
Reimagining Web Activity Tracking for Social Applications

Amy X. Zhang
MIT CSAIL
Cambridge, MA 02138, USA
axz@mit.edu

Joshua Blum
MIT CSAIL
Cambridge, MA 02138, USA
joshblum@mit.edu

David Karger
MIT CSAIL
Cambridge, MA 02138, USA
karger@mit.edu

Paste the appropriate copyright statement here. ACM now supports three different copyright statements:

- ACM copyright: ACM holds the copyright on the work. This is the historical approach.
- License: The author(s) retain copyright, but ACM receives an exclusive publication license.
- Open Access: The author(s) wish to pay for the work to be open access. The additional fee must be paid to ACM.

This text field is large enough to hold the appropriate release statement assuming it is single spaced in a sans-serif 7 point font.

Every submission will be assigned their own unique DOI string to be included here.

Abstract

Today, browsing the web generally requires providing one's web activity to corporations without consent or sometimes even awareness. While corporations that gather this data benefit greatly, very little of the data is accessible to the people that generate it for their benefit. Taking inspiration from social media systems, we reimagine a mode of web activity tracking where people share aspects of their browsing history to others in order to accrue individual and social benefits. We present ongoing work exploring privacy design and social applications surrounding this data, and suggest further research for the HCI community to consider.

Author Keywords

web browsing; privacy; web tracking; surveillance; activity traces; social media; web analytics;

ACM Classification Keywords

H.5.3 [Group and Organization Interfaces]: Web-based interaction; Computer-supported cooperative work

Introduction

The current state of activity tracking on the web is problematic in a number of ways. First, a large amount of people's browsing activity is collected in the databases of many corporations. They have been gathered through many different means but often without users' consent or even awareness,

so that users must go through a great deal of effort to resist being tracked. Even if we trust that well-known organizations would use web activity data responsibly, they and lesser known corporations may still be vulnerable to data breaches. Additionally, such organizations collect data indiscriminantly, gathering users' highly sensitive web activity along with more mundane visits.

On the other hand, researchers and open-source developers are at a disadvantage compared to corporations because they do not have access to such data to build applications or conduct research. Oftentimes they rely on public, anonymized datasets or conduct small-scale gathering of private data from participants. However, the first strategy can be disastrous as such data can be de-anonymized, as in the case of the AOL Query Log Dataset. The second strategy can be difficult to scale, and the data cannot be shared with others.

Finally, that many entities go to such trouble to gather this data underscores how valuable it is. Yet in the end, very little of this valuable data is accessible to the end users who generate it. In contrast to this involuntarily-shared and generally inaccessible data, social media tools like Facebook and Twitter allow users to voluntarily share details about their whereabouts, interests, emotions, and other aspects of their day-to-day life, for purposes such as building a public persona and interacting with friends.

If users could also publish chosen parts of their web activity traces, then instead of only benefiting organizations, shared data could directly benefit its sharers. The information could provide new social opportunities that users could leverage, much as with social media. Furthermore, a successful model where users choose to share their data pushes back against the current assumption that the only way to collect browsing data is to take it surreptitiously.

In prior work, we conduct interviews and user studies to explore potential benefits of web activity sharing [1]. In that work, we describe a prototype system, Eyebrowse¹, that uses a real-time global web log to let people “bump into” their friends on the web, discuss what they are seeing in real time, leave marks of their presence for others to see, and examine and visualize visits in aggregate. Going forward, there are many more opportunities for the use of such data for personal and social benefits. We invite the community to research and build new applications of their own from such data, and to that end, we release the data from our application in an API².

These potential benefits also potentially incur a cost in lost privacy. In our prior work, we explored domain-level whitelisting in user studies as well as interviewed about people's privacy preferences. However, more work is needed in this area to understand how to allow the sharing of such data while preserving users' desired level of privacy.

While there is widespread revulsion for the *involuntary surveillance* of web activity imposed by corporations and governments for their own benefit, we are interested in exploring the *voluntary sharing* people may wish to engage in to benefit themselves, their friends, or society. We have found little work considering cost-benefit tradeoffs around letting end-users leverage the global web trace. We believe this is an exciting opportunity for the HCI community to reimagine the state of web activity tracking today.

REFERENCES

1. Amy X. Zhang, Joshua Blum, and David Karger. 2016. Opportunities and Challenges Around a Tool for Social and Public Web Activity Tracking. In *Proc. of CSCW*.

¹<http://eyebrowse.csail.mit.edu>

²http://eyebrowse.csail.mit.edu/api_docs