# Eyebrowse: Real-Time Web Activity Sharing and Visualization

**Max Van Kleek**
MIT CSAIL
32 Vassar St.
Cambridge, MA 02139
emax@csail.mit.edu

**Christina Xu**
MIT Media Lab
20 Ames St.
Cambridge, MA 02139
ckx@mit.edu

**Brennan Moore**
MIT CSAIL
32 Vassar St.
Cambridge, MA 02139
zamiang@csail.mit.edu

**David R. Karger**
MIT CSAIL
32 Vassar St.
Cambridge, MA 02139
karger@mit.edu


*CHI 2010*, April 10–15, 2010. Atlanta, Georia, USA
ACM 978-1-60558-930-5/10/04.

## Abstract
In this paper, we explore the potential for letting users automatically track and selectively publish their web browsing activities in real time on the Web. We developed a system, Eyebrowse, with three goals: first, to provide a means for individuals to better understand how they spend time on the web through visualizations and statistics; secondly, to foster social discovery and awareness through real-time web activity sharing; and finally, to build a large public corpus of web browsing trails using this method. We gathered user impressions of Eyebrowse, including perceived usefulness, feelings of self-exposure, and privacy concerns, for ascertaining ways to improve the system.

## Keywords
personal informatics, awareness, life-tracking, information visualization

## ACM Classification Keywords
H.5.2 Information Interfaces and Presentation: User interfaces – Evaluation/ methodology

## Introduction

As the web has developed, we have become increasingly reliant upon it for the gamut of our daily activities, whether work-related, social, or personal. As a result, our web browsing trails are increasingly reflective of our interests, needs, and what we do in our daily lives. Yet, information about our web access patterns is not leveraged in the browser beyond simple "recently visited" features. Moreover, we are poor at keeping track of how much time we spend viewing particular pages, and often forget when and which sites we visited more than a few days in the past. Furthermore, despite the social nature of web browsing, individuals have little awareness of what their friends or others are looking at, at any particular moment, or of their long-term browsing habits and activities.

While web analytics providers[1] collect long-term statistics using server-side logging, these statistics are not well-suited for end-users for several reasons. Much of the data that is most relevant to users is simply not available in server-side logs, such as how long a user viewed a particular site, or even (due to caching proxies) all the times the site was accessed. Secondly, such analytics are often restricted to one particular site, and do not capture web browsing history across sites. Finally, this information is often closely safeguarded by site owners – and is rarely divulged to end-users.

We are exploring an alternative approach which focuses on public, opt-in, longitudinal logging of individuals' web browsing activities using client-side browser instrumentation instead of server-side log analysis. The goal of our system, called Eyebrowse, is to provide an approach that is more palatable to end-users, because it gives users more control, and makes the capture process more visible, than server-side logging. Secondly, to demonstrate several new ways in which this data, once captured, can be applied to help users in understanding their own browsing habits and activities, and to facilitate social discovery and real-time awareness.

## Related Work

Eyebrowse incorporates features of both time-management and social life-tracking tools. Much like activity-tracking time-management systems, such as Slife[2], Eyebrowse tracks the user's web browsing activity and produces easy-to-read statistical summary visualizations. Also like life-tracking sites such as Last.fm[3] and Google Latitude[4], Eyebrowse shares that information publicly in near-real-time. Finally, like other predictive recommenders and social website tracking services (WebWatch[1] and Nebul.us[5]), Eyebrowse uses the data collected on the user's past browsing history to recommend web sites; however it is quite unique that it uses context (previously viewed sites, location, time of day and day of week) rather than web content, in performing the recommendation. We postpone discussion of the recommendation algorithm for later work.

## Eyebrowse

Eyebrowse consists of a Firefox add-on and a website. The add-on component, when installed, introduces a small icon in the bottom-right corner of the browser UI. This icon indicates whether the current site being visited is being logged and shared, and reveals the main control panel when clicked. This

[1] such as Google Analytics - google.com/analytics, and Alexa - www.alexa.com
[2] http://www.slifelabs.com/
[3] http://www.last.fm/
[4] http://www.google.com/latitude
[5] http://nebul.us/

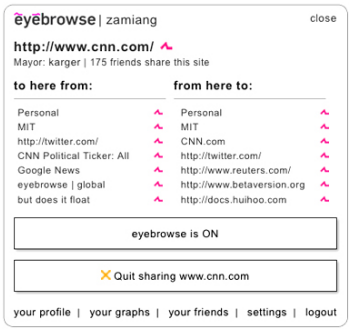


**Figure 3:** View of the Eyebrowse plugin for a user on cnn.com showing socially and contextually related information.

panel (Figure 3) allows quick access to all of Eyebrowse's features, including the visualization gallery, a thumbnail version of statistics related to the page being viewed (Figure 2), as well as buttons for controlling sharing.


**Figure 1:** (a) Top 20 URLs for day of week and time of day; (b) Timeline of pages visited over the course of 1 week (c) Timeline over 20 days.

After signing up and installing the plugin, the user is brought to their settings page where they can initially populate their "whitelist", the list which controls what sites are logged by the system. Sites can be later added by clicking on a button in Eyebrowse's panel. During testing, we experimented with other access control schemes, including blacklists (i.e., logging all URLs except those in the list are logged), but found whitelisting less confusing. When active, the plugin records the page URL, title, and non-idle time spent viewing the page, and sends this information to the Eyebrowse server, which collects this information and updates statistics pertaining to visualizations.

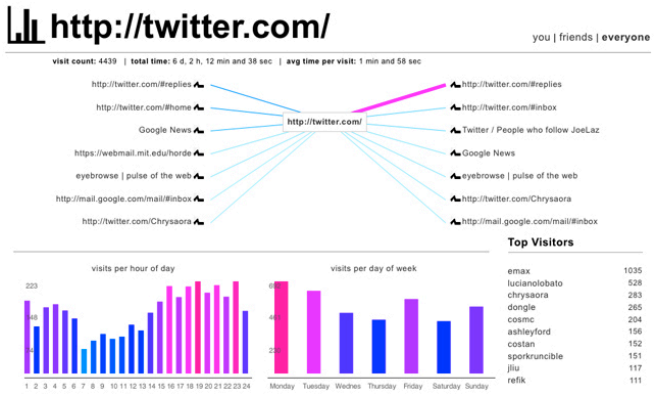


**Figure 2:** Statistics about a site/page, including pages viewed immediately before and after, common times accessed, and top visitors.

Once a site added to the whitelist to be shared, it is included in a on the user's profile page (Figure 4) and statistical summary graphs page (Figure 1). When visiting their own profile page, a user is presented with a view of the last 20 URLs they visited and a week 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 further exploration through their timeline graph (Figure 1b) and "20 days" graph (Figure 1c). With "day by day" (Figure 1a), users can view daily activity patterns and most visited

URLs. Colors are assigned using edit distance such that color similarity correlates to URL textual similarity.

When viewing others' profiles, the graphs facilitate social discovery by showing recent activity, general overviews of how users spend their time online, and web viewing trends over the past week. Users can directly compare themselves to others by through the "you minus me" or "me minus you" visualizations, which visualize areas of overlap and differences in two individuals' aggregate browsing statistics.

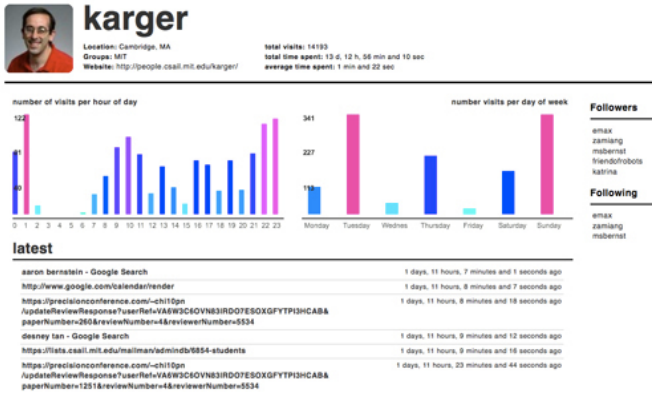


**Figure 4:** A user's profile page, showing a most recent and a macroscopic view of the users browsing habits.

## Initial Study

Prior to developing Eyebrowse, our colleagues mentioned having self-reflection needs surrounding web browsing activity, including wanting to *"find out how much time I waste on a few websites"* and *"figure out how much time I spent on certain projects."*. As revealed in the next sections, however, social uses were ultimately perceived as more useful than Eyebrowse's visualizations of self.

We surveyed 13 out of 121 active users of the system (defined as having at least 3 sites shared in their whitelists and 100 accesses logged apiece). The following sections comprise a summary of survey results and analysis of use by all active users.

*Social browsing versus self reflection*

We asked users to rate the usefulness of each element of Eyebrowse on a 5 point scale. While, on average, users reported finding Eyebrowse equally useful for social browsing and self-reflection, users reported finding the views containing information about others (e.g. social) more useful than views containing their own. Among the visualizations of the user's own activity, the list comparing top 25 URLs between weeks was reported to be the most helpful (with 6/13 of users reporting positively). When asked if Eyebrowse was useful for discovering new people, 9/13 responded with a 1 (not really) or 2 on 7-point scale. However, 6/13 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 3/13 of users found the "page stats" page useful, the lowest of any feature, the plugin view version of this visualization (which displayed the same information with less detail) was the most positively received (5/13). We believe this indicates that social browsing is most useful when contextually relevant.

*Whitelisting, exposure and privacy*

Users had an average of 62 domains in their whitelists (min:3, median:18, max:401) and logged an mean 3719 instances of browsing (min:101, median:1669, max: 36889). The most commonly whitelisted sites were Wikipedia and Twitter; social networking sites were considerably further down the list, with Facebook and Gmail coming at the 12th and 13th place, respectively.

The fact that Eyebrowse explicitly requires users to opt-in to sharing each unique web site led us to expect that users would rarely revise their recorded browsing history or remove elements from their whitelists. However, 69% (83) users removed entries from their whitelists at least once, while 21% (25) removed at least one record from their browsing history. Surveyed users who removed entries reported changing their minds after discovering that Eyebrowse logged unexpectedly informative page titles. This is a likely explanation for why e-mail services like Gmail and social networking sites like Facebook were shared less often; titles of pages accessed on these sites often contained information such as private e-mail subject lines, and the names of profiles and events the user viewed. Comments in our survey expressed a sensitivity to this information: *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.*

Tufecki reported in a study of online social networking sites [2] that users who were more concerned about online exposure and privacy tended to choose non-easily identifiable pseudonyms instead of their real names. To determine what percentage of Eyebrowse users may have deliberately chosen usernames to obscure their identity, we searched the usernames of each of the 121 users using a popular search engine, to see if we could uniquely identify an individual's true name within the top 10 hits returned. This experiment revealed that 55% (67) users chose easily identifiable usernames (either containing their name directly or a uniquely identifying nickname), while the remaining 54 were not identifiable. Comparing these groups, although there was an observable difference between the number of distinct sites each group on their whitelist ($\mu_1 = 72$ vs $\mu_u = 48$) this difference was insignificant; however, the easily identifiable group logged significantly more page views ($\mu_i = 4839$ vs $\mu_u = 2438$, $p < 0.02$). This may indicate that those users who chose identifiable names used the system longer, while perhaps those who chose to go incognito were more concerned about their degree of exposure and did not use the system as much.

When asked "What privacy concerns did you have about Eyebrowse?", 11/13 users reported social- or work-related privacy concerns; several comments indicated a desire to not have the intent behind their web browsing activity misinterpreted. For example, one user wrote: *I was mostly concerned that people would be curious as to why I was looking up certain individuals [or information] and draw the wrong conclusions.* Thus, the cause of privacy concern for these individuals surrounded not the disclosure of the subject/content of viewed pages, but the potentially damaging impact that the act of viewing these pages might have on their reputation.

*Public corpus*

As there is substantial need for activity datasets in personalization and adaptive systems research (surrounding, for example, recommender systems, information retrieval, context-aware computing), we made all data collected by Eyebrowse available to the public. During our initial deployment, we collected 500,000+ page views, which are now available (along with updated nightly snapshots) in various formats (CSV, XML). We hope that, like the datasets provided by *data.gov*, Eyebrowse data will encourage researchers to build new tools to look at their browsing data.

## Conclusion & Future Work

Based on upon the substantial uptake and continued use of our system, we are encouraged to develop Eyebrowse to further understand potential needs and implications

surrounding the capture, archiving and sharing of high fidelity logs of longitudinal web browsing activity.

For example, we wish to better understand needs surrounding self-understanding and reflection on past activity. Despite strong initial interest in being able to reflect upon activity, self-visualizations were rated less useful in Eyebrowse than social features. After using the system for several weeks, one user commented about about how he felt when viewing his own browsing history:

> I think I learned a little bit about myself [...] but there is a part of me that doesn't WANT to know how much time I spend on certain sites (e.g., Facebook)! It's a secret escape from what I'm *supposed* to be doing; if I learned that I spent an hour on it every day, I'd feel bad about it and obligated to try and change my behavior.

Despite the aversion expressed by this particular user towards knowing how exactly he spent his time online, we hypothesize that the visualizations of self-data in Eyebrowse might be more informative and raise less concern if they were more complete, and kept entirely private, so that users could use the system as a memory prosthesis without worrying about external reputation-related concerns.

Similarly, we would like to investigate whether having more levels of exposure, that would let users, for example, control which of their friend(s) saw particular traffic would change users' willingness to share sites. Unfortunately introducing such intermediate levels interferes with our goal of also creating a public corpus for web browsing research, because kept logs will no longer be necessarily public.

Several users mentioned "not having enough time" to peruse the visualizations, and felt that they were not getting much out of use of the system as a result. Thus we are considering ways that information captured by Eyebrowse could be placed strategically such that it could be used *in situ* while in the course of regular browsing activity. To this end, we added a "new tab" page which appears when the user opens a new tab that displays the user's most likely destinations (predicted given the user's location, recent browsing history and date/time). We wish to extend this sort of just-in-time display of information to aspects of reflection, reminding and discovery.

Finally, we would like to gain a better understanding of perceptions and privacy risks surrounding our approach to voluntary, opt-in client-side logging for collecting datasets for research. Given that the most prominent attempts by companies to release "anonymized" datasets inadvertently compromised individual users' privacy, which, further discouraged other holders of such data from releasing their datasets, we feel that this is a very important problem to address. Thus, we would like to explore better approaches that let users selectively volunteer their data as they please. We feel that many questions remained unanswered however, such as whether end-users will be able to make appropriate judgements surrounding their own personal data to achieve their desired level of privacy. Thus, we plan to continue to study users' perceptions of the data collection process to understand concerns and possible solutions.

## References


[1] C. E. Polk. Webwatch: visualizing web page histories and social monitoring habits. In *CHI*, pages 393–394, New York, NY, USA, 2001. ACM.

[2] Z. Tufekci. Can you see me now? audience and disclosure regulation in online social network sites. *Bulletin of Science Technology Society*, 28(1):20–36, February 2008.