Ignacio Streuly
Natural Language Processing
Professor Adam Meyers
Final Project Proposal

Political Sentiment Analysis

For my final project I'd like to create a system that does a political analysis of a person's writing and assigns them a score for their leanings, e.g. Liberal, Conservative. The idea is that the same system can be applied to a colloquially written paragraph or a Facebook dump of a user's data - inherently the more content provided the more the system should be able to assign an accurate score.

A large problem that I'm currently encountering is finding data already in formatted, or scraping a data from online, that could be a reliable as a training data set. The training set really determines the accuracy of how well it can predict anything at all so ideally I'd want to find a set that contains writings by many members of congress. Unfortunately the congressional archives aren't easy to navigate or find, and what may end up happening is that the training data will simply have to be a conglomeration of presidential speeches, campaign speeches by Trump and others that were in the running for the 2016 president election, and any other data that can be of use. I could ideally use help in this region of my research, as it's probably the more critical part of my project that's not completely in my control.

As far as the programming, most of the methods and algorithms that I would use depend on the nature of the data set that I end up assembling. In its simplest form, I'd rely heavily on Feature Selection, the system would train on n-grams specific to each political leaning. Beyond that, my idea is to implement a modified beam search on submitted documents.

I'd be more than willing to collaborate with other students that have similar projects. None of my work for this is proprietary, or owned by an employer. Ideally I'd like to put this online and have it be something that goes beyond the scope of this class.