Computer Science Faculty Publications

Computer Science

2020

MementoEmbed and Raintale for Web Archive Storytelling

Shawn M. Jones

Martin Klein

Michele C. Weigle
Old Dominion University

Michael L. Nelson
Old Dominion University

Original Publication Citation

MementoEmbed and Raintale for Web Archive Storytelling

Shawn M. Jones, Martin Klein
{smjones,mklein}@lanl.gov
Los Alamos National Laboratory, Los Alamos, NM

Michele C. Weigle, Michael L. Nelson
{mweigle,mln}@cs.odu.edu
Old Dominion University, Norfolk, VA

ABSTRACT

For traditional library collections, archivists can select a representative sample from a collection and display it in a featured physical or digital library space. Web archive collections may consist of thousands of archived pages, or mementos. How should an archivist display this sample to drive visitors to their collection? Search engines and social media platforms often represent web pages as cards consisting of text snippets, titles, and images. Web storytelling is a popular method for grouping these cards in order to summarize a topic. Unfortunately, social media platforms are not archive-aware and fail to consistently create a good experience for mementos. They also allow no UI alterations for their cards. Thus, we created MementoEmbed to generate cards for individual mementos and Raintale for creating entire stories that archivists can export to a variety of formats.

KEYWORDS

web archives, memento, storytelling, visualization, summarization

1 INTRODUCTION

Trying to understand the differences between web archive collections can be onerous. Thousands of collections exist [7], collections can contain thousands of documents, and many collections contain little metadata to assist the user in understanding their contents [8]. How can an archivist display a sample of a collection in order to drive visitors to their collection or provide insight into their archived pages?

Search engines and social media platforms have settled on the card visualization paradigm, making it familiar to most users. Web storytelling is a popular method for grouping these cards to summarize a topic, as demonstrated by tools such as Storify. For this reason, AlNoamany et al. [1] made Storify the visualization target of their web archive collection summaries. Because Storify shut down in 2018 [3], we evaluated fifty alternative tools, such as Facebook, Pinboard, Instagram, Sutori, and Paper.li [4]. We found that they are not reliable for producing cards from mementos. Thus we developed MementoEmbed [5], an archive-aware service that can generate different surrogates [2] for a given memento. Currently supported surrogates include social cards, browser thumbnails (screenshots) [9], word clouds, and animated GIFs of the top ranked images and sentences. MementoEmbed's cards appropriately attribute content to a memento's original resource separately from the archive, including both the original domain and its favicon from the memento's time period, as well as providing a striking image, a text snippet and a title. MementoEmbed provides an extensive API that helps machine clients request specific information about a memento. Raintale [6] leverages this API to generate complete stories containing the surrogates of many different mementos.

2 THE MEMENTOEMBED-RAINTALE ARCHITECTURE FOR STORYTELLING

Figure 1 demonstrates the relationship between MementoEmbed and Raintale. In step 1, the user provides a template and a list of URI-Ms to Raintale. In step 2, Raintale records all template variables. For each provided URI-M, Raintale consults MementoEmbed's API for the value of each variable in the corresponding memento. In step 3, MementoEmbed downloads the memento from its web archive and performs natural language processing, image analysis, or extracts information via the Memento Protocol [10], as appropriate to the API request. In step 4, Raintale consolidates the data from these API responses and renders the template with the gathered data, producing a story constructed from surrogates and other supplied content (e.g., story title, collection metadata). Raintale can produce many output formats, including HTML (Figure 2), Markdown, MediaWiki, Jekyll, and Twitter Threads (Figure 3).

3 CONCLUSIONS

We introduced MementoEmbed for generating surrogates for single mementos and Raintale for generating complete stories of memento sets. We envision Raintale and MementoEmbed to be critical components for summarizing collections of archived web pages through visualizations familiar to general users. We developed MementoEmbed so that its API is easily usable by machine clients. Raintale lends itself to incorporation into existing automated archiving workflows. Archivists can leverage this form of storytelling to highlight a specific subset of mementos from a collection. They can advertise their holdings, feature specific perspectives, focus on individual mementos, or help users decide if a collection meets their needs.

REFERENCES

[1] Yasmin AlNoamany, Michele C. Weigle, and Michael L. Nelson. 2017. Generating Stories From Archived Collections. In *WebSci 2017*. Troy, New York, 309–318. https://doi.org/10.1145/3091478.3091508

[2] Robert Capra, Jaime Arguello, and Falk Scholer. 2013. Augmenting web search surrogates with images. In *CIKM 2013*. San Francisco, California, USA, 399–408. https://doi.org/10.1145/2505515.2505714

[3] Shawn M. Jones. 2017. Storify Will Be Gone Soon, So How Do We Preserve The Stories? https://ws-dl.blogspot.com/2017/12/2017-12-14-storify-will-be-gone-soon-so.html

[4] Shawn M. Jones. 2017. Where Can We Post Stories Summarizing Web Archive Collections? http://ws-dl.blogspot.com/2017/08/2017-08-11-where-can-we-post-stories.html

[5] Shawn M. Jones. 2018. A Preview of MementoEmbed: Embeddable Surrogates for Archived Web Pages. https://ws-dl.blogspot.com/2018/08/2018-08-01-preview-of-mementoembed.html

[6] Shawn M. Jones. 2019. Raintale – A Storytelling Tool For Web Archives. https://ws-dl.blogspot.com/2019/07/2019-07-11-raintale-storytelling-tool.html

[7] Shawn M Jones, Alexander Nwala, Michele C Weigle, and Michael L Nelson. 2018. The Many Shapes of Archive-It. In *iPres 2018*. Boston, Massachusetts, USA, 1–10. https://doi.org/10.17605/OSF.IO/EV42P

[8] Shawn M. Jones, Michele C. Weigle, and Michael L. Nelson. 2019. Social Cards Probably Provide For Better Understanding Of Web Archive Collections. In *CIKM 2020*. Beijing, China, 2023–2032. https://doi.org/10.1145/3357384.3358039

[9] Theodorich Kopetzky and Max Mühlhäuser. 1999. Visual preview for link traversal on the World Wide Web. *Computer Networks* 31, 11-16 (May 1999), 1525–1532.
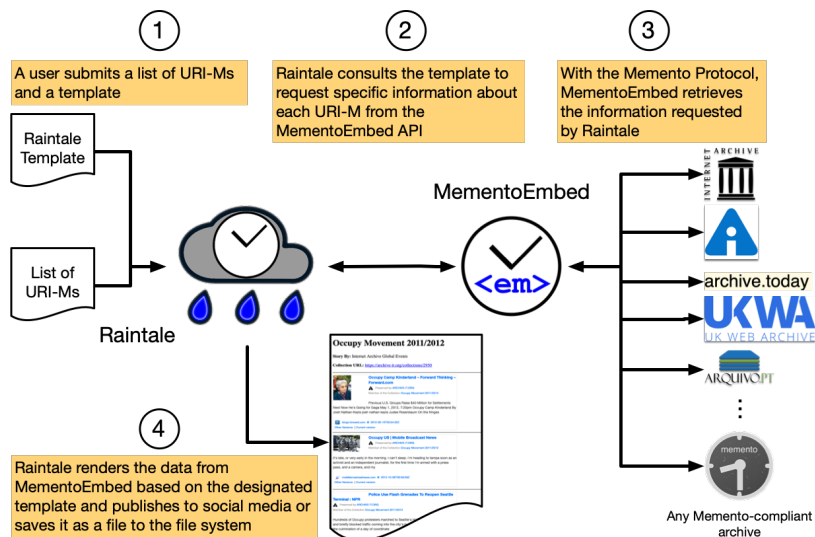
Figure 1: The MementoEmbed-Raintale Architecture for Storytelling

Figure 2: Raintale can render a multiple mementos as an HTML story of cards.

https://doi.org/10.1016/S1389-1286(99)00050-X

[10] Herbert Van de Sompel, Michael Nelson, and Robert Sanderson. 2013. RFC 7089 - HTTP Framework for Time-Based Access to Resource States – Memento. https://tools.ietf.org/html/rfc7089

Figure 3: Raintale can render multiple mementos as a Twitter thread.