Tweaking search user interfaces to web archives
Technical Report

David Cruz and Daniel Gomes
Foundation for National Scientific Computing
Lisbon, Portugal
{david.cruz, daniel.gomes}@fccn.pt

19 April 2013

Abstract

Despite the importance of web archives for the access to historical information published on the Internet, human interaction with web archives systems has not been thoroughly addressed. Research about user interfaces for web archives is still scarce. We present our findings gathered while adapting a typical web search user interface to the context of web archive search. We report how we adjusted the search user interface to address full-text and URL search over web-archived data. We report unexpected problems detected during usability testing of our interface and expose current limitations for future work. The web archive search user interface presented on this paper was derived from several rounds of development and usability testing over the Portuguese Web Archive search user interface (available at [archive.pt](http://archive.pt)). We believe that our work can be applied to improve the quality of the services provided by other web archives.

1 Introduction

Most web users are not acquainted with web archives and accessing to archived web data provide a significantly different user experience than accessing to the live web. Current users demand ready-to-use applications and are getting less tolerant to usability barriers. Although there is a significant number of web archives available [7], only preliminary research has been done about the design of user interfaces to gain access to temporal web data. The adoption of inadequate user interfaces to gain access to web archives jeopardizes the return of the investment made to preserve historical web content. User interfaces for web archives must be carefully designed and tested to respond to real-world user requirements and provide functional features specific to the exploitation of Web-archived content. The design of new user interfaces, to gain access to web archives, must focus on user requirements and the specific features of archived
web content. This approach would expand the scope of web archives to new kind of users.

The Portuguese Web Archive (PWA) started in 2008 and is a public search service with over 1'131 million web files archived since 1996, that aims to preserve web content of interest to the Portuguese community (archive.pt). After witnessing the difficulties of our users, we increased our effort on improving their experience and satisfaction while using the PWA.

In this study, we share our obtained results and how we adapted a typical live-web search user interface to support web archive interaction. The lessons that we learned while doing this study would have been helpful to “kick start” our web archive in 2008. Because of that, not only we share our findings but the developed code is also available freely since we believe that our contributions will help other web archivists to improve the user experience and impact of their services.

2 Methodology

The applied methodology followed an user-centered design approach [1]. This study presents the main conclusions about user interface design to support web archive search derived from the results obtained from several experiments executed during several stages of the development of the PWA. We applied user-centered techniques such as requirement analysis gathered from group brainstorming using the KJ-method [5], user feedback gathered from low-fidelity paper prototypes [2], usability testing with a think-aloud protocol, and with video and screen recording [6], interviews and user satisfaction questionnaires.

The usability of our search interfaces was tested in collaboration with HCI experts from the Human Computer Interaction and Multimedia research group from the University of Lisbon. We tested the initial prototype of our web archive but each change triggered a new round of tests. Each testing round consisted of ten tasks presented to six users with no experience of using web archives. Each of the users executed the test individually in the presence of an usability expert. We recorded the screen, audio and participants’ facial expressions for later analysis. Participants had to fill a questionnaire before the test, to inform about their proficiencies in Internet usage, and a post-test satisfaction questionnaire [4]. We finished each test with a debriefing session to further explore users’ difficulties and clarify doubts in our observations. We obtained feedback from 21 users with distinct profiles. We analyzed the obtained results using a Likert scale from “1” (strongly unsatisfied) to “7” (strongly satisfied). The average user satisfaction increased from 3.6 on the first version of our interfaces to 4.9 on the latest version.

We also obtained results from anonymous user satisfaction questionnaires filled during dissemination events by users after freely trying the PWA, where we obtained a 5.9 score. This evaluation methodology occurred in an environment less controlled than in our laboratory usability testings and used simpler questionnaires not to overload users. On the other hand, the results are obtained
on an usage environment closer to reality.

3 Anatomy of a web archive search user interface

Through usability testing on the first versions of the Portuguese Web Archive, we made two determinant observations. The first observation was that searching historical web content was an awkward concept to most web users. The existence of a website (web archive) that provides access to pages that are no longer available on their original websites, is a perception that requires technical knowledge about the functioning of the Internet that is far beyond the skills of common web users.

The second observation was that obliging the users to choose between a URL and full-text search interfaces to gain access to web-archived content was ineffective and confusing to them. Thus, we designed our interaction model to support both types of queries and shift the burden of detecting the type of query to the system. By doing so, the web archive interface becomes similar to live-web search engines, which users are already acquainted with, and guide them from familiar ground to the new context of searching historical web content. The presented search user interface designed for the Portuguese Web Archive explores the users’ familiarity with traditional search engines by offering similar layout and familiarity and enhance it with specific functions and contextual information required from web archives. However, the current scenario for web archives is to use the Wayback Machine to support URL search and NutchWax to support full-text search, but as independent search user interfaces.

3.1 Search homepage

Figure 1 presents the home page for the web archive search service. This homepage presents a search box without any temporal controls and some highlights of archived pages. This simple homepage, as starting point, provide to the users a gradual contact to new features specific of web archives. The highlights are fundamental anchors that allow users to explore curated content, especially if they have no clue of what to search for. The users observe examples that illustrate the type of information that they may find and progressively gain awareness about the potential of searching a web archive, thus reducing the cognitive effort of first-use. The publication of selected archived pages on the home page improved the overall user satisfaction with the service. Watching archived pages with historical value that have already disappeared triggered feelings of nostalgia which increased the positive perceptions, reflected through comments, about the provided service and general usefulness of web archives.

Unlike most live-web search engines, our web archive search home page also includes a fat footer at the bottom. The objective was to provide additional links to information that clarifies users about the context of web archiving and
Having submitted a query, users come across one of the most important interaction aspect of our interface: the combination of full-text with URL history search. Users only have to fill one search box and our system detects the type of query and presents the results in an interface tailored for that query type. When the query is composed exclusively by a URL, the corresponding version history results are returned. The results also include the versions from different URLs that are likely to reference the same content (e.g. www.site.pt, site.pt, site.pt/index.html). If the query is exclusively composed by text, the system returns full-text search results. If the query includes text and a URL, the system returns the full-text search results and suggests a link to the versions history of the URL.

3.2 Full-text search results

Figure 2 shows the interface for full-text queries. It is comprised by a typical search field for the query and a list of search results. However, it also includes date input fields and datepickers to restrict the temporal interval of the queries. The two datepickers define lower and upper limits of the page archive dates to
be searched. The results are shown on a results page similar to traditional live-web search engines. What differs is that we give greater emphasis to the date of archival of each result. We tried several layouts and found that the position where users better recognized the dates was below the result title. Even so, some users ignored the unusual display of the date of archival within the search
results. The interface for full-text search allows users to sort the results by relevance or archive date through the “sort:” operator or the advanced search interface (figure 3). The advanced search also provides additional fields to allow more specific queries where users can restrict for: words, phrases, words that cannot appear, file type, website or number of displayed results.

3.3 URL history search results

![Interface for URL history search deployed on the Portuguese Web Archive.](image)

Figure 4: Interface for URL history search deployed on the Portuguese Web Archive.

Searching for a URL or clicking on the “other dates” links on the full-text results page directs the users to the history view of that URL (Figure 4). The results are presented on a grid layout where each column group the several archived versions of a specific year, starting from the oldest year, supported by the Archive, up to the most recent year.

Each column then lists the available versions for that year, starting from the oldest. The users have an overall view of the versions available for a given URL. Clicking on the date link opens the correspondent version of the archived page. The grid layout approach was well understood by users. The versions from the current year are unavailable because the PWA only provides access to the archived pages one year after their archival so that the accesses to archived content do not concur with the original live-web sites (embargo policy). However, we display the current year column with a notice that explains the embargo policy to the users why the latest versions may not be visible despite been already saved by the Archive.

Figure 5 presents the new Wayback Machine interface used by the Internet Archive. However, our tests showed that users have a better understanding of the history of a page and an easier interaction using the old Wayback interface (that we use in the PWA service). This new Wayback Machine layout only shows one year of history and an overview sparkline. Despite the benefits of
the old layout, it has limitations of how much it can grow horizontally before forcing the users to scroll, which is an uncomfortable interaction for users using pointing devices such as mice.

## 3.4 Reproduction of archived content

Our present interface presents a banner on the top of the archived page with contextual information: the original URL and the archived date. Having an interface element always visible presents consistent hints that archived pages are different and behave differently than live webpages. However, we observed usability problems related to the reproduction of archived pages that deserve further research in future work. Users frequently lost perception about if they were navigating through archived content reproduced by a web archive or the live-web. When they scrolled-down the archived page, they lost visual contact with this banner. On the other hand, the banner may interfere with the layout of some archived pages by appearing on top of important content or links, for example in pages with frameset or when the layout used absolute positioning in CSS.

Figure 6 presents the future interface design for reproducing an archived content. We were peculiarly careful with the requirement analysis for this interface because it is unique to web archive search. Live-web search engines do not have to reproduce pages nor provide historical features, at most they provide a simple “cache” function that displays the textual content of the last version of an indexed pages without further concerns about maintaining its original layout. The presented design was derived through brainstorming using KJ-method \(^5\) and several rounds of usability testing using low-fidelity paper prototypes \(^2\) with a think-aloud protocol. Notice that, unlike the previously presented interfaces, this new interface design has not yet been deployed to production on the PWA.
The internal frame reproduces the archived page and provides additional features without interfering with the archived page original layout. We concluded that the interface for viewing archived pages should provide contextual information about the page (URL, date, help), features for sharing by e-mail and the main social networks (Twitter and Facebook), and for saving a copy of the archived page as image, PDF or compressed file. A sidebar enables the users to switch between versions of the archived page without having to return to the history page. To maximize the viewport devoted to the archived page, those contextual and navigational interfaces can be collapsed to a narrow bar above the archived page with the minimal information needed: The PWA logo that links to the Homepage, the URL, the date of the version presented and a button to expand the interface (see figure 7). Contrary to the old interface, the new one always show contextual information about the archived page, even if it is scrolled by the user.
4 Datepicker inadequacies

The UI element that required the most tweaks was the datepicker. Standard datepickers are conceptually simple, only presenting a grid of the days of the month and left/right arrows to view previous/next months. However, web archives collect data that can span through decades. For example, the Portuguese Web Archive hold pages archived from 1996 to 2012. Thus, traversing this date range using a standard datepicker would require 203 clicks. After several design iterations, we concluded that a web archive datepicker should use drop-down lists to allow a quicker selection of month and year of the time span of the search.

We observed that for tasks with implicit days (e.g., “Movies released during June 2000”), users only specified the month and year but did not specify the day. Then, they either dismissed the datepicker by clicking outside (doing so closed it without saving the date) or became confused hesitating on how to proceed next. For the users, choosing the month was sufficient to communicate their temporal intent to the datepicker and got flustered because they had to do the extra work of choosing and clicking on a specific day. This unsatisfactory user interaction was overcome by adding a “OK” confirmation button and a “Cancel” button to dismiss the datepicker. When clicking on the confirmation button, if no day was selected, the context of the datepicker had to be considered. Thus, we specified two different behaviors for the two datepickers integrated on our search user interface described on figure 2. If the user is defining the lower limit for the search interval through the left datepicker, the first day of the month is selected. If the user is defining the upper limit for the search interval through the right datepicker, the last day of the month is selected. If the user is defining the upper limit for the search interval through the right datepicker, the last day of the month is selected.

Before adding the “OK”/“Cancel” buttons, users were unsure about the consequences of clicking outside the datepicker area. With these buttons, users gained a strong visual anchor to decide unambiguously how to submit a new date or close the datepicker without any change to the current date.

We also observed that some participants on the usability tests clicked first on the day before adjusting the month or year. The default behavior for the datepicker was to close when the day is selected without leaving the opportunity for further adjustments. We think that this user behavior is triggered by the date format they are familiar with: e.g., 24 December 1996. They interact with the datepicker not according to the visual organization of the information but according to their mental model of the date.

We distilled a list of design tweaks for datepickers to comply with users’ expectations. Web archive datepickers need:

- Navigation arrows to jump to nearby months;
- Dropdown lists for quick access to specific months or years;
- That clicking on a day selects it without closing the datepicker;
- An “OK” confirmation button to confirm the selection of the input date and close the datepicker. If no day is selected, it chooses one according to
the context—the first day of the month if it is the starting date; the last
day if it is the end date;

• A “Cancel” button to close the datepicker without changing the date.

The customization presented in this paper can be seen in our web archive ser-
vice at [http://archive.pt](http://archive.pt). All the code and interface resources are freely avail-
able to be reused and improved at [code.google.com/p/pwa-technologies/](http://code.google.com/p/pwa-technologies/)

5 Conclusions

Search user interfaces for web archives and live-web search engines should be
similar to ease the adoption of web archives by non-expert users. On the other
hand, following this approach strictly does not provide the temporal information
nor specific mechanisms required to exploit historical information.

The obtained results, through usability testing, showed that the concept of
using web archives to search for archived pages that are no longer available
on the live Web is confusing to most users. Web archives have to balance the
familiarity of search engine interfaces with their specific temporal concepts.

For web archives, several aspects have to be specifically addressed: how to
handle each type of queries; the result interfaces for each type of query; how
to make users notice the temporal information of the archived pages; handling
and input of dates; the display of archived pages, so users know they are on an
archived copy of a page.

Therefore, the PWA uses an interface based on traditional search engines and
empowers it with features to show and allow users to manipulate the temporal
dimension. Not only should web archives allow the familiar full-text queries, but
they also have to support URL queries, in a similar familiar interface, since they
respond to different users’ needs. Despite starting with familiar interfaces, each
type of query requires specific interfaces that meet those users’ expectations.
The date of results and the archived pages have to be explicitly shown and
the users must be able to control easily. For the PWA, we use datepickers to
restrict the date range of searches. But despite being a familiar mean to interact
with dates in websites, standard datepickers are not adequate to be used in web
archives and need several adjustments to be successfully used.

By presenting the adjustments that we made to our interface and explaining
their rationale, and also by providing our work under a free open-source license,
we expect to raise the awareness of the importance of the interfaces for the
success of web archives. The interfaces of web archives are still an unexplored
research field that could help transforming the foreign concept of web archives
to something that Internet users understand and deemed necessary. New user
interfaces and data visualizations will offer users new ways of seeing and working
with temporal data, and thus unfold the full potential of web archives.
References


