

Walter Sellars

Csci4140 – Assign 1.

Jan 30, 2019

The genres that were selected were government, religion, lore, news, and humor. I chose these because I wanted to look at the content of documents that would be distributed as “formal” works and compare them to informal cultural works to determine how often the “quality” of life was discussed.

I started by looking at the set of words [life, liberty, pursue, work, poor, children, and death] with the assumption that the gov’t collection would focus exclusively on positive concepts, but ignore intangible ones like happiness. After initial testing, I found that some of my choices of words were not getting any hits, but words like life, work, children, and love were heavily populated. I decided to switch my set of words to [life, freedom, love, wealth, happy, work, poor, children], and found the changes to be lackluster. I changed wealth to money, and added growth to get more keyword hits across possible terminology. This gave a better split in the data and more information as to what the different documents focused on.

The noticeable trend started to appear, with government documents focusing less on personal positive qualities of life, and more on work and growth, which I assume is meant economically. Things like love, happy, and poor are also noticeably missing many entries, while freedom gets more hits than any other of the words. I supposed freedom sells?

