

Kathryn Hite – Mini Project 3: Writeup and Reflection

Project Overview

The goal of this project was to determine the positive or negative sentiment of books within the four periods of Ancient Roman literature to illustrate the sentiment of the civilization during these periods. The books were pulled from Project Gutenberg based on a list of prominent authors during each age (Early, Golden, Augustan, and Imperial). These books were processed using the sentiment analysis tool and an average sentiment was computed for each period. The visualization of this information can be used to illustrate the sentiment of Rome during each time period as related to the historical events of the time.

Implementation

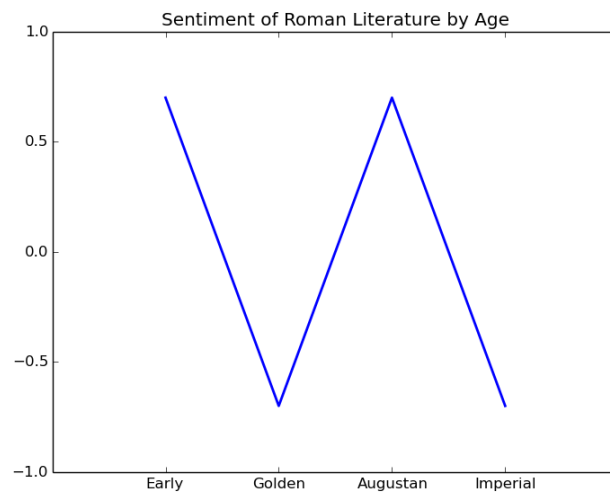
All of the book data was pulled from Project Gutenberg based on a dictionary of time periods with a list of authors for each time period. This allows the analysis to be scaled up to as many authors and categories as the user would like to input. To overcome the Project Gutenberg download limit the texts are saved to a file when they are downloaded so that they can be loaded later to avoid downloading the same book more than once. I chose to implement the choice to run the program in the saved text mode or not using a global variable that the user can change to indicate whether it is necessary to download the books before processing them. This provides an efficient way to change the mode without having to change the function or search for the variable.

The texts are analyzed using linguistic post-processing to determine the sentiment of each individual text. These sentiment values are used to compute an average sentiment for each age and the authors within the category. This data is plotted to provide a visualization of how the sentiment changed from each period to the next, which can be mapped to the historical context of each period.

Results

The figure below shows the test results used to ensure that data could be loaded and processed correctly. The Early and Augustan Periods were given “good” in the text file associated with each author, and the Golden and Augustan Periods were given “bad” in each

author's file. The plot shows that the data was accurately analyzed to determine the average sentiment for each period.



Ninja Note: The project Gutenberg site went down completely Wednesday (I was unable to access it from any computer I tried on campus), so I wasn't able to run with the actual texts. I did confirm that the downloading code was working before the site went down though. As soon as I can get access to the texts, I'll update this section with the actual results.

Reflection

I believe the code was well planned and laid out, and the documentation provides all of the necessary information to understand the process. In the future, I would ensure that I had all of the necessary testing data as soon as possible to avoid being affected by the Gutenberg site going down. The project was appropriately scoped to provide functionality to allow a user to visualize the data in a helpful way. A larger scope for future work would be to search Wikipedia for the authors, sort them into periods based on the years they lived, and then use links to their texts as the data. This would enable using a larger data set without searching for more authors.