

TEXT ANALYSIS WITH VOYANT

Voyant is a tool that allows you to do multiple text analyses at once. This document provides links to the data we'll use during the session and step-by-step instructions for today's activities. I also provide information and links to help you install and explore Voyant after the session.

Data

All data for this workshop is available in this Dropbox folder: <http://bit.ly/2dROUfV>

Smaller text sources:

Noah's Flood: Genesis 6-9, King James Version of the Bible via [Gutenberg Books](#)

Utnapishtim's Flood: Epic of Gilgamesh via [Internet Archive](#)

Large corpus source:

Full collection download from the [North American Slave Narratives](#) project.

Voyant: To Download or Not to Download?

There are two ways to use Voyant:

1. No installation required: A web-based tool (www.voyant-tools.org) using remote servers to crunch data. Depending on your computer and internet connection, it may not handle large files well.
2. Installation required: A browser-based tool using a local server installed on your machine. The server is a Java application. This version typically performs better with large data files. We will be working with Voyant 2.1 for this morning's session. If you have not already done so, you can download Voyant from GitHub here: <http://bit.ly/2dhfI7a>.

Our two exercises today will allow you to explore both versions of Voyant. We'll start with the web-based tool that requires no installation.



Text Analysis with Voyant by [Abby Scheel](#) is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](#).

Exercise 1: Web-based Tool (no installation)

1. From the [Dropbox folder for this workshop](#), download the files Genesis6_9_GutenbergKJV.txt and GilgameshFlood_clean.txt.
2. Go to www.voyant-tools.org
3. Click the Upload button. Navigate to the folder you created in step one and use Shift to select both files. Click open.
4. Voyant will present you with five tools:
 - a. Cirrus: a word cloud that displays the highest frequency terms in the corpus.
 - b. Reader: an infinite scrolling reader for the actual text in the corpus (this fetches the next part of the text as needed).
 - c. Trends: a visualization of word frequency across the corpus (or it can be re-launched for word frequency within each document)
 - d. Summary: a high-level summary of data from the corpus.
 - e. Contexts: a listing of each occurrence of keywords with some of the associated text (context).
5. The first thing we'll want to do is see if there are any non-essential words that are affecting the output. Voyant has already applied a default stop-word list to omit words like the, an, and. There may be other words that were not caught by the default list or that are not interesting for us in this case. For instance, shall is the third most common term but is not a terribly interesting word. Let's add it to the stopword list to omit it from the analysis.
 - a. Hover over the Help icon (question mark) in the upper right hand corner of the Cirrus panel to open a menu. Click on the slider icon next to the question mark to open the options menu.
 - b. The Auto-detect list should be selected next to Stopwords. To add a term to the stopword list click Edit List next to it.
 - c. In the stopwords editor, hit enter to start a new line and add shall. Click save.
 - d. Make sure the box is checked next to 'apply globally' to apply the change to all the tools in Voyant. If this box is not checked, the stop words will only be applied to the tool you opened.
 - e. Click Confirm and the panels should refresh. The word shall should disappear from all the modules.
6. The menu in the upper right corner of each panel includes two more options in addition to the options menu:
 - a. Replacement (four squares): an option to select another tool to replace the current one in the panel. A tools index is available in the [Voyant Tools Documentation](#).
 - b. Export (box with an arrow): options within this vary by tool, but you can export the results to another tool, to a shareable URL, export the image, or a bibliographic reference for the results.



7. Each of the default tools also has one or two alternate tools you can easily switch between in that panel. The Help icon will explain each tool.
8. Explore what the Voyant tools can do. You can click on individual words in Cirrus to focus on that word in other panels. In the Context panel, type a word of interest in the text box to search for the context of that word.
9. What tools in Voyant might help explore these questions? (Recommendations are available at the end of this document.)
 - a. Ark is one of the most common terms in the Genesis story while ship is the term used in the Epic of Gilgamesh. Are these words used similarly in the two accounts?
 - b. Noah is the main figure of the Genesis story. Utnapishtim is the survivor in the Epic of Gilgamesh. What can you guess about the narrative perspective of each of the accounts based on how often and where their names appear in each account?
10. Note that there are other ways you could analyze texts with Voyant. For instance, you could enter each flood account separately into Voyant and compare the resulting data and visualizations. You could also compare the Genesis or Gilgamesh accounts to a broader range of flood stories gathered together in a corpus.

Exercise 2: Voyant Server (requires installation)

Important note: Before exiting the Voyant Server, click Stop Server! I'll explain why later, but it's important enough to say so from the outset.

1. Open Voyant Server on your computer
2. When you open the Voyant Server, the server will open and then automatically open the browser-based tool.
3. You have the option of adjusting the amount of memory allocated to the Voyant Server. The default is usually 1024 MB. If you find that Voyant is taking a while to process you may want to increase the amount of memory allocated. Let's go ahead and increase the amount of memory for this exercise. In the server window, click Stop Server. Enter a new amount of memory (2048 MB should work for today) then click Start Server. Click Open Web to re-open the web-based tool if needed.
4. We'll use Voyant Server to analyse all the narratives in the North American Slave Narratives Project. The compressed file is available in the Dropbox folder for this workshop.
5. In the Voyant web portal, click Upload then click Add. Navigate to the folder where you downloaded the files for this workshop and select the na-slave-narratives-text compressed folder. Click Reveal. --This will take longer to process than exercise 1.--
 - a. Note: Voyant only uploads compressed folders.
6. The Voyant tools interface is very similar to the web-based version. You should have access to all the same tools and functions.



7. Explore what Voyant can do with hundreds of texts. In the Summary panel, choose one of the files under Distinctive words and click on the file name. Notice that the Reader panel now displays that text and the Trends panel displays frequencies for that text. Change the Contexts panel to focus on this text too by clicking on Scale > Documents > select your document title.
8. What tools in Voyant might help explore these questions? (Recommendations are available at the end of this document.)
 - a. How are words like slave, master, freedom, enslavement, and race connected?
 - b. In which narratives are children referred to the most? Which children are being referenced?
9. VERY IMPORTANT: On the server window, click Stop Server before closing. Otherwise the server will continue to run and eat up your computer's memory.

Responses to questions

Flood Myths

1. Ark is one of the most common terms in the Genesis story while ship is the term used in the Epic of Gilgamesh. Are these words used similarly in the two accounts?
 - a. Contexts: Can help understand the context of term occurrences. Search for 'ship' and 'ark' to see how the terms are used.
 - b. Collocates: Allows you to see which terms are often used together. Search for 'ark' then 'ship' to see if they share any collocates.
2. Noah is the main figure of the Genesis story. Utnapishtim is the survivor in the Epic of Gilgamesh. What can you guess about the narrative perspective of each of the accounts based on how often and where their names appear in each account?
 - a. Reader: Search and read the full text. Also provides word frequency graphs over the course of the text. Search for 'Noah' and 'Utnapishtim' to see at which points during the story each figure is referred to.

North American Slave Narratives

1. How are words like slave, master, freedom, enslavement, and race connected?
 - a. Links: Shows connections between terms. Enter each of the terms in the Links box, hitting enter after each one. Increase the number of connected terms with the Context slider.
2. In which narratives are children referred to the most? Whose children are being referenced?
 - a. Document Terms: Can help identify the document with the greatest number of occurrences of a term. Search for 'children' and then sort the list by Count.
 - b. Bubblelines: Can help trace the usage of a term within a document. Search for 'children' and hide the other terms. In the Documents menu select the



document(s) with the highest number of occurrences (from the Document Terms panel).

- c. Contexts: Can help understand the context of term occurrences. Search for 'children.' Then specify a document in the Scale menu.

More Info and Credits

This exercise draws heavily from an exercise from the Humanities Institute for Learning and Teaching over the summer. The Voyant exercise prepared by Brandon Locke, Thomas Padilla, and Dean Rehberger from Michigan State University is excellent and has more information about Voyant and its features: <http://bit.ly/1KBZj3I>. They also provide a few scenarios that will allow you to explore the functionality of Voyant further.

In addition, Emory Libraries has posted a wonderful explanation of how Voyant can be used to analyze a set of texts by looking at the two sermons from the Martyred President Project: <http://bit.ly/1Js3MFa>.

For even more information about Voyant, visit the Voyant Tools Documentation: <http://docs.voyant-tools.org/>



Text Analysis with Voyant by [Abby Scheel](#) is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](#).