDO techniques record the statistics of menu-selections, the count of spreadsheet cell recalculations and time spent reading documents (e.g., email, reports, source-code) in a line-by-line manner, summing over sections and whole documents. Relying upon induced equivalence classes of web resources, Wittenburg et al. [16] describes an automatic social-filtering method for recommending web sites within a work group by cross-indexing bookmark files and hot lists.

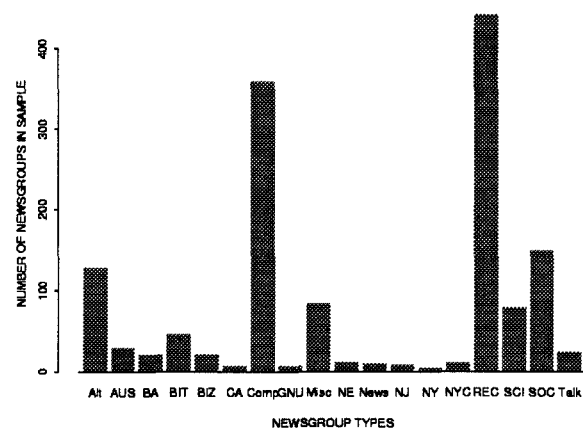
Various researchers have sought a middle ground between user annotations and interaction monitoring, in particular by exploring the use of ratings. Allen [1] reported unencouraging results on one of the first investigations into personal ratings for HCI-type user-modeling. More recently, Resnick et al. [13] have designed and implemented a social filtering architecture based upon personal ratings and demonstrated its application to filtering of netnews. A field test of the system involving 200 participants so far is under way. Hill et al [7] and Shardanand and Maes [15] describe other successful ratings based systems for recommending movies and music. In a study of eight users reading 8K netnews messages, Morita and Shinoda [12] observed strong positive correlations between time spent reading messages and personal interest ratings of those messages. Their work suggests it might be possible for time-on-task measures to stand in for ratings.

DATA ANALYSIS

The Data Set

Our data set consists of messages from 1438 Usenet news groups that received on average at least 3 messages per day and for which we observed at least 30 consecutive days of messages.

Figure 1 Observed Active Newsgroups categorized by their place in the netnews hierarchy.

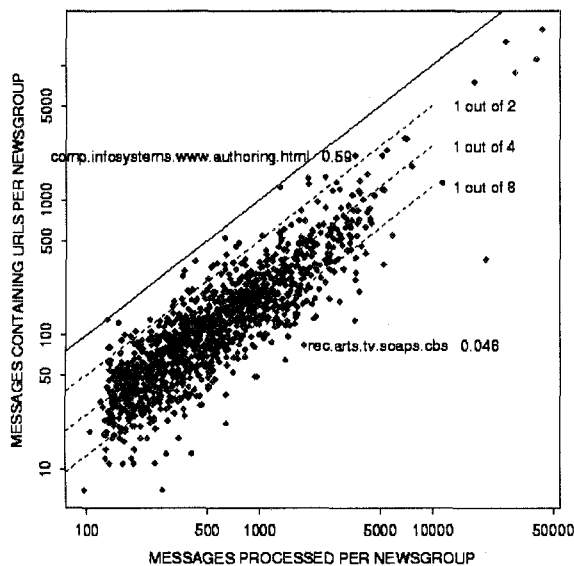


Because they were not a sources of textual recommendations, we excluded "*"binaries" groups. Figure 1 shows the proportion of active groups by Usenet top level category. Most of the groups we observed were in "comp" and "rec".

Url Mentions: How Often?

Of the 1.3M messages we observed in the 1438 active newsgroups, 305K of them contained url mentions (e.g., as indicated by the string “http://”) in the message body - 23.04%. Figure 2 plots (using a log-log scale) the number of messages seen vs. number of messages containing url references for each of the 1438 groups. Each dot represents a single newsgroup. The long solid diagonal represents the 1:1 ratio where all messages in a group would contain the mention of an url. Note the bent galaxy shape of the data. Dotted lines mark the 1:2, 1:4, and 1:8 ratios of messages containing “http://” to total messages.

Figure 2 Scattergram of number of messages processed per newsgroup and number of messages contain “http://” per newsgroup.



We have labeled two of the outliers. On the high percentage side we see “comp.infosystems.www.authoring.html” with 59% of its messages containing urls, on the low side we see “rec.arts.tv.soaps.cbs” with 4.6% of its messages containing urls. The oval shape of the cluster illustrates that the correlation of total number of messages with number of messages containing urls is strongly positive, $r = 0.91$. As mentioned, on average 23.04% of netnews messages contain url mentions, but there is wide variation from near 0% to near 100%.

The 305K messages that mentioned urls account for 478K url mentions. So, on average, messages that mention urls contain more than one mention (1.57). The mode, however is one. We observed about 1% of messages to be “public service” recommendation lists which included multiple urls per message.

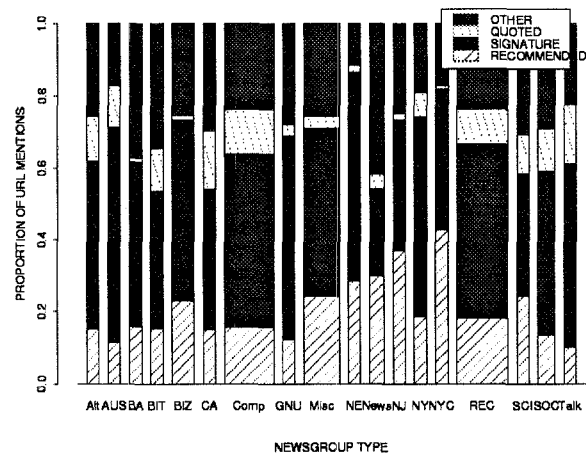
Url Mentions: Why

For social filtering purposes, we need mentions of urls that were recommendations or announcements. A cursory examination of url mentions reveals that a majority of them are not recommendations at all; rather, they are personal or corporate

home pages, usually in a machine-inserted signature at the end of the message.

We determined 29 syntactic features of messages to use in to classifying url mentions as part of a quote, a signature, a recommendation, or something else. We then hand-categorized 1000 random url mentions. We input the 29 features and categorizations to William Cohen’s rule-learning algorithm RIPPER [3], which outputs classification hypotheses. We modified the conjectured rules to remove clauses that seemed due to sampling error, then began using the rules to classify all the url mentions coming down the netnews pipe. (We should note that we tried to make the categorization rules conservative, i.e., to seek precision at the price of recall. Since there are so many references to resources, we can afford to miss a few recommendations. And since the majority of references are not recommendations, we must be careful about what we let in.)

Figure 3 Purposes for mention web resource urls categorized by Newsgroup hierarchy type.

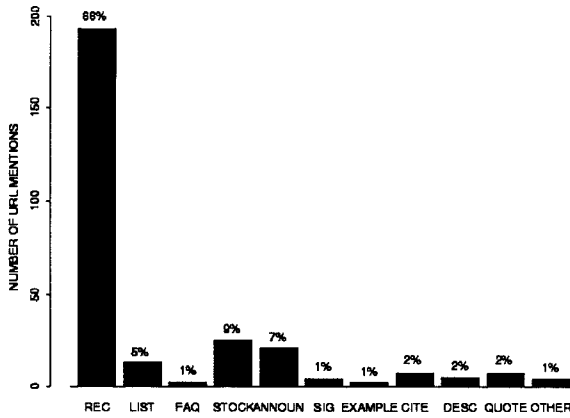


We obtained the following results. As shown in figure 3 (from top to bottom), 26% of url mentions were classified as “other”, 8% were mentioned as part of a quoted message, 46% were part of the poster’s signature, and 19% were recommendations or announcements. The equal heights of the bars in figure 3 represent 100% of urls mentions within each type of newsgroup. The four portions of each bar add up to 100%. The widths of the bars correspond to the number of url mentions observed in each newsgroup type. Thus we see that the “comp” and “rec” groups were most prevalent.

We performed a precision validation study on our classifier rules by selecting 30 newsgroups at random. We then hand-classified each mention of the three most frequently mentioned urls for each group. We confined our validation to most frequently recommended urls because we were most interested in studying this group of urls and their mentions. We wanted to see whether url mentions that our algorithm classified as recommendations truly were recommendations. This is because a social filtering system incurs more of a penalty for a false positive classification as a recommendation than a false negative. Figure 4 shows that 90% of the url mentions

machine-classified as recommendations or announcements were correctly classified (the left 5 columns total 90%). The inter-rater reliability was 0.84.

Figure 4 Precision validation results: Hand reclassification of 283 machine-classified recommendation type url mentions from 30 randomly selected active groups.

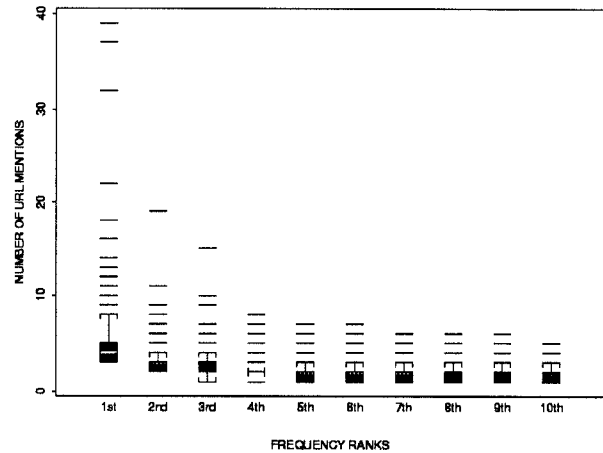


We classified five types of url mentions as recommendations. Of the 283 url mentions classified by our algorithm as recommendations, 68% were isolated url recommendations, 5% occurred in recommended lists of web resources, 1% were FAQs, 9% were “stock” recommendations - the use of a standard text or banner to recommend an url, and 7% were announcements (ANNOUN in figure 5), recommendations where there was a clear relationship between the poster of the message and the web resource recommended.

Frequency Distribution Of Url Mentions

Figure 5 shows, for a second set of 1344 active newsgroups studied from Nov ‘95 to Feb’96, the distributions of the most frequently mentioned url through the 10th most frequently mentioned url. In each vertical column 99% of the data fall within the whiskers. Thus most of the values for the most frequent (the leftmost column) fall between 2 and 8. Horizontal lines above and below the whiskers are outliers. For example, on one newsgroup 39 different people recommended the same url. We interpret these dispersion data tentatively since we have only 3 months of experience. We conclude that there is some variation in frequency-of-mention among newsgroups, with most frequency values in the low single digits. In 484 of the 1344 active groups, the most frequently mentioned url was mentioned by at least 3 different people, and the second most frequently mentioned url was mentioned by at least 2 different people.

Figure 5 Distribution of one-person-one-vote counts for most frequently mentioned urls, most frequent to 10th most frequent, 1344 active groups.



For the time period of our data collection we identified three types of distributional patterns of url recommendations. A majority of newsgroups, primarily those with a smaller flow of messages were “url dribblers”. Posters would recommend urls, but we observed no significant buildup in the frequency with which any given url was recommended. Therefore we saw no evidence that there were any accepted community resources in these groups. soc.college is an example of this kind of group. We observed scores of a second type of frequency distribution for url recommendations that was characterized by a “spike”. One url received virtually all the multiple mentions, often a large number of mentions. These newsgroups were, we believe, public conversations occurring in the context of a definitive text. For example the posters to comp.lang.c++.moderated frequently recommended their newsgroup FAQ. A third pattern, one for which we do not yet have an adequate explanation is the “nice curve” where multiple recommendations are distributed over five to ten urls in the shape of a slalom or quarter circle. This pattern seems to occur frequently in the “rec” category of newsgroup.

PHOAKS SYSTEM

We have experimented with interfaces to the machine classified recommendations that we are collecting. We call this effort PHOAKS, which is an acronym for “People Helping One Another Know Stuff”. Our first experiment is a Web-based system that utilizes resource recommendations discovered in netnews. The PHOAKS system (1) harvests references to urls from Usenet newsgroups, (2) maintains databases of recommended resources and associated information, and (3) creates Web pages that summarize the most frequently and recently mentioned resources. In this section, we describe the PHOAKS system architecture and discuss the design of the Web pages.

System Architecture

PHOAKS currently processes each newsgroup in isolation. A database for each newsgroup records resources, their recommenders, and contextual information from each message that

recommended a resource. Another database organizes information for each contributor. The basic processing loop for each newsgroup is to get new data, integrate the new data into the databases, and generate web pages that summarize the most frequently and recently mentioned resources. We next describe the process in a bit more detail.

1. Scan all new messages for (potential) url mentions, as indicated by the tag "http://". For each url mention, save the following information from the message to use as a context in categorizing and evaluating this url mention:

- All header information,
- At least two lines of text before and after the url. If we come across a second url within three lines of a previous url, however, we save all three intervening lines. This lets us catch and tag long url recommendation lists.

2. For each url mention do:

- Transform it into a canonical form, e.g., stripping the default port (":80") and removing any bracketing punctuation, e.g., a trailing "." or "?" This is important both because we must extract the right text to use, and we want to do the best job possible of detecting when two textually different strings actually refer to the same resource (this is the "url synonym" problem).
- Categorize this url mention, using information from the header and surrounding message context. We currently use the four categories of quoted, signature, recommendation, and other.
- If this url mention was either a recommendation or home page, then create or update database records for the resource and the person who posted the message.

3. Sort the resource database based first on the number of different people who recommended each resource and second on the recency of the recommendations.

4. Select the 40 (this is a parameter that we can vary) most frequently and recently mentioned resources to use in creating web pages for the newsgroup. At this stage, our program attempts to fetch content for each url that points at a html text file. There are two reasons to fetch a url before including it in our pages. First, if we cannot fetch the page, we cannot conclude that it is a valid url. The fast turnover of urls as pages move or go off-line, and simple typos ensure that many potential urls are not really valid. (Of course, just because we cannot fetch a page does not guarantee that it does not exist - the server may be temporarily down or network load may just prevent us from getting there. However, recall that we value precision above recall). Second, we extract the title from the page, which we use to describe the resource in our summary pages (see figure 6). We save resource titles in a database, so we fetch each page only once (and we only attempt to fetch pages that are frequently and recently mentioned enough to appear in our summary.)

5. Generate web pages for the selected resources. We currently generate four types of interlinked pages:

- *resource summaries* - these pages present resources, the people who recommended them, and context from the recommending messages (see figure 6),
- *recommender summaries* - the same data, but organized by the people who recommended resources,
- *resource pages* - for each resource, a list of the people who recommended the resource, and
- *recommender pages* - for each recommender, a list of the resources he or she recommended.

Figure 6 An example summary page concerning the newsgroup *rec.music.dylan* from PHOAKS. (<http://www.phoaks.com/phoaks/>)

People Helping One Another Know Stuff
PHOAKS - People Helping One Another Know Stuff

Resources for PHOAKS : rec . music . dylan

Feedback: Give your opinion of the links on this page Add a link relevant to this newsgroup

Resource	Post(s)	Message Age(s) ↕
01) Bob Links	2	▀▀▀▀▀▀▀▀▀▀▀▀
02) Image at 128.39.161.105	7	▀▀▀▀▀▀▀▀▀▀
03) Tour Update	5	▀▀▀▀▀▀▀▀
04) Telegraph Home Page	4	▀▀▀▀▀▀▀
05) Mailing List WWW Gateway	3	▀▀▀▀▀▀
06) Expecting Rain	3	▀▀▀▀▀▀
07) Jan Halberg's WWW page...	3	▀▀▀▀▀▀
08) "" same title, different location	3	▀▀▀▀▀▀
09) Bringing It All Back Home Page	3	▀▀▀▀▀▀
10) C'Dnow : welcome	2	▀▀▀▀▀▀

[Top 11-20](#) [Top 21-30](#) [Top 31-40](#)
[Sort resources by people who posted](#)
[View other relevant resources](#)

Look for rec.music.dylan FAQ at MIT or Ohio State.
Look for rec.music.dylan usage statistics at NIC.

* Note: each square represents the posting of one resource (e.g., URL) by one person. The darker the square, the more recent the post. Clicking on a square will take you to the part of the posting where the resource appeared.

Your feedback is welcome. Please email thefolks@phoaks.com. If inappropriate resources have snuck through our filters, include their names.

Interface Design Decisions

Our current page design reflects two themes: organizing access to content by social information and seeking a middle ground between privacy and connectedness. These themes play out in many ways. On the theme of organizing access by social information:

- Resources are sorted by the number of different people who recommended them.
- Users have access not just to the number of people who recommended a resource, but also to the names of the recommenders and a list of the other resources they have recommended to the newsgroup. This is useful in at least two ways. First, regular readers of a newsgroup are likely to know who else on the newsgroup has opinions that they trust. Thus, they can investigate just those resources that have been recommended by these people. Second, after investigating and liking a resource, one can find other resources recommended by the people who recommended that resource.
- Users have access to opinion about a resource, as expressed in the surrounding message context. This is especially important because our automatic filters use only

very limited semantic information. Another important benefit is that a message may mention many resources, some of which did not make it into our summary (e.g., because we could not verify their existence or because we categorized them as home pages), yet users still get the chance to visit them when they encounter them in the message context.

- Users can see the timeliness of a resource within a community. A “histogram” of shaded boxes is displayed for each resource, with one box for each person who recommended the resource. The more recent the recommendation, the darker the box. Thus, one can get an impression of whether a resource has been mentioned a lot recently, whether it has been mentioned steadily over time, or whether it appears to have fallen out of favor.
- Users can find out more about people who have recommended resources. We make home pages into hot links. Email addresses are displayed in the message context, so users also can email people they find interesting.

On the theme of seeking middle ground between privacy concerns and opportunities for connectedness:

- We chose not to include a email “mailto:” link because we did not want to make it easy to email recommenders.
- On the other hand, we did include hyperlinks to contributors’ home pages - if they had them. Here our idea was that accessing home pages was not so intrusive as email.
- Other Usenet new access systems such as DejaNews and Alta Vista permit users to search globally by user name in the “:from” field, thus accessing all message posts by a person. We deliberately chose not to include such a facility in PHOAKS. As a result, users may only learn of posters’ activities on a single newsgroup.

The experimental PHOAKS system has been on-line for five months now at <http://www.phoaks.com/phoaks> and its users are offering helpful feedback. As of July 15th, 1996, PHOAKS has served 269K pages (7.7 gigabytes) of social filtering *text* to visitors from 80,000+ hosts in 98 countries.

Community Sorted Search

One way to understand the PHOAKS approach is to compare it with the most common method of locating on-line information, search engines such as Alta Vista, Lycos, and DejaNews. Using a search engine, one constructs a query describing the item being sought. The result is a (typically) large set of documents that match the query. PHOAKS, on the other hand, provides access not to the entire universe of possibly relevant documents, but rather to those documents that a community of experts has been talking about or recommending. Documents are ordered by how often the community has mentioned them. At a finer level, independent opinions about the documents, both positive and negative, are available. Finally, PHOAKS makes timeliness of a resource visible.

We consider a combination of keyword search and social filtering to be extremely promising. We call this combination *community sorted search*. The basic idea is to run a query

using a normal keyword search engine, then to filter the results through the databases maintained by PHOAKS. Results are clustered by the newsgroups that mention them, then ordered by the frequency of mention. For example, figure 7 shows a community sorted search for the query “dylan”. Of all the returned items, three were found to have been mentioned in the newsgroup *comp.lang.dylan*, and 15 in the newsgroup *rec.music.dylan*.

Figure 7 An example of Community Sorted Search. Note that returns from query: “dylan” are clustered and ordered by their appearance in netnews groups: *comp.lang.dylan* and *rec.music.dylan*.

People Helping One Another Know Stuff
“Together, we know it all.”

Community Sorted Search Results

Your Query ["dylan"] returned 100 items

3 items mentioned in comp.lang.dylan

1 Hit: 2 Syntax
Next | Previous | Top | Contents | Index | Dylan Reference Manual. 2 Syntax. Overview. Libraries and Modules. Bindings. Macros. Bodies. Definitions. Local Declarations...
<http://www.cambridge.apple.com/dylan/drm/drm-9.html> - size 2K - 18 Oct 95

1 Hit: 6 Functions
Next | Previous | Top | Contents | Index | Dylan Reference Manual. 6 Functions. Overview. Generic Functions. Methods. Parameter Lists. Kinds of Parameters. Kinds of...
<http://www.cambridge.apple.com/dylan/drm/drm-47.html> - size 2K - 18 Oct 95

1 Hit: 7 Conditions
Next | Previous | Top | Contents | Index | Dylan Reference Manual. 7 Conditions. Background. Overview. Signalers, Conditions, and Handlers. Exception Handling, Stack...
<http://www.cambridge.apple.com/dylan/drm/drm-52.html> - size 1K - 18 Oct 95

15 items mentioned in rec.music.dylan

8 Hits: Bob Dylan
Bob Dylan. Albums. Bob Dylan Official CDs.
<http://bob.nbr.net/dob/dylanalbums.html> - size 151 bytes - 31 Oct 95

6 Hits: No Title
From +dylan Tue Aug 9 07:15:48 1994 From: gerkepott@espe.picker.de Date: 27 Jul 94 08:59:25 CET Newsgroups: rec.music.dylan Subject: notes on Europe tour. report about...
http://reality.sgi.com/employees/howells/euro_tour_notes.html - size 8K - 19 Feb 96

6 Hits: Other Dylan Web Sites
Other Dylan Web Sites. All Dylan Links: Bill Pagel's Boblinks page. Yahoo's Bob Dylan Links. The Official Sony Dylan Home Page: Sony's own Dylan page. Comprehensive...
http://reality.sgi.com/employees/howells/other_sites.html - size 3K - 19 Feb 96

Future Work

We are exploring three research areas. First, the quality of our pages depends on an accurate classification or url mentions. Our analysis pointed the way to a number of improvements to our classifier, including (1) a refined set of categories, (2) better detectors for quotes and home pages, and (3) a spam detector. Further, we are refining our methods of detecting when two (textually) different urls actually refer to the same resource (the url synonym problem). Second, we are planning a host of improvements to the interface. For example, rather than presenting resources in a flat list, we will explore grouping resources by features such as their server or by recommendation type. We also are considering several dynamic visual presentations of our data, both of resources and of the activity of a newsgroup community. Finally, we have much more data analysis to do. In particular we want to further analyze the “dribbler”, “definitive text” and “nice curve” phenomena we observed in our newsgroups. We would also like to understand the web resource recommending behavior in known types of newsgroups such as hobbyist groups, support groups and talk groups. We also will explore the relationship between url mentions and other posting behavior on a newsgroup over time. For example, do recommended urls migrate into the

newsgroup FAQ? In addition, we will analyze patterns of access to our web site.

ACKNOWLEDGEMENTS

Paul Resnick, Lynn Cherny, Julia Hirshberg, Steve Crandall, Don Hindle and Steve Whittaker made many valuable comments on this work. System support from Shankar Ishwar helped to get the PHOAKS web site going. Josh Creter helped with data preparation and analysis. Finally we wish to thank all the people who tried out PHOAKS and gave us feedback.

REFERENCES

1. Allen, R.B. (1990) User models: Theory, method and practice, *International Journal of Man-Machine Studies*, 32, 511-543.
2. Brothers, L., Hollan, J., Nielsen, J., Stornetta, S., Abney, S., Furnas, G., and Littman, M., Supporting informal communication via ephemeral interest groups. *Proc. ACM CSCW'92 Conf. Computer-Supported Cooperative Work* (Toronto, Canada, 1P4 November 1992), 84-90.
3. William W. Cohen. Learning to Classify English Text with ILP Methods, In: *Advances in Inductive Logic Programming* (Ed. L. De Raedt), IOS Press, 1995, to appear in IOS Frontiers in AI and Applications series.
4. Furnas, G., (1995) View Navigable Interfaces, private communication.
5. Goldberg, D., Nichols, D., Oki, B.M. and Terry, D. (1992) Using Collaborative Filtering to Weave an Information Tapestry. *Communications of the ACM*, 35, 12, pp. 51-60.
6. Grudin, J., Social Evaluation of the User Interface: Who Does the Work and Who Gets the BENEFIT?, Proceedings of IFIP INTERACT'87: Human-Computer Interaction, 1987, 805-811.
7. 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*. ACM Press, New York City, New York, pp. 194-201.
8. Hill, W. C., Hollan, J. D., Wroblewski, D., and McCandless, T. (1992) Edit Wear and Read Wear. In: *Proceedings of ACM Conference on Human Factors in Computing Systems, CHI'92*. ACM Press, New York City, New York, pp.3-9.
9. Hill, W.C., Hollan, J.D. (1994) History-Enriched Digital Objects: Prototypes and Policy Issues, *The Information Society*, 10, pp. 139-145.
10. Malone, T.W., Grant, K.R., Turbak, F.A., Brobst, S.A. and Cohen, M.D. (1987) Intelligent Information Sharing Systems. *Communications of the ACM*, 30, 5, pp. 390-402.
11. 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*. ACM Press, New York City, New York, pp. 202-209.
12. Morita, M., Shinoda, Y. (1994) Information Filtering Based on User Behavior Analysis and Best Match Text Retrieval, *Proceedings of the 17th Annual International SIGIR Conference on Research and Development*, pp. 272-281.
13. Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., Riedl, J. (1994) GroupLens: An Open Architecture for Collaborative Filtering of Netnews. In: *Proceeding of the ACM 1994 Conference on Computer Supported Cooperative Work*. pp. 175-186.
14. Rheingold, H., (1993) *The virtual community: homesteading on the electronic frontier*, Reading Mass: Addison-Wesley.
15. 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*. ACM Press, New York City, New York, pp. 210-217.
16. Wittenburg, K., Das, D., Hill, W., Stead, L. Group Asynchronous Browsing on the World Wide Web. In: *Fourth International World Wide Web Conference Proceedings*. O'Reilly & Associates, New York City, New York, pp. 51-62.
17. Wroblewski, D., McCandless, T., Hill, W. (1994) Advertisements, Proxies and Wear: Three Methods for Feedback in Interactive Systems, in *Dialogue and Instruction*, Beun, R., Baker, M., and Reiner, M. editors. Springer-Verlag (forthcoming).