**Voyant Tools Text Mining Lab: Visualizing the MJP**

Wed 10/24/2012

This lab introduces students to a number of the tools, practices, and interoperability functions of the Voyant Tools text mining suite. Through a series of tasks and questions, we will explore different ways of analyzing word usage trends in the MJP corpus, reading visualization graphs, and exporting those tools and graphs to other formats. You are asked to write up your responses to this lab worksheet and post them to the course site, with images and embedded tools.

The following links to the Voyant Corpora can also be found on the Downloads page of the course website (<http://courses.utulsa.edu/modmag/?q=content/downloads>).

**MJP Corpora at Voyant Tools**

Chronological corpus of the 9 magazines individualized below

<http://voyeurtools.org/?corpus=1351025844510.61&stopList=stop.en.taporware.txt>

*BLAST*

<http://voyeurtools.org/?corpus=1351006206655.8775&stopList=stop.en.taporware.txt>

*The Crisis*

<http://voyeurtools.org/?corpus=1351006670332.4242&stopList=stop.en.taporware.txt>

*The Freewoman*

<http://voyeurtools.org/?corpus=1351007344024.8615&stopList=stop.en.taporware.txt>

*The New Freewoman*

<http://voyeurtools.org/?corpus=1351007768576.8502&stopList=stop.en.taporware.txt>

*The Egoist*

<http://voyeurtools.org/?corpus=1351007263352.2751&stopList=stop.en.taporware.txt>

*The Little Review*

<http://voyeurtools.org/?corpus=1329252640907.4758&stopList=stop.en.taporware.txt>

*Others*

<http://voyeurtools.org/?corpus=1351008139831.4765&stopList=stop.en.taporware.txt>

*Poetry*

<http://voyeurtools.org/?corpus=1351008290752.187&stopList=stop.en.taporware.txt>

*The Tyro*

<http://voyeurtools.org/?corpus=1351008418048.5198&stopList=stop.en.taporware.txt>

**Part I. Comparing Corpora**

Pick one or two terms that you are interested in analyzing across two or more of the magazines. The number of terms and magazines is up to you, but it’s probably best to start small. Repeat these steps for each of the magazines you decide to look at.

1. Word snapshots: Go to the Voyant Corpus page of your first magazine. In the Summary panel, what words are the most frequent, which ones have the most notable peaks in frequency across the magazine, and which words are the most distinctive? What does this information tell you about the magazine or its context?
2. Document properties: In the same corpus, what does the Summary panel tell you about the number of documents it contains? Which issues are the longest by word count? Which issues have the highest vocabulary density? What does this information tell you about the magazine or its context?
3. Using other tools such as Word Trends, Words in the Entire Corpus, and Words in Documents, what sorts of patterns catch your interest. Do they seem to bear any relationship with what the Summary panel tells you? Why or why not?
4. Take a screenshot of any graphs or visuals that you think are worth posting or discussing. Save locally and be prepared to post to course site. ***Note:*** You may wish to view individual tools in isolation to get better graph readings, by clicking the floppy disk icon to the upper right. Here is a list of commands:

**Screenshot Shortcuts**

(Taken from <http://guides.macrumors.com/Taking_Screenshots_in_Mac_OS_X>)

* *Command-Shift-3*: Take a screenshot of the screen, and save it as a file on the desktop
* *Command-Shift-4*, then select an area: Take a screenshot of an area and save it as a file on the desktop
* *Command-Shift-4*, then space, then click a window: Take a screenshot of a window and save it as a file on the desktop
* *Command-Control-Shift-3*: Take a screenshot of the screen, and save it to the clipboard
* *Command-Control-Shift-4*, then select an area: Take a screenshot of an area and save it to the clipboard
* *Command-Control-Shift-4*, then space, then click a window: Take a screenshot of a window and save it to the clipboard

In Leopard and later, the following keys can be held down while selecting an area (via Command-Shift-4 or Command-Control-Shift-4):

* *Space*, to lock the size of the selected region and instead move it when the mouse moves
* *Shift*, to resize only one edge of the selected region

1. Compare the results of steps 1, 2, and 3 above for at least two magazines. What do data and the visualizations say about these two magazines? What kinds of questions do they raise that you might be interested in investigating further?

**Part II. The Whole Corpus (well of the 9 magazines offered on Sourceforge, anyway)**

Now that you have done some distant reading of at least two individual magazines in the MJP corpus, let’s look at the whole thing. Go to the URL for the chronological corpus of 9 magazines listed above.

1. What does the Summary panel tell you about the corpus in terms of word and document patterns? How does this compare with the magazines you examined in Part I?
2. What do the Summary and analytic tools tell you about the term(s) you investigated in Part I? What kinds of patterns or trends to you see, and how do these compare with the magazines you picked for Part I?
3. What kinds of questions for further investigation do these panels raise? What kinds of potential problems or pitfalls do you see in this approach?
4. Be sure to take screenshots of interesting graphs or patterns that you will want to place in your post.

**Part III. Exporting the Tools and Building a Skin**

*Embedding*

To embed a tool in the course site with your particular view of the data, click on the floppy disk icon in the top right of the tool panel and view the options. You may export the data as a static image for uploading to the site. You may provide a URL for the tool as a standalone entity. You may also select and HTML snippet, which will embed the tool directly into your post. Do do that, copy all of the HTML code that Voyant provides you, go to the post you’ve started on the course site, switch to Source mode, and paste the code where you wish the tool to appear. Switch back to WYSIWYG mode and click on the Preview to see if it looks right.

*Skin Building*

At the very top right of the screen, click on the floppy disk icon (in the blue bar across the top) and select the lower link for the Skin Builder application. This will open the corpus in a new tab or window and allow you to arrange a layout of tools, many of which do not appear in the default layout, for a customized analysis of the corpus. Drag and drop these tools where you’d like them to appear, and then click on the floppy disk (very top right) to export them with a URL. Paste it into your post so others can see what you did.