

Bryan Holcomb  
CS 410: Text Information Systems  
Project Proposal

Team Member: Bryan Holcomb (bryanph2) as individual

Topic: Sentiment Analysis of News Articles Around the 2020 US Presidential Election

For my course project, I plan to do a sentiment analysis of news articles around the 2020 US Presidential Election. There are a number of applications for which we can use this sentiment analysis, including an evaluation/comparison of biases between different news sources, as well as measure the general opinion of the candidates in different audiences and demographics. Additionally, we can evaluate how sentiment towards each candidate has changed over time.

For this analysis, I plan to build a web crawler to obtain a large sample of articles from various websites, including large national news outlets (CNN, NBC, ABC, Fox News), at least one international source (i.e. BBC), and various newspapers from major cities around the country. For each article, I will track the date and source, and then parse each article and perform a sentiment analysis. I plan to use a number of packages to do the work, including MeTa, NLTK, BeautifulSoup, and IBM Watson.

The vast majority, if not 100% of the project, will be done in Python.

Hours Breakdown:

Web Crawler design and setup: ~5 hours

- This could take longer depending on how much customization is needed for each source I use, and depending on my ability to re-use the same procedures and my ability to run my scrapers behind paywalls.

Parsing and database setup: ~5 hours

- Once I have the articles saved as documents, I believe parsing should be very straightforward, as done in lecture and the MPs.

Design of Sentiment Analysis code: ~7 hours

- This will take the bulk of my effort. I believe the most challenging part will be to divide articles that discuss both candidates and group each sentiment with the correct candidate.

Statistical analysis of results: ~3 hours