

# Final Report

Summer 2016

Xiaoyu Tai ([xtai@umd.edu](mailto:xtai@umd.edu))

August 15, 2016

## Contents

<b>1</b>	<b>Internship Overview</b>	<b>1</b>
1.1	Location . . . . .	1
1.2	Project Overview . . . . .	2
<b>2</b>	<b>Team Overview</b>	<b>2</b>
<b>3</b>	<b>Internship Projects and Outcomes</b>	<b>3</b>
3.1	Phrase I . . . . .	3
3.2	Phrase II . . . . .	5
3.3	Internship Impressions and “Lesson Learned” . . . . .	7

## 1 Internship Overview

### 1.1 Location

The Maryland Institute for Technology in the Humanities (MITH, <http://mith.umd.edu/>) is a leading digital humanities center that pursues disciplinary innovation and institutional transformation through applied research, public programming, and educational opportunities. Jointly supported by the University of Maryland College of Arts and Humanities and the University of Maryland Libraries, MITH engages in collaborative, interdisciplinary work at the intersection of technology and humanistic inquiry. MITH specializes in text and image analytics for cultural heritage collections, data curation, digital preservation, linked data applications, and data publishing.

Here are some examples of what they have done in the past:

- Synergies Among Digital Humanities and African American History and Culture - An Integrated Research and Training Model  
([Link to Synergies Among Digital Humanities and African American History and Culture](#))

- Documenting the Now - Supporting Scholarly Use and Preservation of Social Media Content ([Link to Documenting the Now](#))
- Enhancing Music Notation Addressability ([Link to Enhancing Music Notation Addressability](#))
- Shelley-Godwin Archive ([Link to Shelley-Godwin Archive Project](#), <http://shelleygodwinarchive.org/>)
- “O Say Can You See”: the Early Washington, D.C. Law and Family Project ([Link to the Early Washington, D.C. Law and Family Project](#))

## 1.2 Project Overview

- **IIIF Newspaper Viewer Project**
  - **About IIIF:** the International Image Interoperability Framework (IIIF, <http://iiif.io/>) defines two application programming interfaces that provide a standardized method of describing and delivering images over the web, as well as presentation metadata about structured sequences of images. This is an image standard for world’s leading research libraries and image repositories.
  - **About the project:** the goal of this project is to provide a web-based IIIF image viewer that is suitable for viewing newspaper content. The client includes the University of Maryland Libraries and will be used on the library website.
  - This project will be continued as my regular graduate assistant job at the Maryland University Libraries in fall 2016.
- **Research on “Exploring Historical Newspaper”:** after we identified an existing potential candidate for the IIIF image viewer, my project moved to explore ways of discovering historical newspaper. The goal of this research project is to:
  - 1) understand the user flow of discovering historical newspaper in current repositories,
  - 2) present best practices on discovering historical newspaper,
  - and 3) explore novel ways of discovering newspaper repositories with current and future technologies like content analysis (machine learning) and data visualization.

## 2 Team Overview

The Maryland Institute for Technology in the Humanities (MITH) is a leading digital humanities center that pursues disciplinary innovation and institutional transformation through applied research, public programming, and educational opportunities, and MITH is a relatively small group of less than 15 people (<http://mith.umd.edu/people/>). Since it is a small group, I got the chance to know most of them. My ‘team’ is small at MITH since my project is rather self-orientate; however, I do have some input from my co-workers and my supervisor, who gave me many useful tips for my work.

Most of my co-workers have their own research projects and funding from multiple organizations. They have also been working on learning new technologies as a team. For example, MITH has a “demo Tuesday” – a presentation provided by someone who just learned something cool. During my internship, React (<https://facebook.github.io/react/>) and web pack (<http://webpack.github.io/>) had been a focus for them.

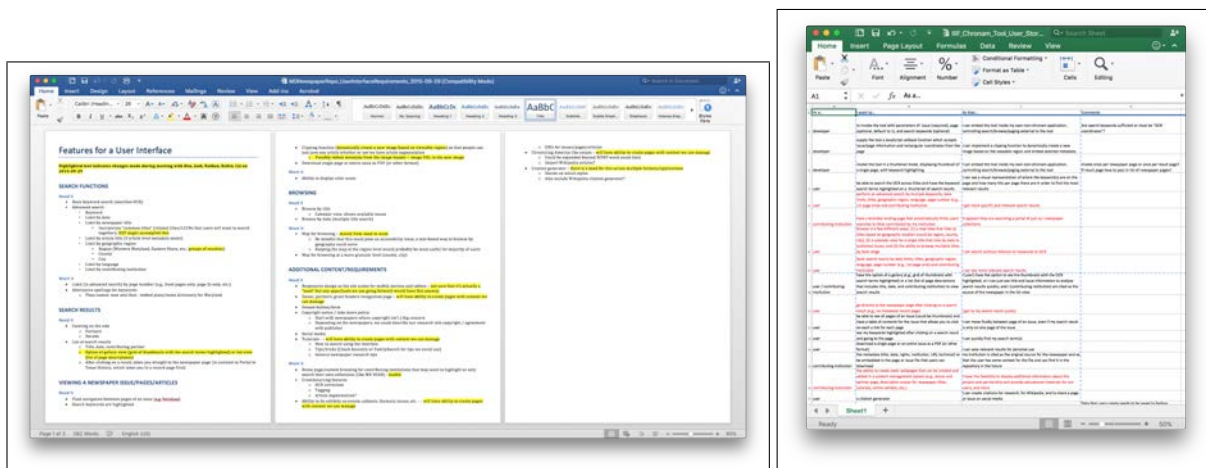
### 3 Internship Projects and Outcomes

The overall impression of this internship is okay; however, a change of direction was surprising: it changed from the newspaper viewer project to research and provide advice on discovering historical newspaper. (though, I will continue the project in fall semester.)

For the sake of documentation, I will separate my internship into two parts: phase I and phase II. Phase I represents the beginning of the IIF newspaper viewer project and the uncovering of the Mirador viewer. Phase II indicates my research on discovering historical newspaper.

#### 3.1 Phrase I

My internship started from two documentations: the first one is a list of user interface requirements for the newspaper viewer, the second document has a list of user stories which also includes many actions for the viewer. My first task is to find existing software and design a mock-up to implement all features from these two documentations.

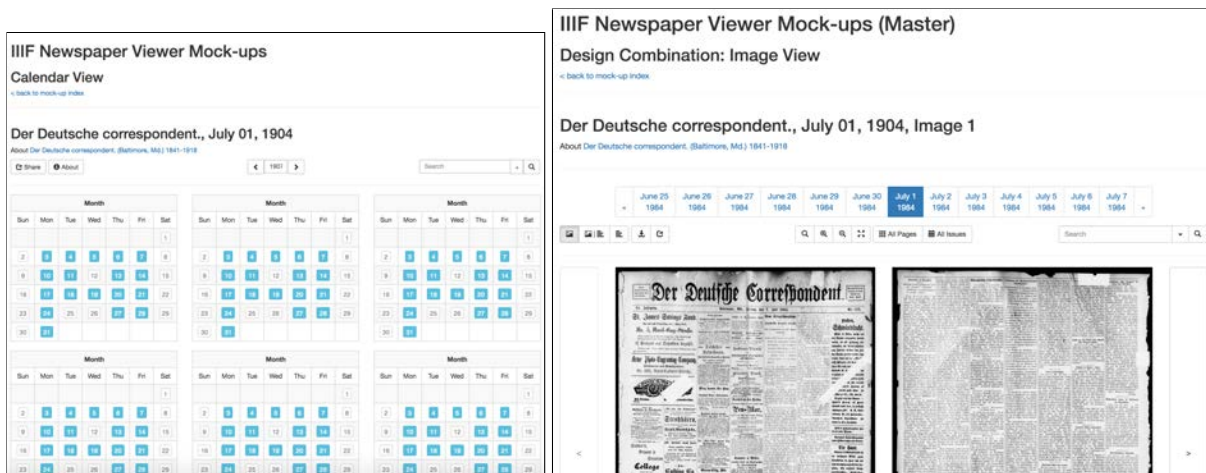
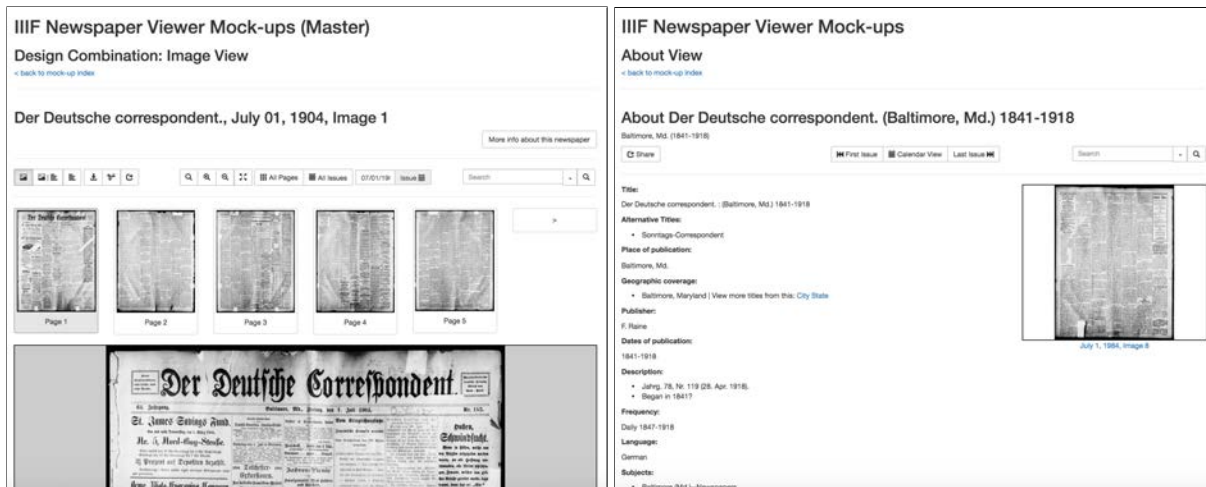


Documentations of user stories and user interface requirements.

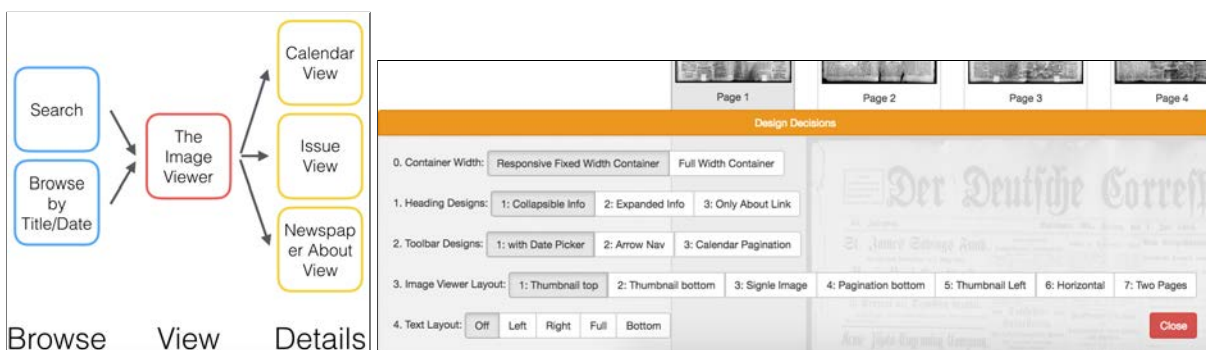
My supervisor at MITH, Trevor, provided me some guidelines on mock-up creating and suggested me to read a list of useful articles about creating mockups and making the final product. The most impressive article is called “Thinking in React”(https://facebook.github.io/react/docs/thinking-in-react.html). Although this article is a tutorial on building React application, the essential of this article is the relationship between mock-ups and final products. This article guided me to create some first-round mock-ups for the newspaper viewer. Screenshots of my mockup are provided below. I created about 10 HTML pages to represent the viewer’s user flow and interfaces, and it can be accessed at <http://iiif.taixiaoyu.com/mockup/>.

Along with creating mock-ups, I was researching existing IIF-supported clients so that the team is not “reinventing the wheel.” Serval open-sourced candidates showed up in the list:

Mirador(<https://github.com/IIIF/mirador>), OpenSeadragon(<http://openseadragon.github.io/>), Diva.js(<http://ddmal.github.io/diva.js/>), and Universal



Screenshots of different views in the mockup.



The user flow of the newspaper viewer; dynamic design variations in the viewer page.

Viewer(<https://github.com/UniversalViewer/universalviewer>). Mirador and Universal Viewer are two well-completed IIIF clients that support various settings; however, none of them meet all the requirements in the feature list and user stories documents. On the other hand, Diva.js and OpenSeadragon are the core implementation of IIIF clients which provide great support for all IIIF APIs, and these two barebone

projects will be useful for a more in-depth development if Mirador and Universal Viewer both turned down by the stakeholders.

After creating the mockups that meets all requirements as the docs and providing implementation of Mirador and the Universal Viewer, I was heading to a stakeholder meeting where I presented my mockups to four of the newspaper viewer stakeholders, Ben Wallberg (wallberg@umd.edu, who is also my supervisor for my graduate assistantship), Robin C. Pike (rpike@umd.edu), Kate Dohe (katedohe@umd.edu), and Elizabeth M Caringola (ecaringo@umd.edu). In the meeting, we cleared many things in the documentations and Mirador seems to be a great fit for the project. However, future developments are needed to meet all requirements. Here is a screenshot of the JIRA main issue (<https://issues.umd.edu/browse/LIBDC-1143>).

Digital Collections / LIBDC-1143  
Project: Mirador/IIIF based Newspaper Viewer

Edit Comment Assign More Mark for Review Resolve Issue Close Issue

**Description**

Project to develop a newspaper IIIF-based viewer for use with Diamondback, Historic Maryland Newspapers, etc. The code base will be the existing Mirador client. Xiaoyu Tai will provide the basic coding with technical guidance from MITH staff.

Mirador is an open-source, web-based 'multi-up' viewer that supports zoom-pan-rotate functionality, ability to display/compare simple images, and images with annotations.

<https://github.com/IIIF/mirador>

**Issues in Epic**

LIBDC-1156	Setup Mirador development environment	IN PROGRESS	Xiaoyu Tai
LIBDC-1145	New Feature: Hit Highlighting	IN PROGRESS	Xiaoyu Tai
LIBDC-1144	New Feature: Image Clipping	OPEN	Unassigned
LIBDC-1146	New Feature: view OCR text side-by-side with page image	OPEN	Unassigned
LIBDC-1147	New Feature: article level highlighting and navigation	OPEN	Unassigned
LIBDC-1148	Add the ability to zoom in while in "scroll" view	OPEN	Unassigned
LIBDC-1149	In single image view, keep zoom levels consistent from page to page	OPEN	Unassigned
LIBDC-1150	Configure Book View to behave like a "true" 2-up display	OPEN	Unassigned

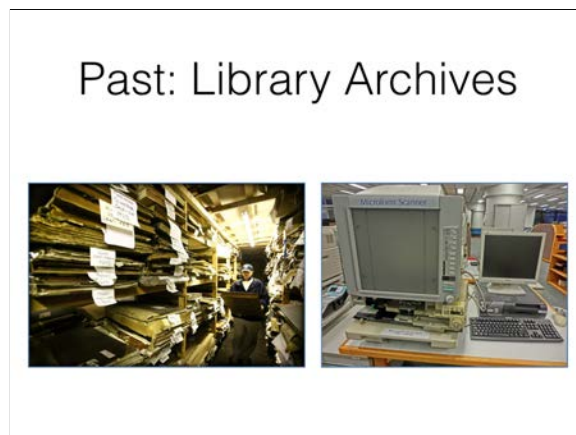
After discussing with Ben and Trevor, the implementation of Mirador became part of my regular GA job, and I was assigned to a new research project to continue as part of my internship.

### 3.2 Phrase II

The second phrase at my internship was researching ways to explore historical newspaper. I was quite excited about it since this research involved user interface designs as well. The final outcome of this research is a research presentation.

My supervisor, Trevor, helped me to get start with this project by providing me some examples as a newspaper expert: for instance, currently, many newspaper archives only provide two ways to discover historical newspaper: browsing and searching. The first question we want to answer in this project is what are the design alternatives of browsing digitized newspaper. The second research question is how to explore

newspaper dataset with modern approaches.



Past newspaper archive - actual newspaper collections and microform films.

The beginning of this research include case studies with five major and different historical newspaper archives: *Chronicling America* (<http://chroniclingamerica.loc.gov/>), *Newspapers.com*, *Trove* (an Australian historic newspapers archives: <http://trove.nla.gov.au/>), *Veridian* (a newspaper digitization service, California Digital Newspaper Collection was used as the study case: <http://cdnc.ucr.edu/cgi-bin/cdnc>), and *Google Newspapers* (<https://news.google.com/newspapers>)



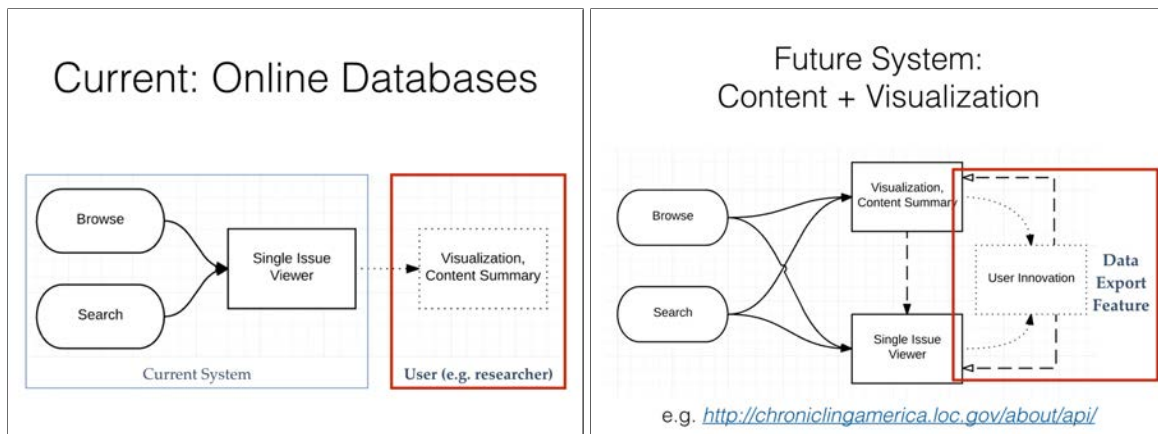
Five major historical newspaper archives.

These five existing applications represent most approach toward browsing newspaper archives. The research is focused on the user interface of browsing, searching, and the viewer. Several interaction methods were highlighted in the research: column view, interactive breadcrumbs, interactive map, mini map, and content awareness.

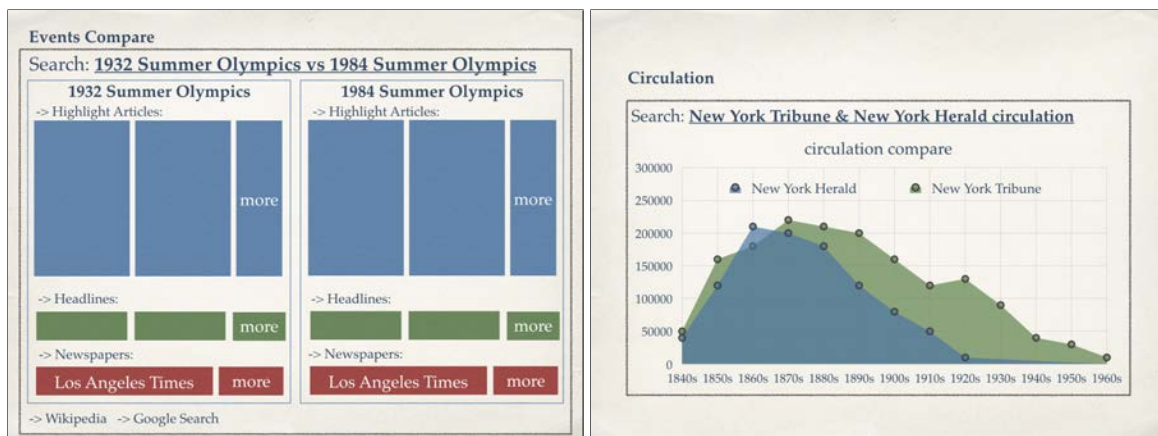
The second half of this research is focused on discovering new approaches of the system. In the past, people have to go to the library to browse newspaper collections and archives. Now, we have multiple online systems that offer people the opportunity to search and browse collections, but advanced techniques are required to summarize findings from these collections; therefore, the insight of this collection is not accessible to the public audience.

With technology became more and more powerful, we have the ability to index the collection more efficiently and provide more insight about the newspaper database. For example, in the research, I proposed a way to display history event summaries by providing clips from different historical newspapers, therefore, to show a completed picture of history from newspapers to the general public. Another approach to the data set is to provide more data visualizations such as language N-gram and circulation data.





The potential use of machine learning and data visualization in the future newspaper viewing system; case studies of five major and different historical newspaper archives.



A comparison of history events from newspaper clips; a visualization of newspaper circulation data.

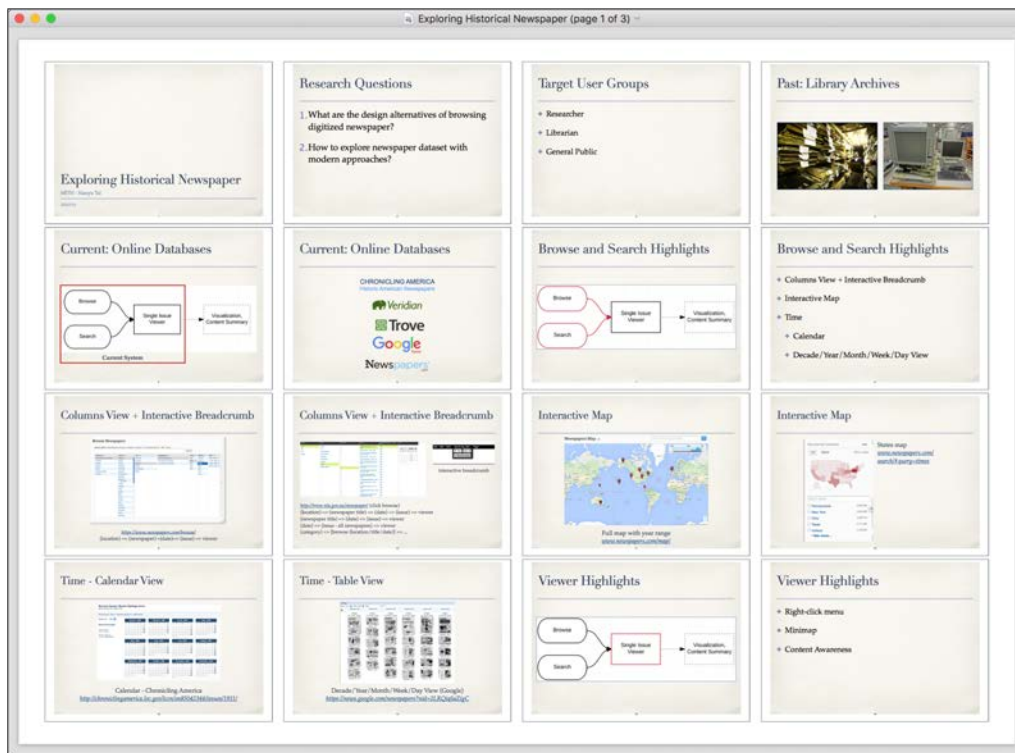
The overall goal of the future system is to present more data and insight to the general public and save time for researchers by doing part of their job.

The research is concluded by possible solutions in the near future. The conclusion provides important user interface improvements for the IIF newspaper project and potential usage of future and existing technology in the project. An overall of the final presentation keynote is presented as follow.

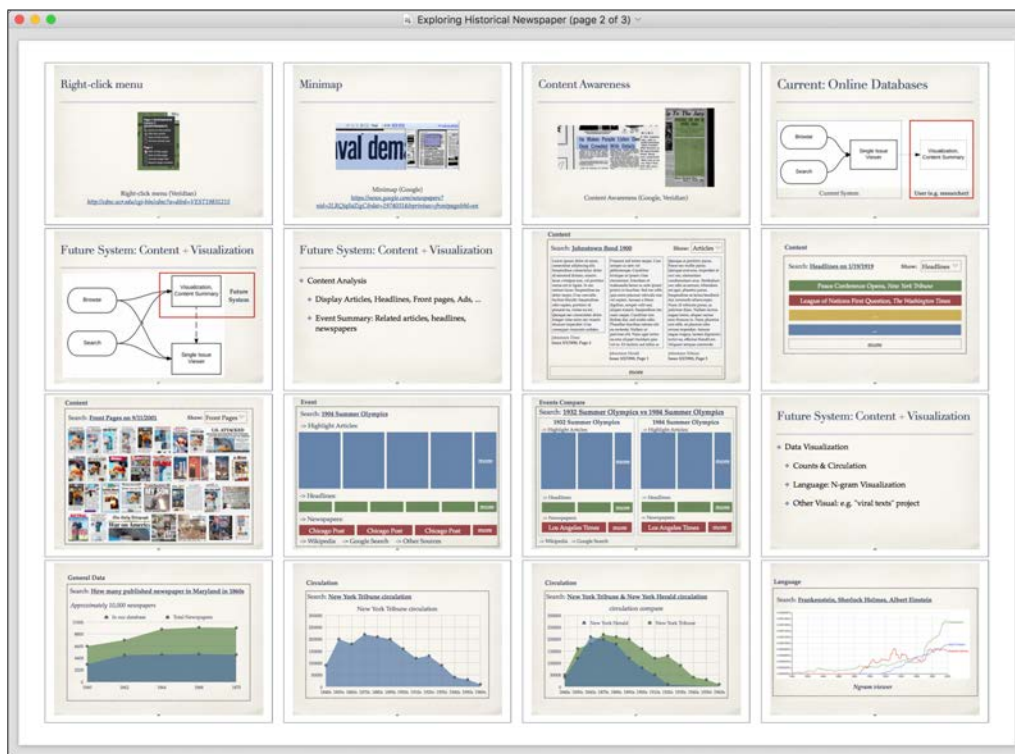
### 3.3 Internship Impressions and “Lesson Learned”

My first impression on this internship is quite good; the working environment is very friendly and people are very nice to me.

I learned a lot from this internship since I got two parts of experience as a UI designer and a researcher. It was interesting that how these two parts same and different. Both designer and researcher need skills to create mockups to convey ideas; however, as a designer/developer, a person needs more advanced-level skills on mockup creating and implementing. A researcher, on the other hand, is required less technique to

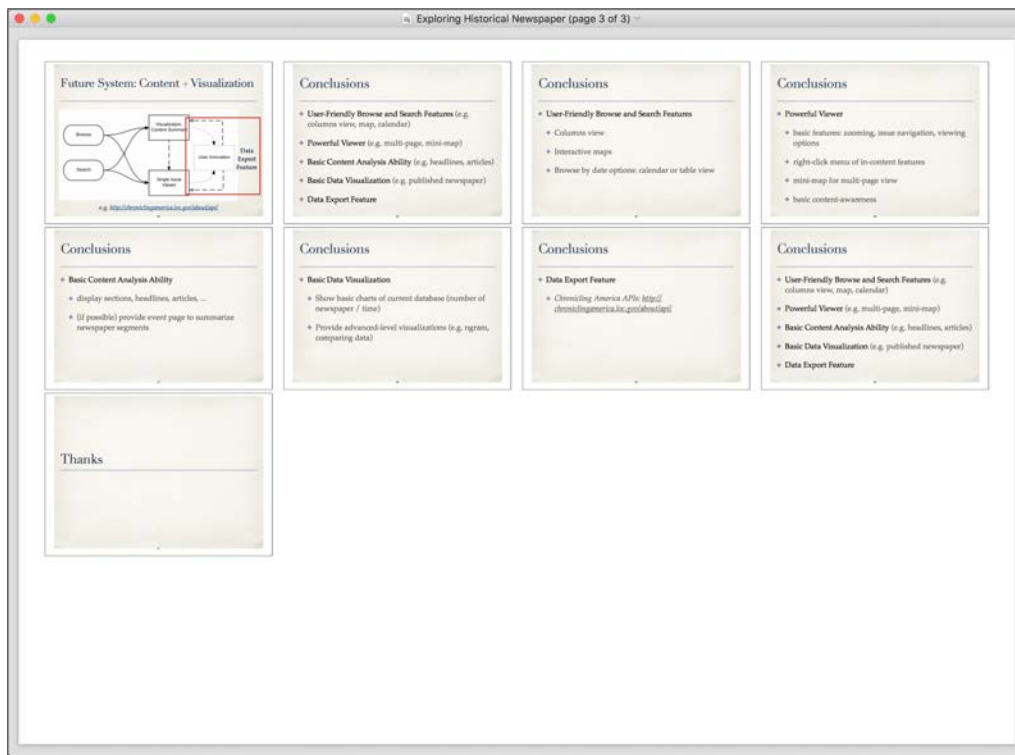


The completed version of the final presentation (1 of 3).



The completed version of the final presentation (2 of 3).





The completed version of the final presentation (3 of 3).

create mockups.

The most important lesson learned from this internship was not to spend time on creating something that is existing, and spending the time to do research is more important in the beginning of a project.



Nice office environment with a historical personal computer collection.