

2020

TMVis: Visualizing Webpage Changes Over Time

Abigail Mabe
Old Dominion University

Dhruv Patel
Old Dominion University

Maheedhar Gunnam
Old Dominion University

Surbhi Shankar
Old Dominion University

Mat Kelly

See next page for additional authors

Follow this and additional works at: https://digitalcommons.odu.edu/computerscience_fac_pubs



Part of the [Computer Sciences Commons](#)

Original Publication Citation

Mabe, A., Patel, D., Gunnam, M., Shankar, S., Kelly, M., Alam, S., Nelson, M. L., & Weigle, M. C. (2020). TMVis: Visualizing webpage changes over time. Web Archiving & Digital Libraries Workshop 2020, Virtual, August 5, 2020.

This Conference Paper is brought to you for free and open access by the Computer Science at ODU Digital Commons. It has been accepted for inclusion in Computer Science Faculty Publications by an authorized administrator of ODU Digital Commons. For more information, please contact digitalcommons@odu.edu.

Authors

Abigail Mabe, Dhruv Patel, Maheedhar Gunnam, Surbhi Shankar, Mat Kelly, Sawood Alam, Michael L. Nelson, and Michele C. Weigle

TMVis: Visualizing Webpage Changes Over Time

Abigail Mabe, Dhruv Patel,
Maheedhar Gunnam, Surbhi
Shankar
Department of Computer Science
Old Dominion University
Norfolk, VA 23529 USA
{amabe002,dpate006,mgunn001,
sshan001}@odu.edu

Mat Kelly
College of Computing & Informatics
Drexel University
Philadelphia, PA 19104 USA
mkelly@drexel.edu

Sawood Alam, Michael L.
Nelson, Michele C. Weigle
Department of Computer Science
Old Dominion University
Norfolk, VA 23529 USA
{salam,mln,mweigle}@cs.odu.edu

ABSTRACT

TMVis is a web service to provide visualizations of how individual webpages have changed over time. We leverage past research on summarizing collections of webpages with thumbnail-sized screenshots and on choosing a small number of representative archived webpages from a large collection. We offer four visualizations: Image Grid, Image Slider, Timeline, and Animated GIF. Embed codes for the Image Grid and Image Slider can be produced to include these visualizations on separate webpages. This tool can be used to allow scholars from various disciplines, as well as the general public, to explore the temporal nature of webpages.

1 INTRODUCTION

Users of web archives may be interested in a quick overview of the topic of a particular webpage in a collection or in observing how that webpage has changed over time. Many web archive interfaces provide only a textual list of archived versions, or *mementos*, which requires that each one be explored individually. This can be a tedious process as the number of mementos for each webpage grows. As a first step to address this problem, we have developed a web service, TMVis, providing visualizations that show how a single webpage has changed over time. We do this by using the list of mementos of a webpage (*i.e.*, a TimeMap) and choosing to display the screenshots, or thumbnails, of the most unique mementos, as a visual summarization.

In our previous work [1], we developed a technique to determine which mementos are useful to display. The algorithm compares the amount of change (based on SimHash) in the HTML of consecutive mementos and selects the most unique. This process is much more efficient, in both time and space, than generating all thumbnails and then performing image processing. We use this same process for TMVis. To improve the response time for large TimeMaps with over 1000 mementos, we first sample a maximum of 1000 mementos before continuing the selection process.

The web service is available at <http://tmvis.cs.odu.edu/>, and the source code is hosted on GitHub at <https://github.com/oduwsdl/tmvis>. We have published a tech report with a full description [2] and a blogpost¹ with more images and a video walkthrough.

2 TMVIS WEB SERVICE

The TMVis web service asks the user for a URI and allows them to choose to process TimeMaps from the Internet Archive or from a

collection in Archive-It². After clicking the “View TimeMap” button, the TimeMap is loaded and the user is presented with a histogram displaying the number of mementos available over time. The user may use the histogram to select a date range in the TimeMap to summarize, or they may choose to summarize the entire TimeMap.

After choosing the desired time range, the user clicks the “Calculate Unique” button. The service downloads the HTML source and computes the SimHash of the requested mementos. Then the SimHashes of the mementos are compared according to AlSum et al. [1]. We present the user with several options for the threshold of “uniqueness” so that they may create larger or smaller visualizations as desired. Once the user chooses their desired number of thumbnails and clicks the “Generate Thumbnails” button, the service uses Puppeteer³ to render each memento and capture a thumbnail screenshot. Once the thumbnails are captured for all the mementos, they are presented to the user with the four visualization widgets shown in Figures 1-3. The default view shows the Image Grid, however users can switch between tabs to view the other visualizations: Image Slider, Timeline, and Animated GIF.

Image Grid. The Image Grid shows the thumbnails of the representative mementos arranged in a left to right, top to bottom manner. Clicking on any thumbnail loads the source memento. On the top right of each thumbnail in the grid is a refresh button to allow users to regenerate the thumbnail if it appears incomplete and an ‘X’ to allow users to remove thumbnails from the visualizations. Figure 1 shows an example of the Image Grid for <http://www.odu.edu/>.

Image Slider. The Image Slider imitates the photo slider functionality used in Apple’s iPhoto⁴. By moving the cursor across the thumbnail image, the next thumbnail is displayed. As with the Image Grid, clicking on the thumbnail loads the source memento. The user can cycle through the thumbnails by clicking arrow buttons to the left and right of the slider. Figure 2 shows a static example of the Image Slider for <http://columbia.edu/cu/english/>.

Timeline. The Timeline view arranges the thumbnails according to the mementos’ datetimes (time of capture). The Timeline view includes zoom, next, previous, next unique, and previous unique buttons to allow the user to navigate between the unique and regular mementos. The unique mementos are represented with yellow stripes and the others are represented with gray stripes on the

¹<https://ws-dl.blogspot.com/2020/05/2020-05-21-visualizing-webpage-changes.html>

²TMVis can be expanded to support any Memento-compatible public archive or Memento aggregator.

³<https://developers.google.com/web/tools/puppeteer>

⁴<http://web.archive.org/web/20150101033528/http://apple.com/mac/iphoto/>

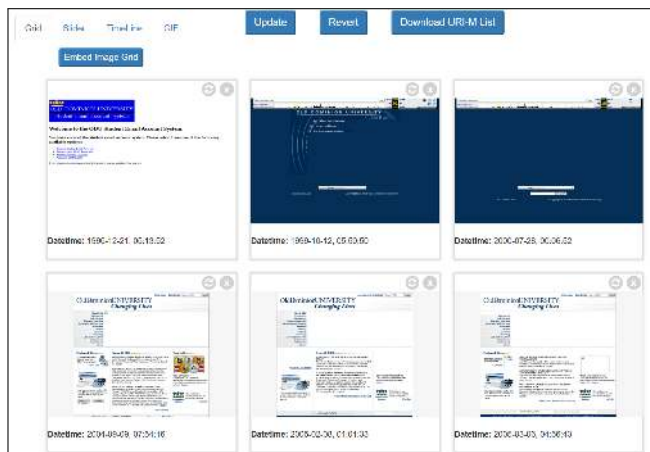


Figure 1: Image Grid for <http://www.odu.edu/>

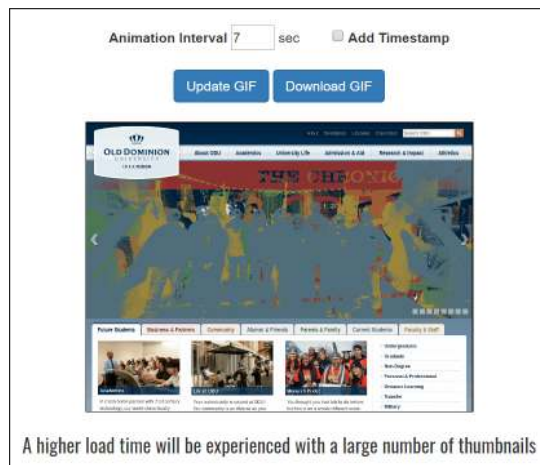


Figure 4: Animated GIF for <http://www.odu.edu/>

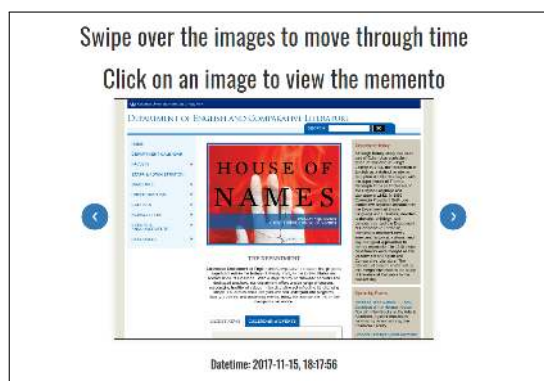


Figure 2: Image Slider for <http://columbia.edu/cu/english/>



Figure 3: Timeline view for <http://www.odu.edu/>

timeline. Non-unique mementos do not have a screenshot, but are represented by the thumbnail of the previous unique memento with an indication that this memento is “similar to” the previous unique memento. The Timeline view is based on Timeline Setter library⁵, developed by ProPublica. Figure 3 shows an example of the Timeline view for <http://www.odu.edu/>.

Animated GIF. The Animated GIF visualization, shown in Figure 4, is built from the thumbnails in the Image Grid using the GifShot

⁵<http://propublica.github.io/timeline-setter/>

library⁶. The user can include a timestamp on each thumbnail or adjust the time interval between each frame of the animation. This GIF can be downloaded by clicking the “Download GIF” button.

Embeddable Visualizations. We have also provided options for users to embed the Image Grid and Image Slider visualizations in their own webpages as well as an option to download the animated GIF for inclusion in other webpages. An example of embedding these elements in a simple webpage is provided at <https://ws-dl.cs.odu.edu/vis/tmvis/embed-examples.html>.

3 CONCLUSION

We have presented a description of TMVis, a web service that provides four visualizations of how a webpage changes through time. We compare the difference in the SimHashes of the HTML source of pairs of mementos to determine which set of mementos is the most unique. Then we render and capture thumbnail-sized screenshots of the chosen mementos. These are then displayed as an Image Grid, Image Slider, Timeline, and Animated GIF. We hope that these visualizations will just be the beginning and will provide a starting point for others to expand these types of offerings for users of web archives.

ACKNOWLEDGEMENTS

This work has been supported by a NEH/IMLS Digital Humanities Advancement Grant (HAA-256368-17). We are grateful for this support and for the input from our partners at the Frick Art Reference Library, New York Art Resources Consortium, and Columbia University Libraries.

REFERENCES

- [1] Ahmed AlSum and Michael L. Nelson. 2014. Thumbnail Summarization Techniques for Web Archives. In *Proceedings of the European Conference on Information Retrieval (ECIR)*. Amsterdam, 299–310.
- [2] Abigail Mabe, Dhruv Patel, Maheedhar Gunnam, Surbhi Shankar, Mat Kelly, Sawood Alam, Michael L. Nelson, and Michele C. Weigle. 2020. *Visualizing Webpage Changes Over Time*. Technical Report arXiv:2006.02487. <https://arxiv.org/abs/2006.02487>

⁶<https://yahoo.github.io/gifshot/>