

Exploring Web Browser History Comparisons

Mark Bilezikjian, John C. Tang, James “Bo” Begole, Nicole Yankelovich

Sun Microsystems Laboratories

901 San Antonio Road, MTV29-230

Palo Alto, CA 94303

Mark.Bilezikjian, John.Tang, Bo.Begole, Nicole.Yankelovich@Sun.COM

ABSTRACT

This work explores how comparing web navigation histories between two people and presenting the results to them might allow them to gain insight about each other. We developed a prototype that presents web matches sorted according to frequency, recency, and web site. Interviews with users of the prototype suggest that common interests and preferences can be inferred from these comparisons.

Keywords

Web navigation histories, web usage mining, shared context, interpersonal context, CSCW

INTRODUCTION

Web browser navigation histories are a rich source of information about an individual's personal and work-related interests. We conjecture that comparing two peoples' web histories can suggest commonalities and interpersonal context to them. This information can help formulate an effective way to approach an unknown person. It can also help people who are already familiar with each other to discover, access, and document their shared context.

A huge benefit of using web histories or browsing patterns to indicate user interests is that the application needs no additional user input. Other “zero-input interfaces”[3] in this space include a web page recommender application [1], a browsing agent [3], and a prediction model for web caching and pre-fetching [4]. Unlike those studies, which are focused on augmenting a user's browsing experience based on collaborative filtering or predictive models, our system infers nothing about the user's preferences or future actions. Instead, it presents what context two individuals have in common by displaying their web history matches, from which the user can make inferences.

To explore this concept, we wrote an algorithm that compares your navigation history with another person's and presents all the matching URLs between them, sorted according to various metrics. After collecting navigation histories from several colleagues, we conducted exploratory interviews with some of them as they viewed their matches.

COMPARING WEB HISTORIES

For our preliminary study we asked colleagues and friends for permission to use their Netscape navigation histories. These files cache the URLs of all the web sites that a person has browsed in the past nine days (default value). Each entry in a history file includes a URL, page title, most recent date visited, first date visited, and the number of times the user has visited the page since the first recorded visit.

Since nine days is only a limited window of navigation history, we supported concatenating multiple history files to accumulate a longer history.

Our matching algorithm finds and lists all URL matches between the entries of the navigation histories for any pair of people. During this process, the algorithm tries to filter any advertisements from the match results by removing any URLs that match a list of known ads.

We distinguish two types of matches in the results. Exact matches indicate that the two individuals visited the same exact URL. Partial matches indicate that the pair visited URLs on the same web site but did not visit the same exact URL (web page). We display the partially matching URLs up to the point at which the two URLs diverge.

The matching algorithm creates three uniquely sorted lists of the matches that were found. The full list of matches groups them according to web site host. Another list displays 25 matches sorted by the recency with which the person you are matching against last visited the site, most recent first. The final list displays 25 matches sorted by overlap frequency, the largest number of times that both parties visited a particular page (i.e., the lesser of the two people's visit frequency), largest overlap first.

INTERVIEWS AND RESULTS

We conducted interviews with four of nine colleagues that sent us at least two of their history files (giving us a minimum of eighteen days of browser history). Each participant was shown at least two match results: one with someone he knew, and one with a person unknown to him. Most of the results contained few matches because of the short time span that the navigation histories represented.

The first encouraging observation was that all the participants expressed intrigue as they began to view the matches. Most participants used the list sorted by overlap frequency. One participant observed that the most interesting matches had the lowest overlap frequency, since those were the obscure matches. Another participant commented that the matches in the middle of the overlap frequency range were most interesting. Thus, people had differing reactions about what regions of the overlap-frequency sort were of interest.

Only one of the participants had more than 25 matches. When he used the list sorted by web site to view all the matches, he seemed to find interesting matches easily. Scanning through a list of sites allowed him to see a broader spectrum of matches more quickly than scanning through individual URLs. We expect that once the number

of matches increases, more users will view the full listing by host. We also expect that after collecting history files for a longer period of time, matches sorted by recency may become more useful.

One subject voiced strong concerns about the privacy implications of viewing web matches. Another subject was concerned, and two were not concerned at all. We discussed some responses to their privacy concerns. First, since the list displays web matches that are common between the two viewers, they are mutually complicit in viewing any questionable material in the list. Unlike Collab-Clio[2], which allows you to search someone else's entire web history, our system only presents web pages that both users viewed. All of the participants agreed that this mitigated concerns about spying on another person's browsing.

Furthermore, the prototype requires people to explicitly opt in to using it. Two participants needed more privacy protection before they would opt in for an actual matching service. The other two were excited about the work and said that they would join the service. We proposed creating an editing tool to allow users to delete items from their navigation histories. Nobody seemed interested in taking time to edit their history files.

Matches With Familiar vs. Unknown People

Matches for comparable web sites evoked different reactions depending on whether they were matching with a familiar person or someone unknown to them. Upon seeing a match for the satire site "The Onion" with a colleague, one of the participants was not interested, surprised, or newly informed that his coworker also visited the site. Conversely, seeing a match for the same site with an unknown person prompted a participant to comment, "Obviously a tasteful person." The match evoked a playful reaction suggesting that the participant enjoyed finding a personal taste that he shared with the unknown person.

Unknown People

Other sites that caused participants to comment on matches with unknown people were more general sites like "Google," "MyYahoo," and "CNN," which were often the matches with the highest overlap frequency. There were differing opinions on the usefulness of these matches. One participant commented that these sites are so large and general that few interesting inferences could be made from the site level matches, but a specific article within a large site caught his interest. Another participant used the unfamiliar person's choice of the "Google" search engine to infer that the user is a technically savvy person due to his use of a "superior" search engine. This participant made several other inferences from the matches, but stated that they were all preliminary inferences that might be reinforced upon making personal contact with the person.

We categorized inferences about unknown people into two groups. Participants inferred discrete facts like employer, job role, etc. They also made value-based inferences where the value that a viewer places on a particular web site was

projected onto the unknown co-worker who also visited that site. Negative prejudices will not likely be projected since the viewers only see web sites that they also visit.

Familiar People

Participants viewing matches with familiar people tended to find the obscure matches more interesting since many of the matched sites came as no surprise. Some matches evoked a subtle effect of discovering meta-information about the match. Upon seeing obscure matches, one participant began to speculate as to how both he and his friend had visited that site. In one case he was able to recall a third party who had circulated the URL and deduced that his friend must have received it from the same person. Many participants were excited at the possibility of finding common interests with friends or being able to use the tool as a collaborative web investigation utility.

SUMMARY

Our observations from the interviews suggest that viewing a history comparison can help someone form initial expectations of an unknown person. Users' comments also suggest that viewing matches against a familiar person can produce interesting discussion topics and suggest meta-information not shown directly by the match. These two observations of the utility of the match results support our conjecture that people can use them to discover interpersonal context between themselves and others.

Potential future work for this project includes efforts to increase the number of matches by collecting more history files and by seeding the navigation histories with the URLs from bookmarks. Our interview data suggest that people are interested in contextual information about colleagues, but that the web histories alone may not provide a sufficient level of detail. We are considering embedding the history comparisons into a broader application that includes multiple sources of contextual information. For example, embedding web history comparisons in a corporate directory could give you access to a more personal dimension on the relationship between you and the person you look up.

ACKNOWLEDGMENTS

We thank Paul Marquess for his help with DB_File and all participants in this investigation.

REFERENCES

1. Fu, X., Budzik, J., and Hammond, K.J. Mining navigation history for recommendation. In *Proceedings of Intelligent User Interfaces 2000*. ACM Press. 2000.
2. Lau T., Etzioni, O. & Weld, D. Privacy Interfaces for Information Management. *Communications of the ACM* 42(10), October 1999. pg. 89-94
3. Lieberman, H., Fry C., and Weitzman, L. Exploring the Web with Reconnaissance Agents. *Communications of the ACM* 44(8). 2001. 69 – 75.
4. Yang, Q., Zhang, H., and Li, T. Mining Web Logs for Prediction Models in WWW Caching and Prefetching. In *Proceedings of ACM SIGKDD International Conference 2001*. ACM Press. 2001.