Text Analysis Using Voyant Tools

Important Terminology

- **Corpus** (plural: corpora): a text or collection of texts that is used for analysis.
  - Example: One could create a corpus of all of Frederick Douglass's speeches to trace his language use over time.
- **nGram**: A continuous sequence of *n* items in a text.
  - For example, in Frederick Douglass's speeches, a bigram (or 2 continuous tokens) could be 'United States' and a trigram (3 tokens) could be 'justice for all'.
- **Stop words**: words that appear frequently in a language, like pronouns, prepositions, and basic verbs. These are often removed for computational analysis.
  - Some English stopwords include: a, the, she, he, I, me, us, of, is, would, could, should, etc.
- **Computational Text Analysis**: Text analysis is making inferences based on textual data. Computational text analysis (CTA) involves a computer drawing out patterns in a text, and a researcher interpreting those patterns. CTA includes methods such as word count frequency, nGrams, and sentiment analysis. CTA is similar to statistical analysis, but the data are texts.
- **Word Count Frequency:** Counting the total times a word appears in a text/corpus or the percentage of how often it appears.

Voyant Tools website: https://voyant-tools.org/

What is Voyant?

Voyant is a powerful web-based text analysis platform "designed to facilitate reading and interpretive practices for digital humanities students and scholars as well as for the general public." With Voyant, you can analyze one or more text files and use multiple visualization tools. Voyant can read .pdf and .doc, but .txt is recommended.

For Voyant documentation see: https://voyant-tools.org/docs/#!/guide/about

Using Voyant:

1. On the Voyant front page, click on 'upload' and navigate to the chosen corpus to upload text files—then click on 'reveal.'
   a. Alternatively paste URLs or full text into the textbox.
   b. Advanced options and help can be found in the top right of the text box.

2. The results are given in five default panes that can be changed.
3. To change a pane:
   a. Hover mouse over one of the panes.
   b. Click the panes button:
   c. Use the dropdown menu to navigate to different tools.
4. The "Trends" tool allows you to examine the usage for a word or set of words across a text or corpus.
   a. Appears in the top-right corner in the default Voyant results display.
   b. It will default to showing the most frequent (non-stopwords) terms in your input document(s).
      i. You may fill in any terms you would like to query in the search box.
   c. You can experiment with the settings using the gear icon.
5. The "Contexts" tool allows you to examine keywords in context.
   a. Appears in the bottom-right corner in the default results.
   b. You can input query terms in the search box and use the arrows to sort by the contexts to the left or right of your input term.
   c. This is not the most sophisticated way to examine words in context, but gives a sense of how this approach works.
6. The "Wordcloud" tool allows you to visualize the most frequently used words in a corpus.
   a. Appears in the top-left corner in the default results.
7. Voyant offers many other tools to explore. Read more about these in the documentation.