

Fundamentals of Text Mining: Analysis Activity 6 - Religion

Summary: In this activity we will explore different ways to examine and visualize your datasets using [Voyant](#).

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

Instructions:

- **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**
 - 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 Sunday School Books in Nineteenth Century America dataset, consisting of 166 texts, including Sunday school books published between 1809 and 1887.¹
 - The file-naming convention is yearpublished_partial title. There is also a collection of undated texts.
 - The material reflects the emerging diversity of Protestant Christian denominations in the United States during that period. Additionally, texts included also mark the appearance of a theologically inflected genre of juvenile literature, which was published by a variety of sectarian presses.
 - More contextual information is available here:
<https://digital.lib.msu.edu/projects/ssb/?action=introessay>
- **Most Frequent Words comparison**
 - Open two Voyant windows:
 - Load your entire content set in one window, and load a single text in the other Voyant window. Compare the word clouds.

¹ Sunday School Books in 19th Century America. East Lansing: Michigan State University Libraries Special Collections. <https://www.lib.msu.edu/ssbdata/>. Dataset: <https://lib.msu.edu/ssbdata/>

- Are the most frequently used words in the single document the same ones that appear most frequently in the larger corpus? Describe any differences you observe.
- **Decade-by-decade comparison**
 - To work efficiently on this comparison, consider splitting the work between each group member, then compare your analysis results. Consider which terms appear most frequently decade-by-decade. Do you see any recurrent words, phrases or patterns emerging?
- **Thematic Analysis**

Researchers have identified a number of themes in the Sunday School datasets. We'll focus on two: animals/natural world and moral tales. There are two subsets for each theme, and relevant texts have been extracted for you to explore using Voyant.

Animals/Natural World

Background: *Sunday school teachers dedicated a good deal of their time discussing with students the world around them. In this section one will find three types of texts, those providing more or less straightforward information about animals and plants or natural wonders (e.g. Child's Own Book of Natural History), those offering lessons of moral conduct with respect to animals and the natural world (e.g. Kindness to Animals: or, The Sin of Cruelty Exposed and Rebuked), and those that combine the two (e.g. Talks with Amy Dudley; or, What Makes Me Grow?). The first type tended to rely on facts and illustration, the other types on scripture and precept.*

*While one would best look elsewhere for popular debates about evolution one will find that advances in natural science are of explicit concern in these books. Most authors go out of their way to reconcile scientific understanding of the universe with knowledge of God. For example, in *What is a Star?* there is a dialogue discussing the astronomical discoveries of Sir John Hershel. "Dear Mamma," the boy exclaims, "if all these stars are worlds like are own earth and other planet how much God must have to take care of!"*

"The discoveries made by the telescopes in these boundless regions, my dear, are well suited to exalt our notions of God's wondrous power."

Other works attempt to reconcile theories of the physical body with man's spiritual nature and provide unusual commentary on such subjects as phrenology, and nervous sympathy. In general, these works seek to ascribe a moral content to nature to man's physical conditions.²

Animals/Natural World Subsets:

- What is a Star?

² <https://digital.lib.msu.edu/projects/ssb/index.cfm?CollectionID=54>

- Man, in his Physical, Intellectual, Social, and Moral Relations.

Moral tales

Background: *Relatively scarce in the early years of the United States, moral tales were increasingly popular in the nineteenth century (see Introduction). Frequently based on a line of scripture of a moral precept, these works functioned to convey the ideals of virtuous conduct (obedience, forgiveness, and responsibility) through narrative means. In this sense, they are companions to conduct literature. Stories appealed to children through imaginative and emotional identification in a way, frequently representing scenes of piety and conversion. Most denominations tailored these stories to suit their own doctrinal needs.*³

Moral tales subsets:

- Advice to Young Men on their Duties and Conduct in Life.
- First Flight From the Nest; or, Nellie's Six Months in Boston.

Questions to consider:

- Can you identify the context for words related to specific themes?
- How does the word frequency vary between the subset at large, and individual documents? Can you identify trends over time?

If you want to explore further, 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.

³ <https://digital.lib.msu.edu/projects/ssb/index.cfm?CollectionID=60>