Mobile Mink: Integrating the Mobile Web with Internet Archives

using the Android Platform and Memento

Wesley Jordan

Governor’s School for Science and Technology

2014 - 2015

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Mentor Signature

Justin Brunelle

903 Enterprise Parkway, Ste 200

Hampton, VA 23666

**Abstract**

lorem ipsum dolor sit amet

**Introduction**

The Internet has been the ever-increasing source for information since its creation in the late 20th Century. However, the information contained on the web is not a permanent resource, as the erasure of old or outdated versions of web pages due to limited storage space is common practise throughout the web, and servers can also go off line and take their stored information with them. This can lead to the loss of valuable historical or legal information. Web archival services such as the Internet Archive solve this potential issue by creating *archived page*s, or lightweight copies that are stored in a single public repository, known as an *archive*. Web archival services use programs known as *web crawlers* that find a web page, make an archived copy, and then follow all the hyperlinks on a page to make copies of them, thereby repeating the process. These systems allow a user to view a web page as it appeared in several points in history.

A major development in the realm of web archival has been Memento, developed by researchers at Old Dominion University and Los Alamos National Laboratory. Memento, like the similar Wayback Machine at the Internet Archive, aims to add a defined temporal dimension to the web by providing a complete archived history of a web resource across many source archives. Memento defines a standardized framework to uniformly describe each known archival point of a web resource, known individually as ‘Mementos’. A user queries an on-line URL from a Memento proxy server and receives a machine-readable list of these Mementos, which can then be parsed and used in other applications (Van de Sompel et al. 2013).

Since the release of the first small-screened Internet-enabled devices, the mobile web has been growing substantially. This growth has been accelerated by the introduction of smartphones, which while have processing power rivaling that of many personal computers, usually feature slower and less powerful JavaScript engines and lack support for common plugins such as Adobe Flash. The proliferation of these small-screened devices has led to standard computing tasks (writing, viewing media, et cetera) previously performed by personal computers moving into the mobile space, as well as a number of websites featuring versions formatted for phones and tablets. The mobile platform possesses an increasing amount of market share, with a 2012 study finding that 69% of respondents had used the mobile web in the past year (Accenture, PLC 2012) and 83% of those from a small-screened device. This change in user habits has been enough to cause a large decline in the sales of stationary computers over the past several years, making mobile users an important demographic to

address.

However, the mobile space is increasingly ill-served by web archives. This is due in part by how mobile editions of sites are often not reachable from hyperlinks from the main site, and can only be reached by changes in a client's user agent. This leads to mobile pages having a semi-dark-web existence, being mostly inaccessible by web crawlers. This means that mobile web pages are drastically under-represented in internet archives, compared to their desktop counterparts (see Figure 1). This is especially concerning, as mobile content can differ from the content presented on desktop versions.

The proposed solution for this problem is Mobile Mink, an Android application based upon the Google Chrome extension MINK by Mat Kelly (Kelly et al. 2014). Like MINK, Mobile Mink uses the Memento framework to present a list of archived versions of a web site to the user, sorted by archival date. Mobile Mink differs from MINK in that in addition to showing a list of archived versions of the user-provided URL, it also shows archived versions of all of the mobile and desktop versions associated with it, using a method described by Schneider (Schneider 2012). Mobile Mink also allows the user to submit all of the versions of site to web archival services, such as the Internet Archive.

The design goal for Mobile Mink is similar to that of MINK: to be an easy-to-use tool that adds a temporal dimension to web browsing in the mobile space, to be able to find archived versions for all versions of web sites across screen sizes, and to be able to submit to archives the mobile versions of sites which are often missed by web crawlers.

**Application Design and Function (Materials and Methods)**

Mobile Mink is an Android application built upon version 21 (v5.0) of the Android framework. Mobile Mink is intended as a mobile offshoot of Mat Kelly's MINK, a Google Chrome extension which uses the Memento framework to show a list of archived versions (“mementos”), from which the user can select and view by selecting a menu in the lower right-hand corner of the screen. Mobile Mink works in a similar way, with the user accessing the application by using the “Share” menu found in most Android web browsers (see Figures 2a and 2b). This opens up the application with the currently open web page set up as the URL. After loading mementos from a Memento proxy server, the application displays a combined list of mementos for all versions of the site (see Figure 2c), with an indicator telling the user if the memento is from a desktop or mobile site. The user can then select a memento from this list and view the archived version it represents in their default web browser. In addition, the user can submit all of the versions of the site to multiple internet archival services for archival (see Figure 2d).

Mobile Mink replicates the availability of MINK's floating menu by being accessible in the share menu available in most popular Android browsers, using Android's Sharing Intents system. The Sharing Intents system allows applications to send out a request to share data (an “intent”) and have multiple other applications respond, from which the user can choose one. In the case of Mobile Mink, the web browsing application shows a list of applications that can handle URLs as parameter data. When the user selects Mobile Mink from this list, the application starts with the currently open URL as a parameter extra.

When the application starts, it uses a pattern matching process described by Schneider (Schneider 2012) to determine if the given URL is a mobile or desktop URL. If it is determined that the given URL is a mobile URL, the URL is converted into a desktop URL and used instead. To get a list of potential mobile URLs, the application uses a list of common mobile URL patterns (Meunier 2010) and generates a list of various permutations of the URL. For each potential mobile URL, the application launches a new worker thread in a background thread pool that first checks if the URL exists (if an HTTP header request returns <400), and if so queries, queries a Memento proxy server for mementos for the URL. Once all of the worker threads in the thread pool have finished, the lists of mementos gathered by the worker threads are combined into a single list, which is then displayed to a user. It is important to note that most of the initial fetching happens in a thread separate from the main UI thread due to limitations of the Android platform and other performance reasons.

After the list is rendered, control returns to the user, who then has the option of either opening a Memento by tapping on it, or submitting all versions of the page to one or more internet archival services. If the user selects to open a page, the application issues an intent to show the archived version in the user's default web browser (if it is set). If the user chooses to submit the pages to the archives, then the application opens up a new screen (see Figure 2d) with checkboxes to allow the user to choose which internet archival services to submit to. The archival services currently supported by the application are the Wayback Machine (through the Internet Archive) and Archive.Today, due to their open public APIs. When the user clicks the 'submit' button on the screen, the application asynchronously submits the URLs to the selected archives.

In order to test if the application met the stated design goals, it was distributed to researchers at Old Dominion University for comments and debugging.

**Design Evaluation (Results)**

lorem ipsum dolor sit amet

**Conclusions**

lorem ipsum dolor sit amet

**Acknowledgements**

Justin Brunelle, for mentoring me throughout the development of Mobile Mink and showing me the ropes of Memento.

Mat Kelly, for creating MINK and for helping to beta test Mobile Mink

Those at Old Dominion University who helped improve my original presentation

The researchers at Los Alamos National Laboratory and Old Dominion University for their work on the Memento project.

**Literature Cited**

Accenture, PLC. 2012. Mobile Web Watch 2012. (accessed November 6, 2014).

Kelly, M., M. L. Nelson, and M. C. Wiegle. 2014. "Mink: Integrating the Live and Archived Web Viewing Experience Using Web Browsers and Memento." *Proc. ACM/IEEE Joint Conference on Digital Libraries*, 2014

Meunier, B. 2010. Dotcom vs Dotmobi: M.\*.com Most Popular Mobile URL Option. Retrieved November 11, 2014.

Schneider, F. 2012. “First Steps in Archiving the Mobile Web: Automated Discovery of Mobile Websites”  *Proc. ACM/IEEE Joint Conference on Digital Libraries*, 2012

Van de Sompel, H., M.L. Nelson, and R. Sanderson, 2013. HTTP Framework for Time-Based Access to Resource States -- Memento. Retrieved November 1, 2014.

Jordan, W., M. Kelly, J. F. Brunelle, M. C. Weigle, M. L. Nelson, and L. Vobrak, 2015. “Mobile Mink:

Merging Mobile and Desktop Archived Webs”

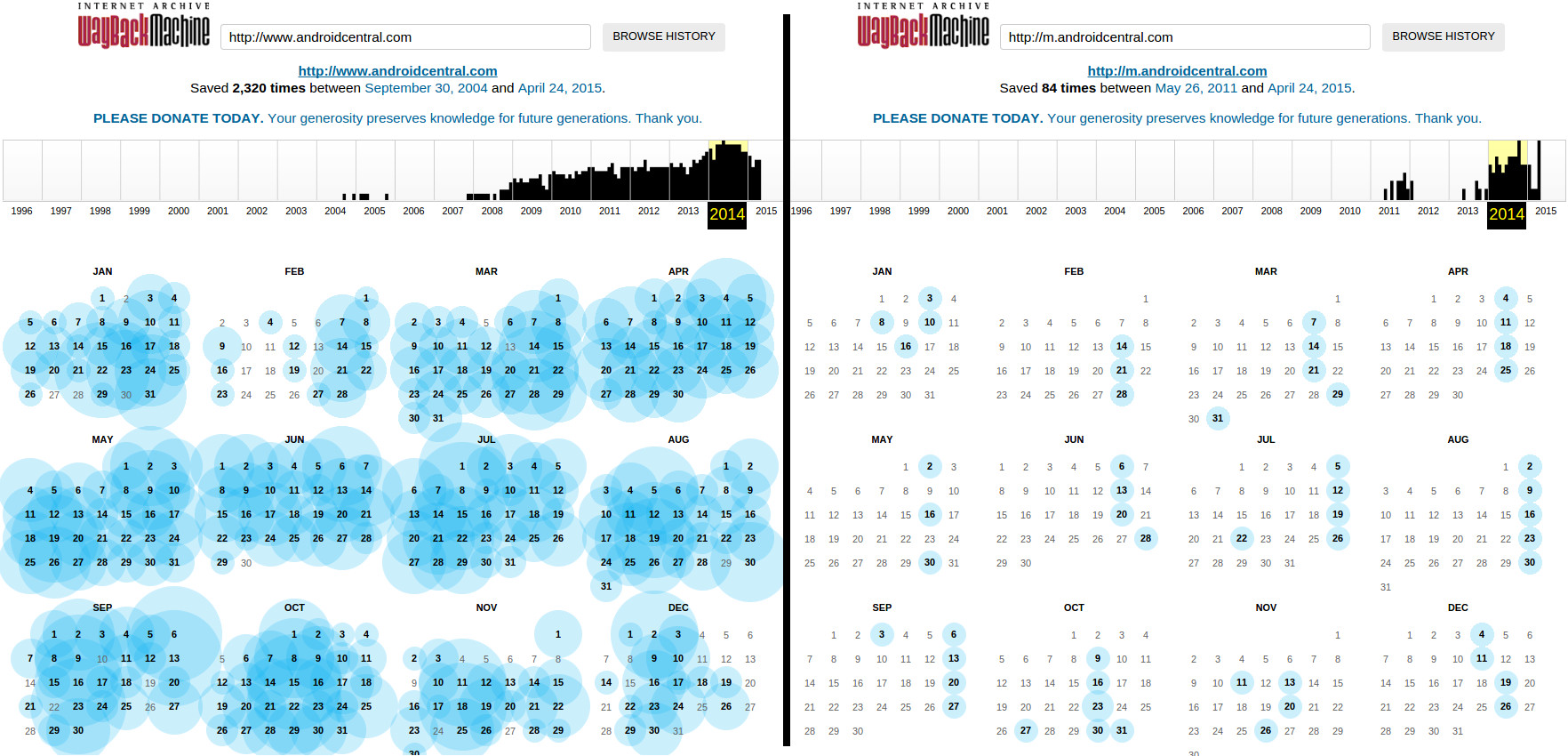
Figure 1: Comparison of number of archived versions of a desktop site (left) with a mobile site (right). A blue circle around a date indicates an archive was made on that day, with the size being dependent on how many snapshots were taken. Due to a lack of links from the desktop page to the mobile page, the mobile page is inaccessible to archival crawlers, leading to the mobile version having a smaller number of archives.

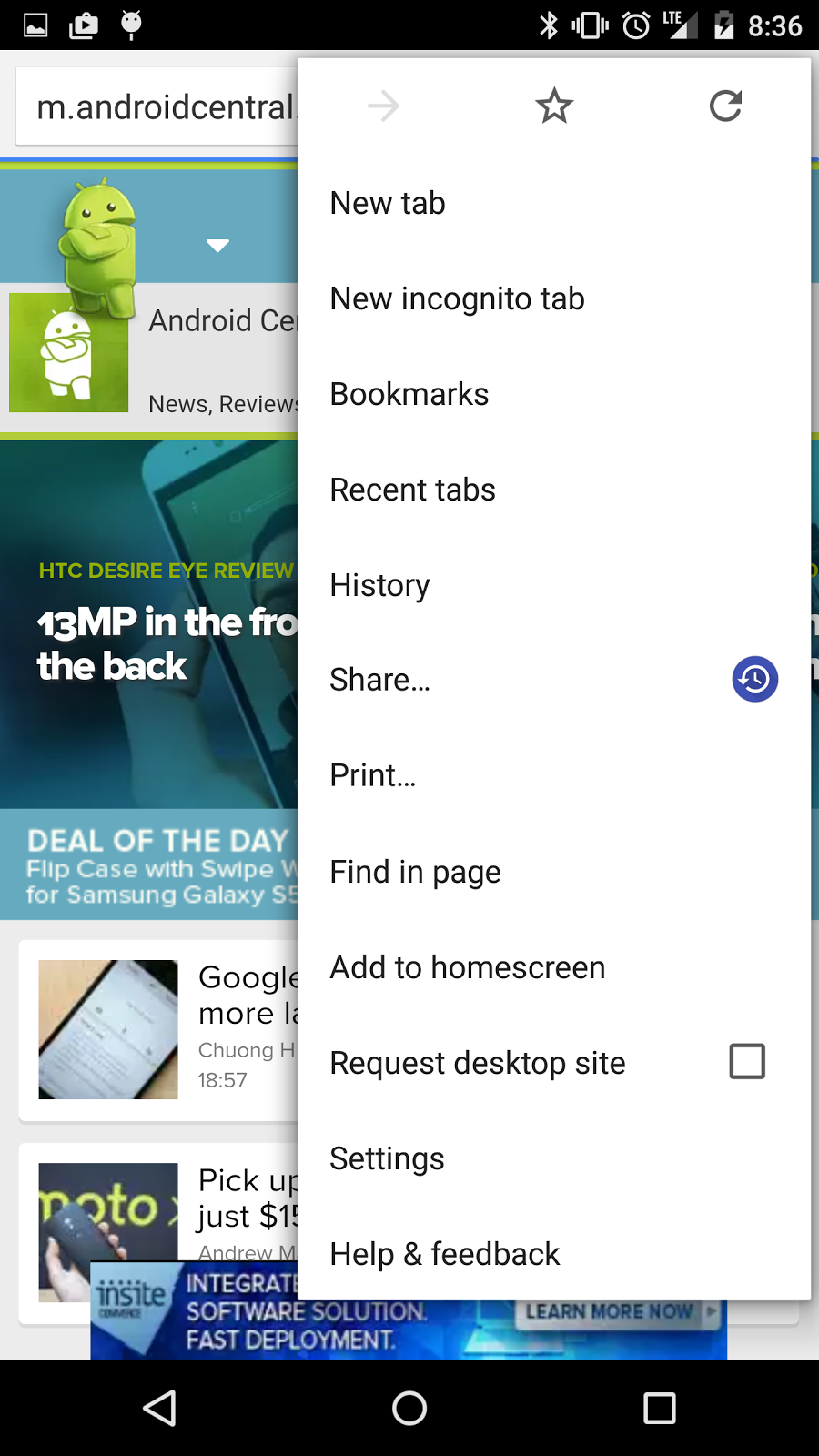
Figure 1a: The user can access the application by opening up the sharing intents menu found in most major Android web browsing applications.

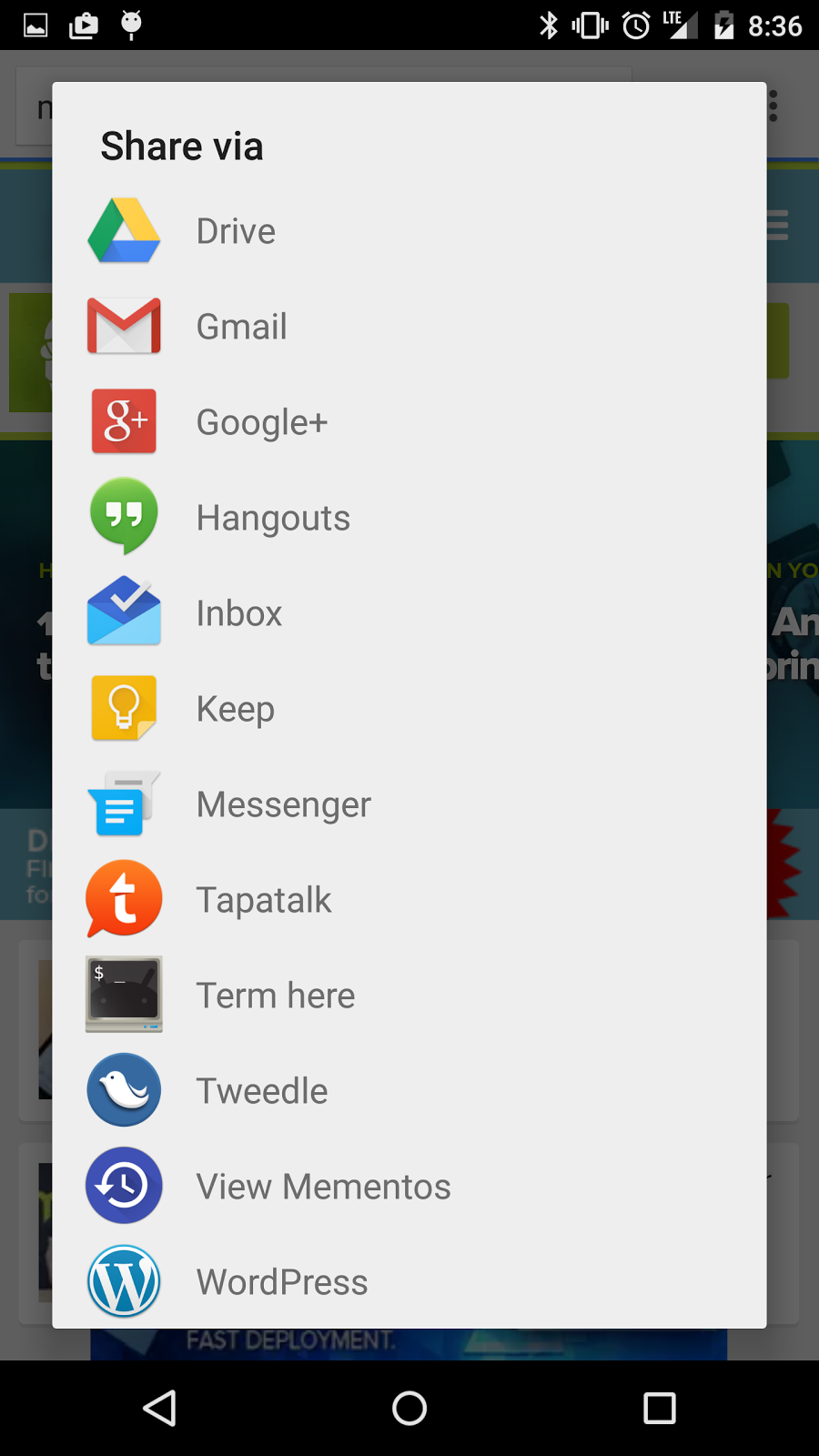
Figure 2b: The user can access the application by opening up the sharing intents menu found in most major Android web browsing applications. In this case, the user opens up the Chrome menu, and selects “Share…”, then selects “View Mementos”.

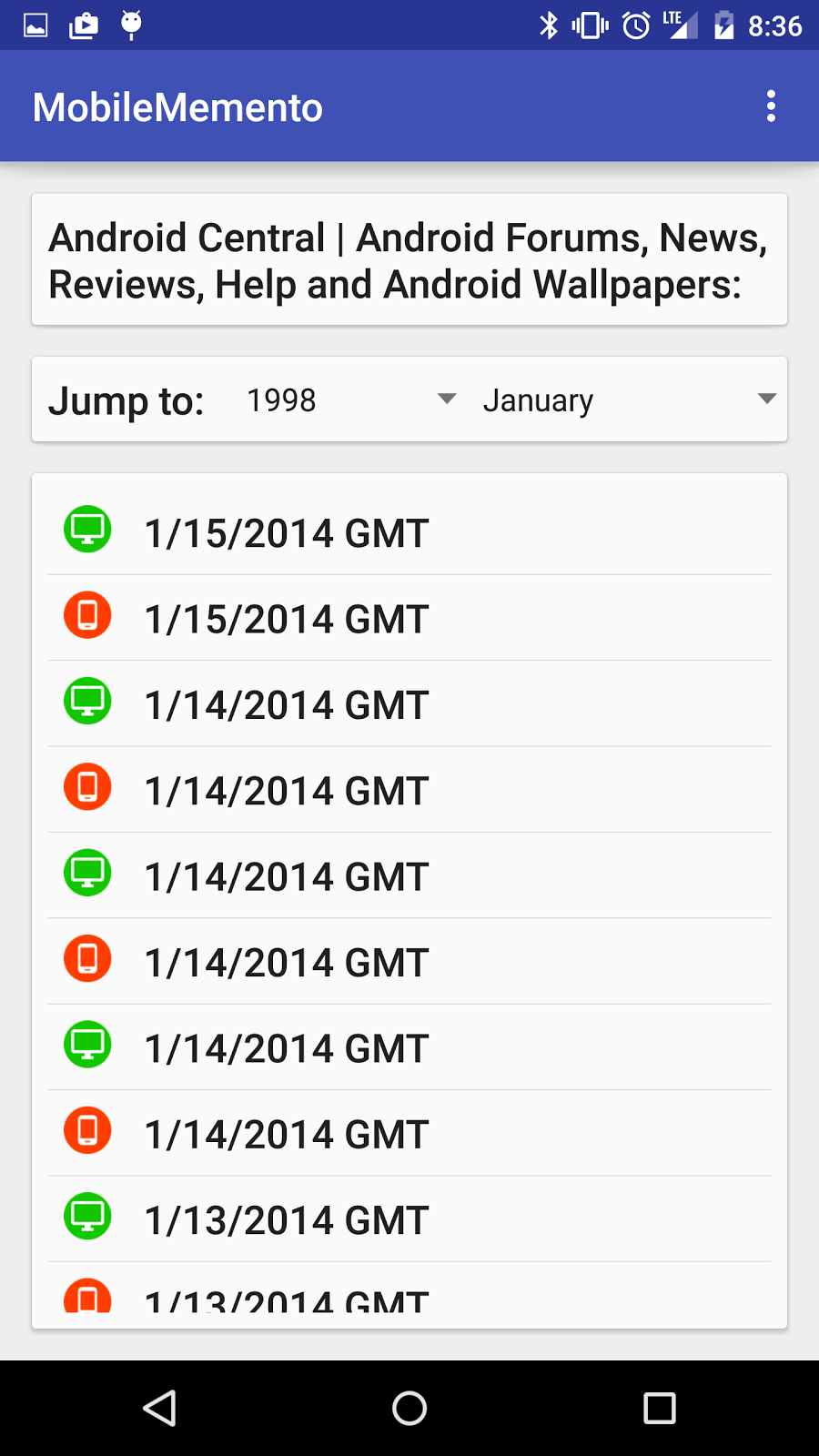
Figure 2c: The application displays a list of the Mementos from all of the TimeMaps fetched by the application. It also informs the user if they are from a mobile or traditional site.

Figure 2d: The user has the option of submitting new Mementos to various web archives.