Rowan Sharman
SoftDes MP 3
2/23/17

Project Overview:

For this project I wanted to objectively compare supposedly liberal-leaning Wikipedia with its right-wing counterpart Conservapedia to identify any bias in either of them. To do this, I scraped the HTML text from the pages for Barack Obama and Donald Trump on both sites and then analyzed each one for positive and negative sentiment using Vader Sentiment Analysis.

Implementation:

My code must be run in four parts in order for it to function correctly. First, TextMiner.py pulls the HTML text from the pages for Barack Obama and Donald Trump on wikipedia.org and conservapedia.com, and stores each result by pickling it. This could be further generalized by adding functionality to input lists of URLs to scrape. The next part of the program, TextFormatter.py looks at the pickled objects and parses them for the relevant parts of the page, removing titles, sites, links, and any other extraneous information. It also removes the embedded HTML tags, leaving a clean string that is easy to analyze. I chose to have this part of the program run separately from the analysis because it really only needs to run once, so allowing the anlysis part of the program to just analyze a string makes the program faster.

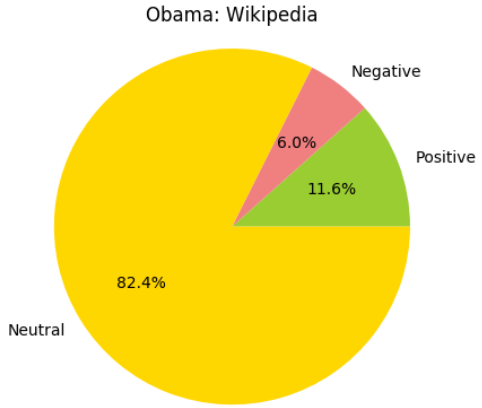
Next, TextAnalyzer.py uses Vader Sentiment Analysis to find positive, negative, and neutral sentiment in the page, then prints the results in a readable fashion. It also pickles the result again for use by the next function. GraphMaker.py reads the pickled result of the analysis and plots it in four pie charts which show up in a graphics window.

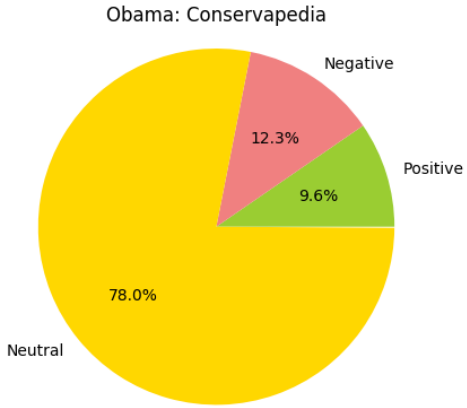
Results:

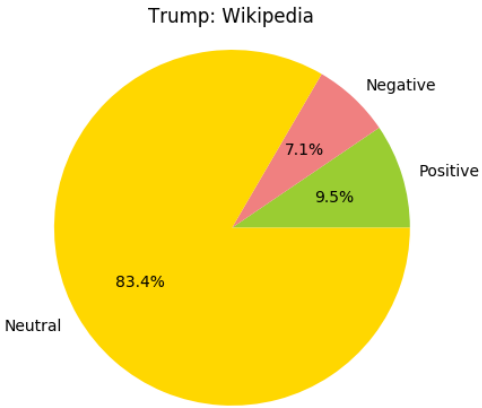
I found, not too surprisingly, that Wikipedia seems to look more favorably on Obama than Conservapedia, and Conservapedia looks more favorably on Trump than Wikipedia. Both sources also show some amount of bias in that Conservapedia looks more favorably on Trump than it does on Obama, and Wikipedia looks more favorably on Obama than Trump. The plots are shown on the next page.

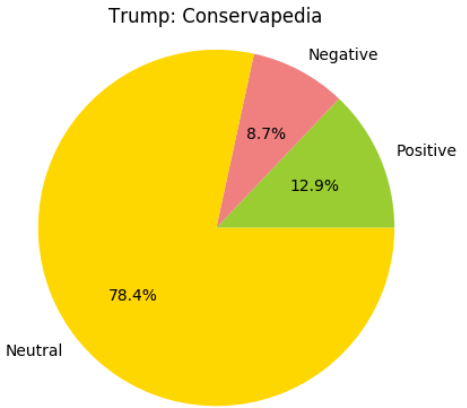
Reflection:

I felt that this project went fairly well, but I wish I had started earlier so that I could have added some functionality. I would have liked to add a database of more politicians to analyze and look at a larger sample size. I also found myself writing code and then later putting it into a function. If I had been able to think ahead and plan more I would have saved myself some of this labor. My unit testing, while not necessarily apparent in my code, works well for me because I use incremental development. I definitely learned about some very useful tools like matplotlib, pickle, and BeautifulSoup which I'm sure I will use in the future.


Obama: Wikipedia
Negative
6.0%
Positive
11.6%
82.4%
Neutral


Obama: Conservapedia
Negative
12.3%
Positive
9.6%
78.0%
Neutral


Trump: Wikipedia
Negative
7.1%
Positive
9.5%
83.4%
Neutral


Trump: Conservapedia
Negative
8.7%
Positive
12.9%
78.4%
Neutral