

A Memento Web Browser for iOS

Heather Tweedy
Harding University
Computer Science Dept
Searcy, Arkansas, USA 72149
htweedy@harding.edu

Frank McCown
Harding University
Computer Science Dept
Searcy, Arkansas, USA 72149
fmccown@harding.edu

Michael L. Nelson
Old Dominion University
Computer Science Dept
Norfolk, Virginia, USA 23529
mln@cs.odu.edu

ABSTRACT

The Memento framework allows web browsers to request and view archived web pages in a transparent fashion. However, Memento is still in the early stages of adoption, and browser-plugins are often required to enable Memento support. We report on a new iOS app called the Memento Browser, a web browser that supports Memento and gives iPhone and iPad users transparent access to the world's largest web archives.

Categories and Subject Descriptors

H.3.7 [Information Storage and Retrieval]: Digital Libraries – Collection.

General Terms

Design.

Keywords

Mobile web, web archiving, web browser, Memento.

1. INTRODUCTION

The Memento framework provides transparent client access to archived web resources [7]. It introduces a modification to the HTTP protocol that allows a client to request a particular time version of web resource without the user having to navigate through various archives in search of the resource. The archived resources discovered with Memento are called *mementos*. A number of web archives have implemented Memento, including the open-source Wayback Machine software [8], but client support has mainly been limited to the Firefox browser running the MementoFox add-on [3][4].

We introduce a new iOS app called Memento Browser [2]. As pictured in Figure 1, the Memento Browser looks like the stock iOS browser except that it introduces a toolbar that allows the user to 1) view all the mementos available for a particular URL, 2) navigate to previous and next mementos, and 3) view the web page as it looks right now. The mementos are discovered using a Memento TimeGate running at Los Alamos National Laboratory (LANL) which aggregates archived URLs from Internet Archive, WebCite, UK Web Archive, and others. Figure 1 shows the Memento Browser displaying an archived web page from the Internet Archive. Memento Browser has also been implemented for Android; both browsers implement Memento's "simple" mode which we introduce in this poster. We also discuss some of the challenges viewing the Past Web on a mobile device.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

JCDL'13, July 22–26, 2013, Indianapolis, Indiana, USA.

ACM 978-1-4503-2077-1/13/07.



Figure 1. Memento Browser showing a cnn.com archived page from Internet Archive

2. MEMENTO CLIENT MODES

Web pages are normally composed of an HTML file and other supporting files: style sheets, JavaScript, images, etc. Most web archives do their best to archive all these resources at the same time, but this is technically challenging because web crawlers must be "polite" when crawling the web, meaning they must delay a few seconds between requests. Archiving a website can thus take hours, days, or weeks depending on its size. During this time resources might be modified, be moved to new URLs, or even disappear. Therefore, most archived websites experience various degrees of temporal incoherence [6].

When a client is using Memento to discover and display a web page from an earlier datetime T , it will request a Memento-enabled web server for the web page as it existed at time T . There are three HTTP requests involved in this request, and the details of this process can be found in [4]. The end result of these requests is that the client will have a URI where the archived resource closest to datetime T can be found. The resource may be on the web server or in some web archive. When the HTML for the page is retrieved, the client should make subsequent requests for the images, style sheets, etc.

using the same process (three HTTP requests) for each of the resources. In other words, datetime content negotiation should be performed for *all* components making up the web page. This strict application of Memento across all requests is what we call **thorough mode**.

To speed up the process of rendering a memento web page, a client may use **faster mode**. In this mode, the client uses Memento to discover the HTML for the web page, but it does not make Memento requests for all the embedded resources. Instead, it takes advantage of the fact that many web archives rewrite the URLs of the embedded resources so they point back to the archive. So the embedded resources are fetched as-is, and the returned resources are checked to ensure they are also mementos (by checking the link rel HTTP response headers or checking against a white list of archive URL patterns). If the archive does not rewrite the embedded URLs, they will point to the “live” web, and they will not be used to display the memento web page.

Although thorough mode requires three HTTP requests per resource, it is more likely to produce better approximations of historical web pages because the web page will be composed potentially from resources from multiple archives that are closest to the desired historical datetime. Faster mode requires fewer HTTP requests, but it may produce poorer approximations of historical web pages because the entire page is composed of resources coming from only one archive. It also requires the web archives producing the HTML to rewrite all embedded URLs. MementoFox supports both thorough and faster modes.

When developing Memento Browser for iOS and an earlier version for Android, we were forced to support a new mode that is somewhat like faster. We used the stock controls provided on both these platforms for rendering web pages. The iOS version uses the `UIWebView` control, and the Android version uses the `WebView` control. Neither of these controls support Memento, and they do not allow the programmer to trap the outgoing HTTP requests and incoming responses for embedded resources. The iOS app will therefore get a `TimeMap` using the LANL proxy server (discussed below) in order to discover all the available mementos for the web page’s URL that is currently being viewed. It will then display the memento matching the datetime chosen by the user.

So Memento Browser only uses Memento for discovering web page mementos, but it does not use Memento for making the requests for all embedded resources. We call this new mode of Memento support **simple mode**. Simple mode is just like fast mode except that the embedded resources are not checked to see if they are mementos. It is easy to implement but problematic when a web archive does not rewrite the URLs of embedded resources that point to the “live” web.

Memento is not yet widely implemented in web browsers and web servers, so it is not uncommon for Memento-enabled clients to use a proxy server which speaks Memento. Both MementoFox and Memento Browser use a proxy server at LANL which aggregates archived URLs from Internet Archive, WebCite, UK Web Archive, and others. If the proxy were asked for a web page as it existed at datetime T , it will reply back with a `TimeMap` listing the archived URLs that are closest to T from the list of supported archives. Because of the temporal coherence issue mentioned earlier, the web page’s images and style sheets closest to T might be found in different archives. In other words, a web page from datetime T

could be composed of HTML from Internet Archive, an image from WebCite, and a style sheet from UK Web Archive.

The value of having access to resources from a variety of archives can be demonstrated by a simple example [1]. Querying the LANL proxy for `http://www.cnn.com/ mementos` on January 30, 2013, reveals 17,188 archived copies. A large majority of these pages (98.5%) come from the Internet Archive, but there are only 77 total archived copies from 2012; they come from the UK National Archives (58.4%), Archiefweb (33.8%), and Archive-It! (7.8%).

3. PAST WEB ON A MOBILE DEVICE

Many websites will redirect a mobile browser like Memento Browser to a mobile-friendly website that uses a different URL. These mobile sites are usually much newer than their stationary counterparts, and they are less likely to be archived [5]. For example, there are only 185 total copies of `http://m.cnn.com/` in the Internet Archive compared to 17K for `http://www.cnn.com/`. Therefore the Memento Browser will sometimes discover far fewer mementos than MementoFox running on a desktop would. We are interested in allowing Memento Browser to see historical mobile pages and traditional web pages.

4. ACKNOWLEDGMENTS

We would like to thank Steve Baber for writing the initial version of Memento Browser for iOS. We also thank members of the Prototyping Team at the Research Library of the Los Alamos National Laboratory for their feedback on Memento Browser. This research was supported by the National Science Foundation (IIS 1008492 and 1009392) and the Library of Congress.

5. REFERENCES

- [1] Ainsworth, S., AlSum, A., SalahEldeen, H., Weigle, M., Nelson, M. L. 2012. How much of the Web is archived? In *Proceedings of the 11th annual international ACM/IEEE Joint Conference on Digital Libraries (JCDL '11)*. ACM, New York, NY, USA, 133-136.
- [2] Memento Browser for iOS and Android. <http://code.google.com/p/memento-browser/>
- [3] MementoFox Add-on. <https://addons.mozilla.org/en-US/firefox/addon/mementofox/>
- [4] Sanderson, R., Shankar, H., Ainsworth, S., McCown, F., Adams, S. 2011. Implementing time travel for the Web. *Code4Lib Journal*, Issue 13 (Apr S2011). <http://journal.code4lib.org/articles/4979>
- [5] Schneider, R., McCown, F. 2013. First steps in archiving the mobile web: automated discovery of mobile websites. *Proceedings of the 13th annual international ACM/IEEE Joint Conference on Digital Libraries (JCDL '13)*. ACM, New York, NY, USA.
- [6] Spaniol, M., Denev, D., Mazeika, A., Weikum, G., Senellart, P. 2009. Data quality in web archiving. In *Proceedings of the 3rd Workshop on Information Credibility on the Web (WICOW '09)*. ACM, New York, NY, USA, 19-26.
- [7] Van de Sompel, H., Nelson, M. L., Sanderson, R., Balakireva, L., Ainsworth, S., Shankar, H. 2009. Memento: Time Travel for the Web. Technical Report (Nov 2009). <http://arxiv.org/abs/0911.1112>
- [8] Web Archive Access Utilities. Wayback. <http://sourceforge.net/projects/archive-access/files/wayback/>