Selected Papers of AoIR 2016:
The 17th Annual Conference of the
Association of Internet Researchers
Berlin, Germany / 5-8 October 2016

USING WEB BROWSING HISTORIES TO FACILITATE MULTI-METHOD RESEARCH

Ericka Menchen-Trevino
American University

Extended Abstract

If someone has any relationship at all to their web browsing history, it is probably best summarized in one word, “delete.” All major web browsers keep a log of the websites they access. This is needed for the normal functioning of the “back” button that enables users to return to previous sites, and to show users which links they have visited on a web page by displaying visited links in a different color, as well as auto-filling the URL of a recently visited site when partially typed. Although the functions browsing histories enable are quite banal, these logs are perceived as having a high potential for embarrassing the user if publically exposed (Bogart, 2015). It is likely that many a romantic relationship has ended after one partner combed through the other’s browser history, and certainly the employment of some has been jeopardized by browsing history log analysis as well.

Despite the sensitive nature of this data, if given the ability to explore and better understand their own data and remove what they choose, some web users may opt-in to sharing this data with researchers they trust. In fact, the informed consent process could demystify this often poorly understood source of information and give individuals better tools for understanding and controlling their own browsing data logs (Menchen-Trevino, 2016). Furthermore, this user-focused data crosses between platforms where data is often siloed, such as different social media sites and web services making it particularly useful for extending platform-based studies of digital traces (Menchen-Trevino, 2013).

The Herodotus (name changed for anonymous review) project has developed an open-source web browser extension with the goal of informing users of the insights available in their browsing history data through visualizations and analytics. The extension allows users to opt-in to share their data with a research project. Participants not only submit their browsing data but are immediately directed to a survey. This survey can inquire about the context of the browsing they submitted (e.g. “Does more than one person use this web browser?”), as well as provide questions relevant to the study. The survey also allows participants to be compensated for their participation and enables participants to

Suggested Citation (APA): Menchen-Trevino, E. (2016, October 5-8). *Using Web Browsing Histories To Facilitate Multi-Method Research*. Paper presented at AoIR 2016: The 17th Annual Meeting of the Association of Internet Researchers. Berlin, Germany: AoIR. Retrieved from http://spir.aoir.org.

choose whether they would allow the researcher to contact them for a follow-up interview via online video. Critically, this approach allows for the collection of both observed and self-reported information about web browsing. This can help researchers assess the accuracy of self-reports, but perhaps more importantly it can address questions about the impact (or lack thereof) of attitudes on behaviors in a more valid way than self-reported behavior data can provide.

This approach extends the methodology the author developed in 2010 using a proxy server to collect real-world web browsing data with the informed consent of participants (Menchen-Trevino & Karr, 2012; Menchen-Trevino, 2012). Browsing histories provide similar data to the proxy system and the data is created before a participant agrees to be part of a study. The participant can be better informed about what data they are providing since they can explore it before deciding to participate. The exploration process itself has been developed. Interactive data visualizations enable informed consent and also provide knowledge that is a direct benefit to participants about their own web browsing habits and behaviors (see Figures 1-4). Each visualization can focus on a specific time period using a calendar feature for begin and end date.

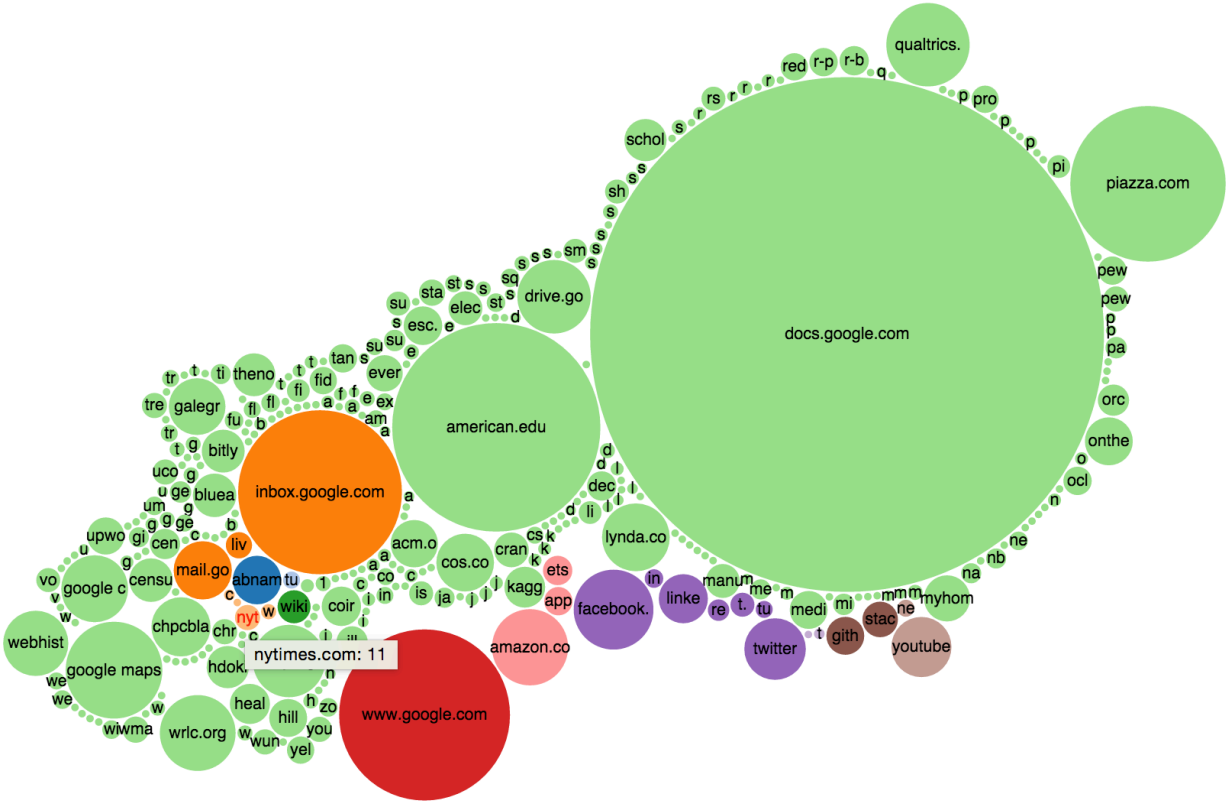

Figure 1. Websites Visited: Circle size represents the number of visits to a domain. Tooltip identifies the domain name and number of visits.

Figure 2. Search Words: Word size increases when the word is used in different search terms. The tooltip shows the search terms in which the word appears (here the search terms containing javascript are shown).

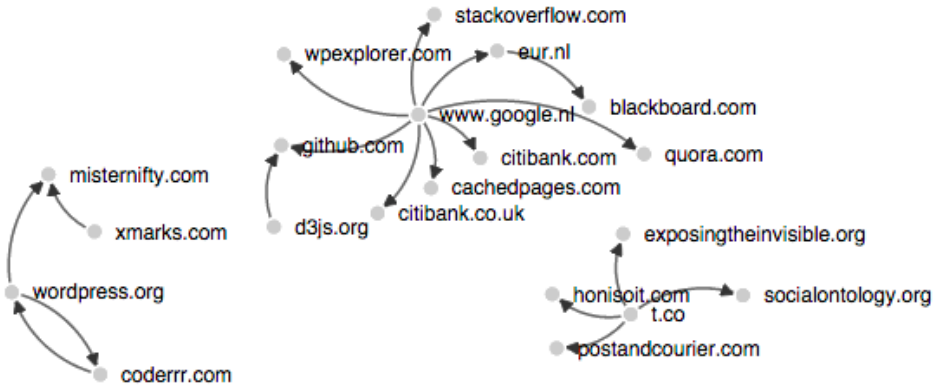

Figure 3. Network: Arrows connect domains where the user browsed from the origin to the destination domain. Hovering over the node or label allows the user to reposition the node manually.

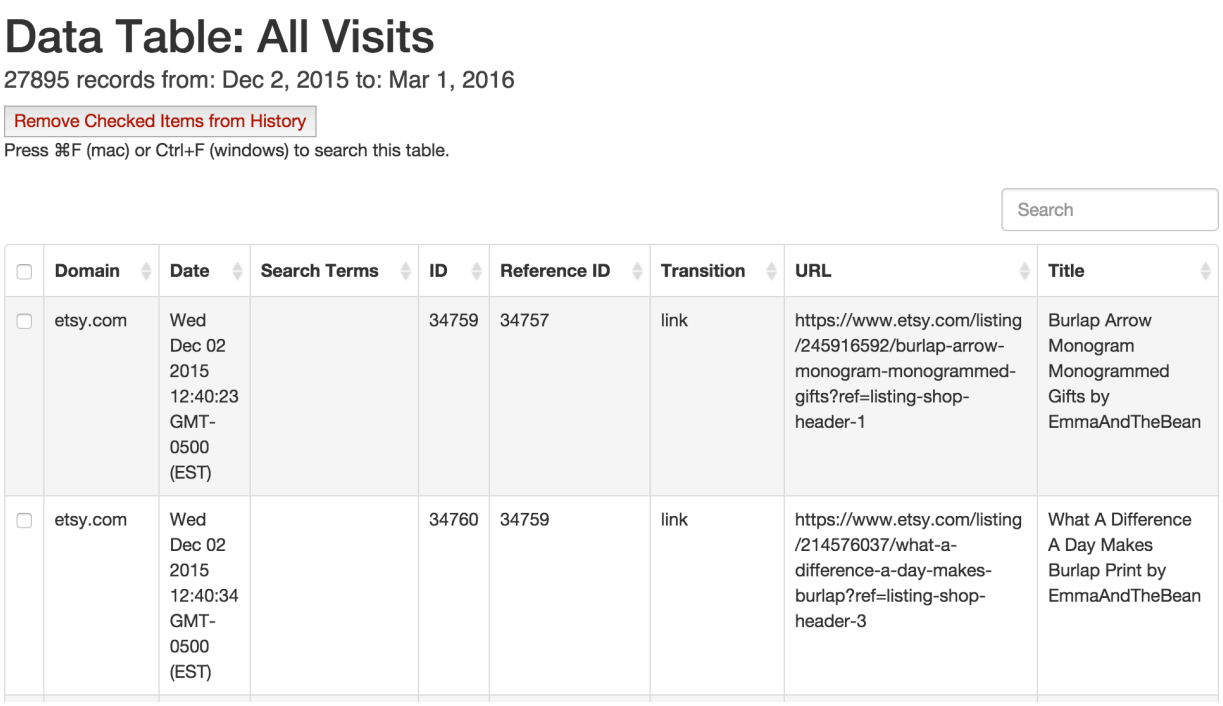

Figure 4. Data Table with two visit records: Checking the box in the remove field allows users to remove records by pressing the “Remove Checked Items from History” button. Columns can be sorted by any field, and the search field quickly narrows the results to items containing the search term in any field.

Based on the research questions of the investigators, the browsing history and the survey data may become the basis for interviews regarding the context of web browsing and its role in the topic of interest. That is, the Herodotus project can support research focusing on web history data itself, the relationship between attitudes (measured on a survey) and web history data, or questions about the role of the web in broader social processes or in the context of individuals or communities. A key limitation for those interested in studying large populations is that some people who are concerned about sharing private data will not participate, and perhaps these users browse the web differently than others. These same people may also be less likely to complete telephone surveys, or research of any kind. As refusal rates rise this is a problem faced by all researchers who need to partner with participants to complete their studies. An important benefit of this approach is that it may be possible to get anonymized or summarized browsing data for a large population of users such that the differences between volunteers and non-volunteers can be observed.

In an interview context the visualizations aid participant recall. The participants can validate the accuracy of the data using the visuals and may identify missing data. Researchers can define a list of websites about their topic of interest to focus on and highlight these websites within the visualizations, and/or they can ask participants to identify websites they use for particular purposes.

This work fits Dubois and Ford’s definition of a trace interview (2015) where participants are provided data visualizations as part of a qualitative interview process. They point out the complexity and challenge of understanding visualizations of digital traces and the importance of introducing the visualizations clearly, as well as the kind of rich contextual

information that can be gained from aiding participant recall in this way. This project will build on the work of Dubious and Ford by using a different and more individual-focused form of data.

Currently the Herodotus browser extension is undergoing beta testing to facilitate wider deployment. Nine preliminary interviews have been conducted with young adults in the Netherlands, and additional interviews will be conducted in the U.S. in the coming weeks to refine the utility of the visualizations for interviewing purposes. The final paper will report the methodological findings regarding the browsing data, surveys, and interviews.

References

Bogart, N. (2015, August 20). Bank statements, browser history “most embarrassing” info to leak online: survey | Globalnews.ca. Retrieved March 1, 2016, from http://globalnews.ca/news/2176339/bank-statements-browser-history-most-embarrassing-info-to-leak-online-survey/

Dubois, E., & Ford, H. (2015). Trace Interviews: An Actor-Centered Approach. *International Journal of Communication*; Vol 9. Retrieved from http://ijoc.org/index.php/ijoc/article/view/3378/1416

Menchen-Trevino, E. (2012, December). *Partisans and Dropouts?: News Filtering in the Contemporary Media Environment*. Northwestern University, Evanston, Illinois.

Menchen-Trevino, E. (2013). Collecting Vertical Trace Data: Big Possibilities and Big Challenges for Multi-method Research. *Policy & Internet*, *5*(3), 328–339. https://doi.org/10.1002/1944-2866.POI336

Menchen-Trevino, E. (2016). Web Historian: Enabling multi-method and independent research with real-world web browsing history data. Presented at the iConference, Philadelphia: IDEALS. Retrieved from http://hdl.handle.net/2142/89388

Menchen-Trevino, E., & Karr, C. (2012). Researching real-world Web use with Roxy: Collecting observational Web data with informed consent. *Journal of Information Technology & Politics*, *9*(3), 254–68. https://doi.org/10.1080/19331681.2012.664966