

Fundamentals of Text Mining: Analysis Activity

5: What Make a Hit Song?

Summary: In this activity we will explore different ways to examine and visualize your datasets using [Voyant](https://voyant-tools.org).

About Voyant

Voyant is a web-based reading and analysis environment for digital texts. It is freely available here: <https://voyant-tools.org>

The tool is open source, and is widely used by Digital Humanists using text analysis and visualization. It is an ideal tool to experiment with in order to familiarize yourself with the process of importing textual data, analyzing it, then creating and interpreting the results as visualizations. One drawback of the tool is that it is hosted on the McGill University servers, and so its ability to process very large datasets is limited. It is possible, however, to install it on a home or local server.

Key features include:

- Importing documents in various formats (plain text, HTML, XML, PDF, RTF, MS Word, ODF, etc.)
 - Several tools for studying term frequencies and distributions within documents and within a collection of documents (a corpus)
 - A full-text reader that supports very large texts and includes interactive features
 - Interaction between the tools that facilitates navigation and exploration at different scales (from "close reading" to "distant reading")
 - A mechanism for bookmarking and sharing instances of Voyant Tools (specific texts and tools) through persistent URLs
-

Exploring Voyant:

- **Upload to Voyant**
 - To upload a group of documents, you must first create a zip file of your dataset folder
 - Upload the zip file to Voyant:
 - From the landing page, select the zip folder you have just created.
 - Click 'upload'.
 - Voyant will do the work of expanding the archive and processing all of the documents in your dataset.
- **Understanding the Dashboard View**
 - Familiarize yourself with the dashboard.

- List three pieces of information about your content set that you can see at a glance from the dashboard view.
 - What are your overall impressions of the Voyant dashboard? Do you find it intuitive and user friendly? If not, what do you find unclear or challenging?
 - What is a stopwords? [hint: read the 'help' documentation]
 - How would you add a tool to the dashboard that is not included in this default dashboard view?
 - **Voyant Suite of Tools**
 - Voyant provides a range of tools and options for text analysis. What information can you learn from the following tools and visualizations? [hint: Voyant help documentation is useful].
 - Cirrus
 - Document Terms
 - Mandala
 - Contexts
 - Choose your own tool
-

Explore your Dataset using Voyant

- Now you'll have an opportunity to explore your datasets using the tools embedded in Voyant. The goal is for you to experiment with your data, to customize tool options and to create a visualization or two.
- **Your dataset**
 - In this dataset, you will find the song lyrics for the Billboard #1 hits for four artists or bands who have had the most #1 hits on the Billboard charts.
 - Beatles: 20 No. 1s
 - Michael Jackson: 13 No. 1s
 - Mariah Carey: 18 No. 1s
 - Madonna: 12 No. 1s
- **What makes a hit song?**
 - **Compare the MALE artists**
Objectives:
 - Explore The Beatles lyrics as one dataset
 - Explore the Michael Jackson lyrics as one dataset
 - Explore both datasets together
 - **Most Frequent Words comparison**
 - Open two Voyant windows:
 1. Load The Beatles lyrics in one window, and load the Michael Jackson lyrics in the other Voyant window. Compare the word clouds.
 2. If there are any words that do not seem meaningful to you in the "Cirrus" tool, hover on the top right corner of the panel and click on "Define options for this tool" to remove those words from the list.
 3. What does each data visualization tell you about the songs?

4. Are there any commonalities amongst the two word clouds? What are the differences?
 5. Identify two or three words in “Cirrus” tool that interests you. Now type them into the “Trends” tool (hint: using a comma after the word will let you enter multiple). What does adding multiple terms into the “Trends” tool reveal?
 6. How might you export these visualizations to a blog, social media, or an academic paper?
- Now, open a 3rd Voyant window:
 1. Load both datasets into the window (i.e combine The Beatles lyrics and the Michael Jackson lyrics)
 2. If there are any words that do not seem meaningful to you in the “Cirrus” tool, hover on the top right corner of the panel and click on “Define options for this tool” to remove those words from the list.
 3. Are there differences between the wordcloud for the individual artists vs. the two artists combined?
 4. If you want to continue exploring, find the four-box icon for the upper right panel to select a different visualization tool for that panel. The index of available tools describes what each option can do.
- **Compare the FEMALE artists**

Objectives:

 - Explore Mariah Carey’s lyrics as one dataset
 - Explore Madonna’s lyrics as one dataset
 - Explore both datasets together
 - **Most Frequent Words comparison**
 - Open two Voyant windows:
 1. Load Mariah Carey’s lyrics in one window, and load Madonna’s lyrics in the other Voyant window. Compare the word clouds.
 2. If there are any words that do not seem meaningful to you in the “Cirrus” tool, hover on the top right corner of the panel and click on “Define options for this tool” to remove those words from the list
 3. What does each data visualization tell you about the songs?
 4. Are there any commonalities amongst the two word clouds? What are the differences?
 5. Identify two or three words in “Cirrus” tool that interests you. Now type them into the “Trends” tool (hint: using a comma after the word will let you enter multiple). What does adding multiple terms into the “Trends” tool reveal?
 6. How might you export these visualizations to a blog, social media, or an academic paper?
 - Now, open a 3rd Voyant window:

1. Load both datasets into the window (i.e combine Mariah Carey and Madonna's lyrics)
2. If there are any words that do not seem meaningful to you in the "Cirrus" tool, hover on the top right corner of the panel and click on "Define options for this tool" to remove those words from the list.
3. Are there differences between the wordcloud for the individual artists vs. the two artists combined?
4. If you finish early, find the four-box icon for the upper right panel to select a different visualization tool for that panel. The index of available tools describes what each option can do.

- What have you discovered through your visualizations?
- What are the differences between the lyrics of male and female artists?
- Are there any commonalities?
- Can you identify any trends that might suggest why these songs were so popular?