Eyebrowse: Real-Time Social Web Browsing

Brennan Moore, Max Van Kleek, Christina Xu
MIT
32 Vassar Street
Cambridge, MA 02139

David Karger
MIT CSAIL
32 Vassar Street
Cambridge, MA 02139
karger@csail.mit.edu

ABSTRACT
In this paper we explore the potential for letting users easily and automatically track, analyze, and publish their web browsing activities in real time to share with others on the Web. We developed a system, Eyebrowse, with three goals in mind: first, to help individuals better understand their own longitudinal web browsing activities through visualizations aggregated over time; second, to facilitate sharing, comparison, and increased social awareness of browsing patterns among friends; and finally, to form a public, democratized corpus of web browsing data for the research community. We examine whether users perceive such logging and analysis as useful from a personal and social perspective, and discuss insights into perceptions surrounding privacy and self-exposure pertaining to the public display of personal web browsing behavior.

Author Keywords
Information visualization, real-time, social browsing, life-tracking, temporal data, personal reflection

ACM Classification Keywords
H.5.2. Information interfaces and presentation

INTRODUCTION
Today, we rely increasingly on the Web for a multitude of everyday activities that run the gamut from simple queries to complex social interactions. As a result, our browsing patterns are starting to reflect the intricate and multi-faceted nature of our daily lives[1], but web browsers retain little of the nuanced richness of this information beyond simple "page histories" of previously visited sites. Analytics providers such as Google and Alexa regularly collect statistics of browsing activity, but such analytics are site-centric and not clustered around individual end-users. Moreover, despite the social nature of web browsing, individuals have little awareness of what others are looking at and how often; while sites like del.icio.us facilitate social exploration, they focus on what people choose to share rather than on their actual habits.

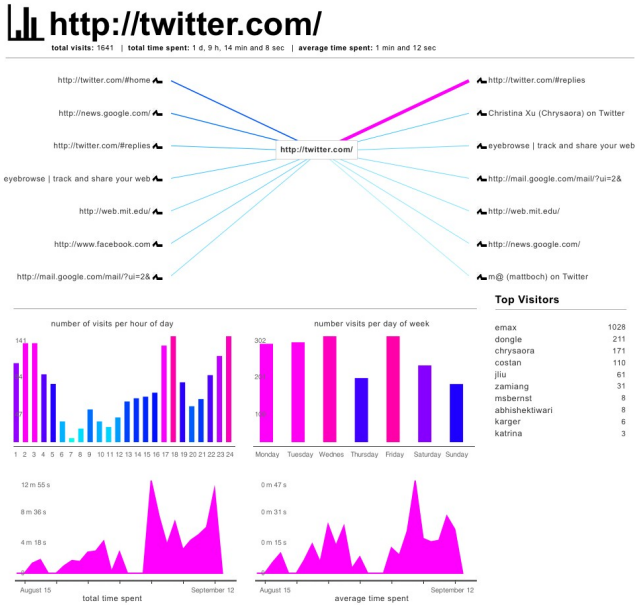

Figure 1: For any input url, top places to go before and after, common times accessed and top visitors.

In this paper, we explore whether the high-fidelity logging of individual web browsing histories can contribute to social awareness and personal reflection. We created a light-weight and non-invasive way for users to share their web browsing activity and examine its qualities and trends graphically. By making web browsing activity public, we maximize its social impact, but we must also take into account the user's privacy concerns. So, we ask, how much are people willing to share, and how does sharing impact the web browsing experiences and habits of the individual?

MOTIVATION
When asked what resources they turned to when they failed to find information through online search queries, the most common responses among fellow researchers, were social in nature, e.g. contacting a friend or posting to a forum. From this initial investigation, we concluded that social and contextual cues fill in the gaps left by semantic search. However, there are currently few effective querying methods for following common and socially related paths through the web. This dearth motivated us to explore

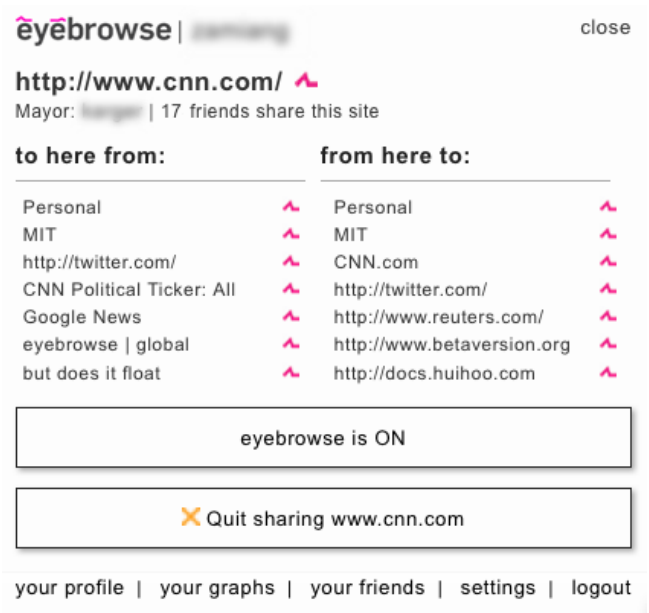

Figure 2: View of the *eyebrowse* plugin for a user on cnn.com showing socially and contextually related information.

querying on a more personal and adaptive level using just-in-time[2] statistics. Furthermore, we were perturbed that only large corporations currently have access to a corpus of web browsing data for research purposes. With Eyebrowse, we wish democratize this information.

RELATED WORK

Eyebrowse occupies a unique intersection between life-tracking and social discovery tools, balancing introspection and public broadcasting. As a tool for self-reflection, eyebrowse borrows longitudinal activity logs and visualizations from time- and task-management software such as RescueTime and Slife. Unlike these services, it publicly displays user activity in the style of social discovery aids such as Last.fm, diigo, and wakoopa. Like FourSquare, it playfully promotes awareness of activities within pre-existing social networks and adds ideas explored in WebWatcher, through using learned web browsing habits to help users navigate the web.

EYEBROWSE

Eyebrowse is a Firefox plugin and a website. The plugin, a small icon in the bottom-right corner of the browser, controls logging and is intended to provide relevant information within the context of browsing while the website facilitates more in-depth personal and social explorations. After signing up and installing the plugin, the user is brought to their settings page where they can add domains to their 'whitelist.' This list controls what URLs are recorded and shared. During testing, we experimented with having both a whitelist and a blacklist (all URLs except those in the list are logged), but decided to choose just one after finding that users often confused the two. A quick

Figure 3: A user's profile page, showing a most recent and a macroscopic view of the users browsing habits

glance at the icon indicates whether the current URL is being recorded. When browsing a page not on the whitelist, a user can add their current URL through the Eyebrowse icon. When active, the plugin records the page title, URL, duration of visit, and tab switches and sends this information to the Eyebrowse server. Clicking on the Eyebrowse icon engages a popup box (Figure 2), which displays social and contextual information about the current page, including which friend spends the most time on the current URL, how many friends have recorded that URL and a ranked list of the most common sites visited before and after. By clicking on the graph icon in the plugin, a user may visit the 'page stats' page for that URL (Figure 1). This page is a means for context-based search, which in addition to showing the information found in the plugin, supports querying and enables the discovery of longitudinal associations.

Once a visit is shared, it is shown in a variety of ways on the user's profile page (Figure 3) and graphs page (Figures 4-6). When visiting their own profile page, a user is presented with a view of the last 20 URLs they visited and a weekly macroscopic view where they can learn about browsing trends and activity statistics. Other statistics, such as the top 25 websites by week and information about frequency and duration of web browsing activity lead to

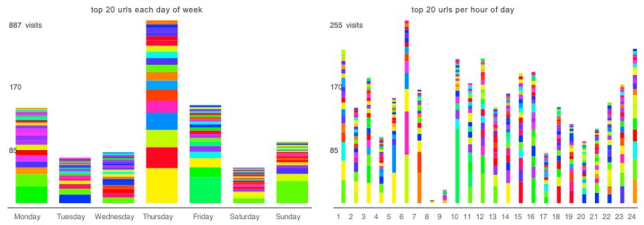

Figure 4: Top 20 URLs for day of the week and time of day

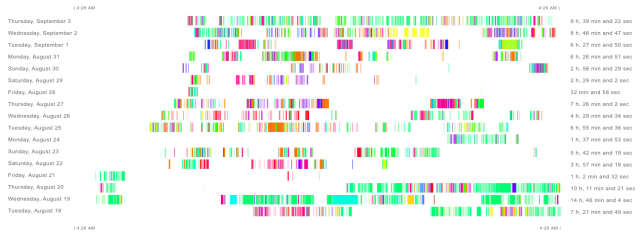

Figure 6: User activity over 20 days and time spent each day

further exploration through their timeline graph (Figure 5) and "20 days" graph (Figure 6). In the "day by day" graph (Figure 7), users can assess daily activity patterns overall, and for their most visited URLs. The interactive visualizations were created using the HTML5 canvas element and are designed for ease of use while compelling users to notice patterns, play with the visualizations, and share more URLs. Colors are assigned such that color similarity correlates to URL edit distance.

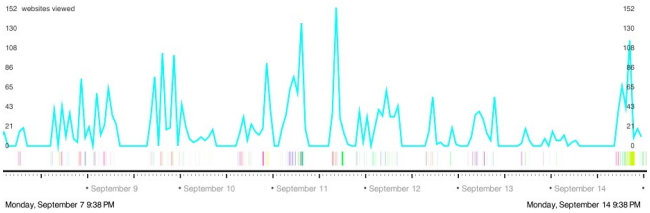

Figure 5: A user's urls visited over the course of one week

When viewing another user's profile, the graphs facilitate social discovery by showing recent activity, a general overview of how that user spends their time online, and web viewing trends over the past week. Users can make social comparisons by clicking 'you minus me' or 'me minus you' to better understand how their browsing habits differ.

USER STUDY

To assess the usefulness of our system and better understand the privacy concerns associated with it, we surveyed 13 of our most active users, a tech-savvy group between the ages of 20 and 40 with 61% using social networking, microblogging, photo-sharing, and blogging sites in tandem.

Social browsing versus self reflection

Prior to sending out the survey, many users expressed interest in a tool that would simply record and visualize their browsing activity locally without sharing it. When asked what Eyebrowse is useful for, we received many responses akin to *"finding out how much time I waste on a few websites"* and *"figuring out how much time I spent on certain projects."*

We asked users to rate each element of Eyebrowse on a 5 point scale from not helpful to very helpful, hoping to learn whether self-reflection or social information discovery was a more important feature of Eyebrowse. Despite that on average, users found Eyebrowse useful for social browsing and self-reflection on similar levels, users reported a wider usefulness for social browsing information. Out of the elements useful for self-reflection, the list comparing top 25 URLs between weeks was reported to be the most helpful with 44% of users reporting positively.

When asked if Eyebrowse was useful for discovering new people, 69% responded with a 1 (not really) or 2 on 7-point scale. However, 47% of users reported that other user's profiles were useful, a higher percentage than the average helpfulness for the self-reflective graphs; user comments indicate that social awareness about established friends was the reason for this. It is interesting to note that while only 25% of users found the "page stats" page useful, the lowest of any feature, the less-detailed but almost identical plugin view was one of the most positively received (41%). We believe this indicates that social browsing is most useful when contextually relevant.

Privacy concerns

Because Eyebrowse allows for domain-level sharing control, we assumed that users were more likely to modify their whitelists and page logs than their browsing habits. 69% of users removed entries from their whitelists while 21% of users removed at least one record. Surprisingly, all of these users cited unexpectedly informative page titles on social networking sites as the reason. This contradictory behavior of adding a URL and then removing it indicates that users are curious about tracking their overall browsing activity on these sites and do not mind sharing general statistics about usage, but find too-detailed information about intrasite activities an invasion of privacy. In comments about whether reputation was a concern for sharing web browsing activity, users expressed a desire to control not only what information is shared but also whom it is shared with. One user elaborated:

If my advisor finds out how much time I spend on Facebook and Gmail, I'm screwed. If my girlfriend sees my habit of clicking random girl photos on Facebook, I'm screwed.

Despite these privacy concerns, the 20 active users (at least 3 domains in whitelist and 10 sites logged) had an average of 56 domains in their whitelist and logged an average of 2,460 instances of browsing over 1 month. Indeed, we would like to think that privacy concerns from our users reflects an increased awareness of what others (e.g. Google Analytics, Alexa) could be collecting about surfing habits.

Public data

By making web browsing public, we hope to help researchers better understand how people use the web and make tools that respect users' privacy. Over the course of our one month deployment, we have gathered a corpus of over 52,000 web browsing trails. For this corpus to be useful for research, data needs to be gathered from a larger user base over a more extend period of time. We hope to greatly expand the corpus to facilitate future studies by any interested parties.

DISCUSSION

Our curiosity as to whether the sharing of socially-sensitive information is more of a privacy concern than other browsing information was answered by conclusive results from our user study. At time of writing, the most common whitelisted sites are nytimes.com, twitter.com, and wikipedia.com; private social sites are surprisingly low on the list, with facebook.com and gmail.com coming in 12th and 13th, respectively. In a free response about privacy concerns, 92% of users reported social- or work-related privacy concerns and their comments indicated a fear of being misrepresented by their web browsing activity, best expressed here:

I was mostly concerned that people would be curious as to why I was looking up certain individuals and draw the wrong conclusions.

This leads us to believe that the cause of privacy concerns is not the web browsing content itself but the activities that the content implies.

We are concerned that the omission of these social sites from a dataset greatly limits the potential uses for web browsing statistics. One potential solution is the creation of a "graylist" for domains that users want to log generally, but with truncated page names that do not reveal specific information. Another solution may be for users to have multiple whitelists that automatically adapt logging by inferring real-world context from the user's web browsing patterns.

CONCLUSION AND FUTURE WORK

Eyerowse distinguishes from previous work by exploring the intersection between life-tracking and social discovery. We have presented evidence that suggests that by using these ideas in tandem, we are capable of helping many users with information-seeking and self-reflection without significant individual burdens on the users.

Instead of speculating about users' privacy limits and interest in introspection, we built a tool and found that there are users who are willing to share their browsing activity and find others' shared information useful. We tested a wide variety of visualizations for making social browsing statistics useful by allowing users to see their own browsing habits in a variety of ways on their own profiles, through viewing other user's profiles and through seeing socially derived information in context of their own web browsing activity. Based on our initial exploration, we are optimistic that making browsing activity social and visible can help users better understand their web browsing habits and navigate the web despite associated privacy concerns. We noted that by making information more accessible and socially- and contextually-aware, it becomes significantly more useful for providing trusted pathways through the web even if it must be less detailed. Although we cannot conclude that Eyebrowse's current features are useful for self-reflection, we see several potential avenues for making them more effective. For example, grouping or tagging URLs by life activity may allow users to provide further context and make more sense of their statistics. Furthermore, in order to achieve the logging fidelity needed for accurate self-observance, it may be necessary to give users greater control over what of their logged activity is shared.

We hope to eventually turn Eyebrowse into a service that supports social browsing through collaborative filtering and other crowd-sourcing techniques, promotes self-awareness among users of both the patterns in their browsing activities, and provides researchers with useful web browsing data without violating the users' privacy sensibilities.

ACKNOWLEDGMENTS

Left blank for blind review

REFERENCES

1. Hogan, B. J. 2009. Networking in Everyday Life. MS thesis, University of Toronto, 2009
2. Rhodes, B. J. and Maes, P. 2000. Just-in-time information retrieval agents. IBM Sys*t. J.* 39, 3-4 (Jul. 2000), 685-704.
3. Narayanan, A. and Shmatikov, V. De-Anonymizing Social Networks In Proc. of *30th IEEE Symposium on Security and Privacy*, Oakland, CA, May 2009, pp. 173-187. IEEE Computer Society
4. Dumais, S., Cutrell, E., Cadiz, J., Jancke, G., Sarin, R., and Robbins, D. Stuff I've Seen: A System for Personal Information Retrieval and Re-Use. In Proc. SIGIR 2003, ACM Press (2003).
5. Armstrong, R., Freitag, D., Joachims, T., and Tom Mitchell, T. WebWatcher: A Learning Apprentice for the World Wide Web. AAAI Spring Symposium on Information Gathering In Heterogeneous, Distributed Environments (1995)