**Pioneer Days in Florida Lesson Plan: Introduction to the Digital Humanities with the Voyant Data & Text Mining Tools**

**Overview**

Students will learn about the Digital Humanities by learning about how to use data and text mining tools (Voyant) to explore and analyze textual materials.

This lesson supports teaching with and about primary historical sources, and teaching about historical research methods, including new and developing methods for massive archives of digital materials, and for thinking about information in the age of Big Data.

- **Time Required:** Half day (4 hours, all computer time)
- **Target Audiences:** Advanced high school, undergraduate college, or graduate students
- **Materials Required:** Computer classroom with web access to go to:
  - *Pioneer Days in Florida:* [http://ufdc.ufl.edu/pioneerdays](http://ufdc.ufl.edu/pioneerdays)
- **Assessments:** Because this is an introductory lesson, it is designed to introduce new terms and topics, including helping students become familiar with primary sources, digital research methods, and concepts (Digital Humanities, Big Data, text mining, stop words, distant reading, etc.). Future assessments will build off of the material presented in this lesson. Assessments on this lesson alone could include methods to evaluate learning the new terms and concepts.
Introduction

About Pioneer Days in Florida

The Pioneer Days in Florida: Diaries and Letters from Settling the Sunshine State, 1800-1900 Digital Collection (http://ufdc.ufl.edu/pioneerdays) includes 36,530 pages of diaries and letters describing frontier life in Florida from the end of the colonial period to the beginnings of the modern state. These first-hand accounts document the experiences of settlers, soldiers, and travelers who trail-blazed Florida during the wars of Indian Removal, the Civil War, and the Gilded Age. These 19th century manuscript materials from the Florida Miscellaneous Manuscripts Collection within the P.K. Yonge Library of Florida History, George A. Smathers Libraries, University of Florida.

Pioneer Days in Florida is a Digital Humanities scholarly work curated archive that builds on existing scholar created intellectual access from finding guides, metadata records, and database information built by scholars to support access to the resources in the collection (first as physical materials, and then as digital). As a digital collection, new opportunities are possible, including text and data mining.

About the Digital Humanities

As defined by Wikipedia (with many competing and complementary definitions in use by scholars):

The Digital Humanities are an area of research, teaching, and creation concerned with the intersection of computing and the disciplines of the humanities. Developing from the field of humanities computing, digital humanities embrace a variety of topics, from curating online collections to data mining large cultural data sets. (http://en.wikipedia.org/wiki/Digital_humanities)

Digital Humanities research builds and uses tools for exploring digital texts, and much more.
Teaching Activities

1. Introduction & Orientation to Pioneer Days
   - Introduction to the Digital Humanities and Pioneer Days in Florida (text above)
   - Orientation to Pioneer Days in Florida: http://ufdc.ufl.edu/pioneerdays
     - Browsing
     - Searching
     - Description, and permanent URLs for referencing

2. Student Activities with Pioneer Days
   - Student exploratory time on the site in pairs or groups, to read about materials
   - Student time to select at least two items of interest, identifying:
     - Title
     - Permanent URL
     - What they think this item will be about, contain, main themes, topics, etc.?
     - How does its physical condition inform what they think?
   - Optionally: to the class and by turns, each pair/group of students may share their notes on their selected items of interest

3. Overview of the Voyant Tools
   - Showing Voyant Tools, web-based reading and analysis environment for digital texts: http://voyeurtools.org/
   - Review of the interface: see http://hermeneuti.ca/voyeur/users for the interface overview from the Voyant Tools website or review the interface with the example
     - Entering sample URL for full PDF:\(^1\)
       http://ufdclimages.uflib.ufl.edu/AA/00/01/35/01/00001/LogBook.pdf
     - Review of the Voyant Tools interface with the sample

4. Discussion and Activity with the Voyant Tools (detailed below, after the overview and introduction to the Voyant Tools)

---
\(^1\) Permanent URL for the full item: http://ufdc.ufl.edu/AA00013501/00001
Voyant Tools: initial screen with two main columns

Left-column, top: Cirrus word cloud
Left-column, bottom: Summary
Right-column: Corpus Reader

Notes on the initial screen:
Click on the small gear icons to adjust settings.
Click on the small expand-arrows next to “Words in the Entire Corpus”, “Corpus”, and the double arrows to the right of the “Corpus Reader” to expand other tools.

2 As explained http://www.tapor.ca/?id=8; “Cirrus is a visualization tool that displays a word cloud relating to the frequency of words appearing in one or more documents. One can click on any word appearing in the cloud to obtain detailed information about its relativity.”
Voyant Tools: main screen with expanded tools showing

By clicking on the small expand-arrows, additional tools display, as well as more expand-arrows for even more ways to interact with the text.
4. Discussion and Activity with the Voyant Tools

Changing Stop Words
Stop words are words which are filtered from results, and are often based on lists of very common words. For instance, in the screenshot before this section, the largest word in the word cloud is “the”.

Click on the gear above the Cirrus word cloud to select different options, and select the Taporware (English) stop words.

Doing so changes the word cloud dramatically by removing common words like “the”:
Next, click on the gear above the “Words in the Entire Corpus” to remove the stop words.

Notice that the most common words are now “boat” and “professor”.

Clicking on “professor” in the “Words in the Entire Corpus” on the left-column then generates the “Word Trends” graph on the far right.

Clicking on any of the terms in any of the panels similarly changes other panels.
**Discussion**

With these panels open and expanding others as interested, how does this relate to other reading experiences?

**Discussion Concept: Distant Reading**

In the Digital Humanities, one approach or method to reading masses of digitized texts is called “Distant Reading.” Distant reading employs quantitative methods to analyze texts in new ways and can offer new ways of reading.

Distant reading differs from close readings, which are focused, deliberate, and sustained readings of texts, often focusing on brief passages of text to conduct the analysis. Close reading and distant reading are research methods or approaches that inform the gathering of data or evidence and analysis.

Other types of readings include casual readers skimming online websites, often going from one site to another, in an exploratory manner.

**Activity**

In considering different types of reading, students (in pairs or groups) should explore the text using the Voyant Tools with the concept of distant reading (and reading methods overall) in mind.

For this activity, students should be asked to verbalize their thoughts on what they think the interface options will do before clicking or making any changes. The other student(s) should also offer their thoughts and discuss any actions and changes.

This pair-group activity format in using interfaces is common for interface and system usability testing, where people in pairs share how they think the system does work, and which informs system design. It’s a useful practice for usability testing and for users to employ in collaborative learning for new technologies and new methods.

This activity format supports learning interfaces together and awareness of and the development of new methods for approaching new tools and readings.
Discussion

After time exploring and discussing the text as read in the Voyant Tools, the description for the logbook (which contains all of this text) is useful to review to see what different information and sense of the material is gained in a quick review.

Logbook description:

A woman traveler’s account of an 1874 voyage on the steamboat Okahumkee of the Hart Line. The memoir appears to be the work of Martha D. (House) Allen, from Pennsylvania, although the name Martha H. Holmes is also associated with the text. Martha House Allen (1827-1900) and her husband Charles J. Allen (1822-1887), of Baring St., Philadelphia, were passengers on the Okahumkee as it traveled up the St. Johns River to the Ocklawha and on to Silver Springs. She describes daily activities, sights, and adventures on the trip, paying close attention to wildlife and local plants. Referring to herself as "the Scribe", she never uses her or any of the passengers' actual names, instead creating nicknames, according to the occupation or personality of the passenger. One of the passengers reads aloud from Harriet Beecher Stowe's Palmetto Leaves, published the previous year, as the steamboat progresses. The Okahumkee, a rear-paddle steamer, was one of numerous steamboats operated by the Hart Line in the 1870s, others being the Osceola and the Hiawatha. It was a popular tourist attraction during the last quarter of the nineteenth century and was still afloat although derelict in the 1930s.

Optional: Student Activities with Voyant Tools for Text Visualization and Data Mining

- Student exploration of text visualization and data mining:
  - Students, in pairs or groups, go to Voyant Tools (http://voyeurtools.org/)
  - Students enter text or a URL for one or more of their selected items of interest

3 http://ufdc.ufl.edu/AA00013501/00001/citation