

Herbert van de Sompel and Martin Klein (Los Alamos National Laboratory), *The Memento Infrastructure to Support Research Using Web Archive Collections*

The Internet Archive pioneered web archiving and remains by far the largest collection of archived web resources, worldwide. However, over the last decade, the landscape of openly available web archives has significantly diversified. We now observe national libraries and archives, for-profit as well as non-profit organizations, academic libraries, and even motivated individuals creating archival collections and hence contributing to recording web history.

Despite this variety, research that requires archived web resources commonly only uses the Internet Archive's collection. Given its temporal span and vast size, it is understandable that it is the first point of reference for scholars. However, it has been shown on several occasions that the utilization of multiple web archives is beneficial for research that requires archived resources (e.g., [1]).

In this presentation we will explain the Memento protocol [2] that specifies interoperable access to archived web resources and is meanwhile supported by all major public web archives. We will provide an overview of the Memento aggregator infrastructure that currently supports simultaneous access to archived resources held by over 20 web archives around the world. We will demonstrate cross-archive tools aimed for manual exploration (Memento for Chrome [3], Memento for FireFox [4], TimeTravel [5]) as well as for machine use at scale (TimeTravel APIs [6]).

To illustrate the benefit and feasibility of using multiple archives for research purposes, we will provide insights into research problems that we have explored using the Memento infrastructure and detail the methodologies we have used. For example, we will highlight our recent publications that assessed the extent of "Reference Rot" in scholarly articles ([7] and [8]) as well as our exploratory effort to apply focused crawling techniques across web archives ([9]).

This presentation aims at widening the horizon of workshop attendees. We will demonstrate that, by utilizing the Memento infrastructure, leveraging multiple web archives isn't harder than leveraging a single one. We will show that, when developing research software based on the Memento protocol, software created for exploring one Memento-compliant archive can be used without changes to explore others. Since consulting a broad variety of archival collections clearly has the potential to make research findings more robust, we hope to encourage workshop attendees to start exploring the diverse web archiving landscape.

[1] <https://doi.org/10.1145/3041656>

[2] <https://tools.ietf.org/html/rfc7089>

[3] <http://bit.ly/memento-for-chrome>

[4] <http://bit.ly/memento-for-firefox>

[5] <http://timetravel.mementoweb.org/>

[6] <http://timetravel.mementoweb.org/guide/api/>

[7] <https://doi.org/10.1371/journal.pone.0115253>

[8] <https://doi.org/10.1371/journal.pone.0167475>

[9] <https://doi.org/10.1145/3201064.3201085>