WebWatcher: Knowledge Navigation in the World Wide Web

Dayne Freitag, Thorsten Joachims, and Tom Mitchell
School of Computer Science
Carnegie Mellon University
August 15, 1995

1 Learning to Navigate

Many have noted the need for software to assist people in locating information on the World Wide Web. Although effective tools exist, they typically rely on brute-force scanning and indexing of Web pages for later keyword-based retrieval. Such tools ignore at least two sources of knowledge which might prove useful in navigation and retrieval: (1) the structure of the Web as a *graph*, and (2) the *sequence* in which human users search the web.

We describe a learning apprentice system, called WebWatcher, which both performs the kind of indexing used by Web catalogers like Lycos, and attempts to exploit the two sources of knowledge listed above.

2 WebWatcher

A *learning apprentice* system is a personal software application that acquires knowledge of its task by observing routine user actions. The task performed by WebWatcher is to serve as a tour guide to Web neighborhoods. It attempts to determine a user's interests, both by asking for an explicit textual goal when the system is entered, and by inferring the goal from user actions. While the user browses, WebWatcher tracks user actions, suggests hyperlinks it believes are promising on the current page, and occasionally suggests pages not adjacent to the current page which it thinks will interest the user or meet the user's needs. Its advice is based on expertise learned as a side effect of giving earlier tours.

From the user's perspective WebWatcher requires little expertise and no special software. Instead of a modified browser, WebWatcher exists as an intermediate server between the client-side browser, such as Mosaic or Netscape, and the machine serving the pages the user wants to see. This approach has several advantages:

1. Entering the system simply involves following an appropriately encoded link to the WebWatcher front door.[1]
2. WebWatcher can be used from any site with normal internet access and a web browser with graphic capabilities.
3. Because WebWatcher knows the page from which any given user started, it can learn about *local* web neighborhoods, in effect acting like many specialized assistants.

[1] One such link may be found on the SCS Front Door, `http://www.cs.cmu.edu/Web/FrontDoor.html`.

Once in the system, the user can browse as usual. WebWatcher makes only the minimal changes necessary to suggest links and provide the user with an opportunity for feedback. These changes are limited to icons inserted around promising links, a menu bar at the top of the page, and occasional page suggestions above the menu bar.

Most importantly, WebWatcher records user actions and learns from user traces how to become a better tour guide. The main thrust of our research to date has involved exploring ways to learn how to identify which links on a page are most likely to be useful to the user.

3 Status and Future Directions

WebWatcher has collected over 1300 user traces, more than 700 of these since its official release in early August.

WebWatcher link suggestions are based primarily on knowledge gleaned from these traces. [Armstrong et al., 1995] describes experiments using a variety of machine learning methods to acquire knowledge for providing advice. We are now beginning a set of experiments to test the effectivenss of the system over a large set of data.

Future research questions include:

- How can we use to best advantage other sources of information in the SCS environment about user search patterns?
- How can WebWatcher exploit knowledge made available by other text learning systems in our local project group?

References

[Armstrong et al., 1995] Robert Armstrong, Dayne Freitag, Thorsten Joachims, and Tom Mitchell, "WebWatcher: A Learning Apprentice for the World Wide Web," *AAAI Spring Symposium on Information Gathering from Distributed Heterogeneous Sources*, 1995.