

Do-I-Care: A Collaborative Web Agent

Brian Starr

Mark S. Ackerman

Michael Pazzani

Information and Computer Science

University of California, Irvine

{bstarr, ackerman, pazzani}@ics.uci.edu

<http://www.ics.uci.edu/CORPS/dica.html>

ABSTRACT

Social filtering and collaborative resource discovery mechanisms often fail because of the extra burden, even tiny, placed on the user. This work proposes an innovative World Wide Web agent that uses a model of collaboration that leverages the natural incentives for individual users to easily provide for collaborative work.

KEYWORDS: computer-supported cooperative work, CSCW, social filtering, collaboration, World Wide Web

INTRODUCTION

Collaborative efforts to discover Internet resources, such as interesting Web pages, require work by users. Unfortunately, users often will not do the extra, altruistic work, leading to adoption failures for social filtering systems.

In the work described below, individuals engage in resource discovery for themselves. Without extra work, other individuals can use that effort.

We do this by using a naturally reoccurring problem with the Web. After one has discovered Web resources, a problem remains: How do you know when to *revisit* those resources for new material? Consider some common post-discovery situations:

- You know the page where a colleague has his publications. You would like to obtain new papers as they are posted. Furthermore, you would like to know about interesting papers, rather than just any papers.
- You have obtained the page for the low cost clothes merchandiser. You would like to know when there is a sale on jeans or Hawaiian shirts.

These are examples of change events that occur somewhat unpredictably and offer no explicit notification. Today, interested parties must occasionally check by hand. This can be burdensome and tedious, and one may forget to do it on a timely basis. Furthermore, we would like an agent to discover *interesting* changes, rather than some insignificant change.

In order to track interesting changes, an agent needs to

know what changes are interesting and where to find them. This paper is an overview of an agent that collaborates with both users and their peers to identify potentially interesting changes. The agent works by soliciting user opinions about changes it finds to train a user model; an individual does this for her own good. The agent can then share these opinions and findings with agents of like-minded users, providing for collaborative effort. This agent, Do-I-Care, and its model of collaboration are described below.

LOCATING INTERESTING INFORMATION

There has been considerable effort devoted to the resource discovery problem on the Web, mostly focusing on three types of technical methods. The first, active browser systems (e.g., Fish Search and WebWatcher), offer navigational suggestions that direct the user along a path. User models are obtained directly from users while they are browsing and require their active participation.

The second, index creation and lookup systems (e.g., Harvest and Lycos), treat the Web as a database of information to be cataloged and indexed. Users must be able to articulate their information goals and may need to periodically resubmit their searches to obtain new or future information.

There are also several agents that monitor Web pages for *any* modification, such as URL-minder. However, many of the changes detected may be irrelevant. Also, the user must still discover the Web sites to be monitored.

SOCIAL RESOURCE DISCOVERY

Although technical discovery methods are valuable, they neglect perhaps the single most important method of discovery that people rely on -- other people. Social resource discovery assumes the existence of others, who have located and evaluated relevant resources. Opinions of like-minded individuals are also assumed to have more discriminatory value than an automated evaluation. The goal of social resource discovery systems is to aggregate and share the fruits of individual activity and knowledge.

Relatively few social resource discovery systems currently exist. The Pointers system [3] facilitates the distribution of links to resources with accompanying context. While the benefits to a pointer's recipient seem clear, the system relies on a provider's desire to be helpful that may not always exist.

Our work is closer in emphasis to Ringo [6] and GroupLens [5]. GroupLens uses newsgroup article ratings to predict an article's relevance to a given user. Because the weights assigned to other's opinions are proportionate to their past correlation with a user's opinions, there is some benefit for users to offer opinions. However, in both systems, the emphasis is on providing collaborative effort, for unclear personal gain.

The Do-I-Care agent attempts to avoid free-loading by exploiting the naturally occurring individual effort in a way that facilitates information sharing without incurring any significant additional costs. It is similar in its incentives to Answer Garden [1].

DO-I-CARE

Do-I-Care was designed to help users discover *interesting* changes on the Web, using both technical and social means. Do-I-Care agents automate periodic visits to selected pages to detect interesting changes on behalf of individuals. They use machine learning to identify by example what changes are interesting, and how often they occur. It is in users' best interests to keep their personal agents informed about relevant pages and about the quality of reported changes. In so doing, they can automatically contribute to the common good, because Do-I-Care agents can also share information with other agents.

Because of space limitations, we can only briefly discuss the system architecture. Each Do-I-Care agent must:

- Periodically visit a user-defined list of target pages.
- Identify any changes since the last time it visited.
- Decide whether the changes were interesting
- Notify the user if the change was interesting.
- Accept relevance feedback on the interestingness of the change and timeliness of the notification.
- Facilitate information sharing and collaboration between individuals and groups.

We assume that Web pages generally change incrementally and that the user has a list of Web pages that will have interesting information. By limiting the search to a small subset of Web pages specifically selected *because* they are likely to contain interesting information, we can greatly improve precision. By accepting or rejecting Do-I-Care's suggestions, the user refines the types of changes that are interesting.

Since users provide relevance feedback, we can view the process of determining whether a change is interesting as a classification task, and the process of learning a user profile is a supervised learning task. We represent each change by a relatively small number of features. We find words that are informative by computing the mutual information [4] between the presence of a word in the change and the relevance feedback. In addition, we use other change attributes, such as size and the number of links. A simple Bayesian classifier [2] then determines whether the change is interesting.

Because different pages may be interesting for different reasons, users may maintain multiple Do-I-Care agents specialized on particular topics or kinds of Web pages. For example, the criteria for an interesting journal announcement may be different from that for a news item.

Once an agent spots an interesting change, the user is notified by e-mail, and the change is appended to the agent's associated Web page. This Web page is also used for relevance feedback and collaborative activity.

SHARING INTERESTING CHANGES

Since all changes are reported through an agent's Web page, agents can use information from other agent's pages for their own change notifications. The technical mechanism is simple. Users who wish to allow their agent's Web page to be cascaded simply notify others of its existence. Cascaded agents' owners continue to use their agents as before, but now they are also sharing both their lists of interesting sites and their opinions about changes with others. Agents can be cascaded any number of times, allowing the creation of group or organizational "new and interesting" pages.

Users of cascaded agents benefit by leveraging the work and judgment of trusted peers. Gathering changes from cascaded agents increases recall, while exploiting other's filtering increases precision. There is little social cost, since no additional work is generated for anyone. Thereby, the perceived costs and benefits favor collaboration.

SUMMARY

Do-I-Care attempts to solve two post-discovery problems: when should a user revisit a known site for new information, and how does a user share new information with others who may be interested. These post-discovery chores are laborious, suffer from inequitable distribution of costs and benefits, and are not both addressed by existing systems. By providing a simple technical mechanism, we have found a collaboration mechanism that provides for group effort through individual incentives.

REFERENCES

1. Ackerman, Mark S. Answer Garden: A Tool for Growing Organizational Memory. Dissertation, Massachusetts Institute of Technology, 1993.
2. Duda, R., and P. Hart. Pattern classification and scene analysis. New York: John Wiley and Sons, 1973.
3. Maltz, D., and K. Ehrlich. Pointing The Way: Active Collaborative Filtering. *CHI 95*: 202-209.
4. Quinlan, J. R. Induction of decision trees. *Machine Learning*, 1 (1), 1986: 81-106.
5. Resnick, P., N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl. GroupLens. *Proceedings of CSCW 94*: 175-186
6. Shardanand, Upendra, and Pattie Maes. Social Information Filtering. *Proceedings of CHI'95*: 210-217.