C. Johnston, 2026

DEPARTMENT OF HISTORY—WESTERN WASHINGTON UNIVERSITY

HIST 490/507: Digital Methods in History

Text Mining Activities

Pre-activity step - Create an account with JSTOR (if you don't have one already)

Voyant Text Analyzer

1. Download a copy of your Archives Worksheet Assignment from Canvas
2. Open the Voyant Tools website: https://voyant-tools.org/
3. Use the "Upload" feature and select the Bamman and Crane reading from Week 4
4. Examine the word cloud generated. Does this graphic represent the contents of the article (e.g., are the most important terms those that reflect the arguments / data discussed)? Are there any terms included that are not meaningful?
5. Now rerun the program using your Archive Worksheet assignment (I would suggest deleting the question texts from the worksheet template if it is still included in your document). Does the word cloud graphic represent the focus of your project? Are the terms included useful for further research?
6. Select one secondary source that most closely aligns with your research project (ideally an article that you have read). Rerun the program and examine the terms that are highlighted for this paper.

JSTOR Text Mining Using Constellate

1. Navigate to the JSTOR Homepage and select "Constellate" under the "Research Tools" menu
2. Select "Create a Dataset"
3. Login using your JSTOR login details (if you do not have a JSTOR login, create one now using your WWU address)
4. Select "Dataset Builder"
5. Enter 4–5 keywords to begin building a dataset. These could include the terms that were identified by Voyant from your Archive Worksheet or the secondary source you used. This could also include other terms
6. Adjust the filters as needed (e.g., if you are interested in results from specific time periods)

7. Take a few minutes and look at the histogram of publications over time. Are there periods of higher or lower publication activity?

8. Use the "Search Within Results" feature. Search related terms to see if there are subtopics of particular focus

9. Refine the search terms with any relevant additional terms identified in step 8. Then select "Build" and enter a name for this dataset

10. Download the dataset metadata. If you have less than 1500 documents, you can select the "Metadata (CSV)" file. If you have more than this, I suggest selecting the "Sampled Metadata (CSV)"

11. Open the CSV File. Save this as an XLSX File.

12. Name the current file "Download". Copy the file and name it "Titles". Open this second "Titles" file and delete all Column aside from the "Title" column (you can select all content using "Command+a", then unselecting the Title column before deleting the others)

13. Highlight all of the contents and use the "Search" function. Select the "Replace" option. Find and replace the following terms with nothing: Bibliography; References; Works Cited; Index; Untitled

Text Parsing Using CLAWS

1. Select the remaining Titles from the downloaded CSV file (you can quickly select all data by selecting the top cell, then holding down "Shift+Command" and clicking the down arrow)

2. Copy this content, and paste it into the entry box on the CLAWS Tagger Website

3. Copy the output, and paste it into a word document

4. Delete all Paragraph Marks by either:

    a. Using the "Replace" function, find all "^p" and "Replace All" with nothing entered in the replacement field

    b. Open the "Advanced Find and Replace" tool ("Command+F" → select "Replace" → select the tools dropdown menu → select "Advanced Find and Replace"). Select "Paragraph Mark" from the "Special" dropdown menu at the bottom right, then select "Replace All" with nothing entered in the replacement field

5. This will create one large text block. Copy this text and paste it into a new excel sheet

6. Select this cell, and under the "Data" tab, select "Text to Columns". Choose the "Delimited" option and click "Next". Select "Space" as the Delimiter and click "Finish". Each parsed component contents should now be in its own cell

7. Transpose this row of contents into a column by selecting all content cells (Shift+Command+right arrow). Copy the selected cells. Then select the first free cell (A2) and select "Paste Special" and make sure the "Transpose" option is selected. Then click "OK". You can now delete the original top row of content

8. Once again, select all cells in this column, and under the "Data" tab, select "Text to Columns". Choose the "Delimited" option and click "Next". Select "Other" as the Delimiter and type in the underscore (or copy it from one of the cells). Click "Finish". Each parsed component contents should now be in its own cell, with the POS code in the adjacent cell

9. Under the "Data" menu select the "Filter". You can now sort the contents based on POS code (you can open the CLAWS Tagset list here)

10. Take a few minutes to explore the contents using the filter and sort options and note down your observations. For example, select all the noun codes (NN0, NN1, NN2, NP0). What terms appear frequently? Is it possible to get a sense of the paper topics from the parsed titles? Are there any terms you want to add to your search list based on these observations

Western Libraries ONESEARCH

1. Create a final search string based on the terms identified through the activities above. Use this search string to look for additional sources related to your topic using the Western Library ONESEARCH database