CMSC 12200
Proposal: What's My Bias? - A Google Chrome Extension
Team: Error 418
Date: 01/24/2017

Stated Objective of the Software:
Our goal is to build a Google Chrome browser extension that uses natural language processing to dynamically analyze the text of a given webpage. First, the extension will look for proper names of political figures like Donald Trump, Paul Ryan, and Elizabeth Warren. Second, it will use sentiment analysis of the sentences in which these names appear to determine whether the author has a positive or negative view of the political figures. Third, it will compare the pattern of positive and negative sentiments to the placement of the figures along liberal-conservative lines (as evaluated by either a neutral organization like GovTrack, or a pair of think tanks with opposite ideologies). The result will be an overall ideological score for the page. It will display this score with a small popup. Upon clicking the icon of the extension, it will explain to the client what names it found and how it scored each sentiment.

Also included in the extension will be a formality index generator. This will take the text of the HTML page and compare each word to a dictionary where a word is mapped to either a 0 (for a formal word) or a 1 (for an informal word). This dictionary will be scraped from the Oxford Dictionary or some other online dictionary that includes whether a word is formal or informal. The fraction of words that are informal will be reported as the informality index of the page. The extension will present this score, along with guidelines on how to interpret it. We will generate these guidelines by comparing results across a large number of webpages. This tool will be most useful for news articles, so this will be the type of page we will focus on.

Work Required:
We believe the work could be divided into three large categories: data collection, data analysis, and presentation of analysis. We will refer to the three categories as phases A, B, and C respectively.

Phase A will require us to explore the capabilities of JavaScript and Chrome extensions by watching tutorials and researching various packages and their documentation. We have already begun work on this phase. In particular, within the Node.js runtime environment, we have found packages called Sentiment and Natural that may be useful in determining the sentiment of sentences. Once we determine the intended structure of our code and the packages required, Phase B will require us to write code integrating the capabilities of these packages into our tasks. Finally, in Phase C we will ensure that the extension operates reliably, quickly, and in a way that is attractive to users.

Timeline/Deadlines:

Provided below is a breakdown of our work plan and timeline by academic weeks:

Week 4 (01/23/2017 - 01/29/2017): Proposal Presentation
Week 5 (01/30/2017 - 02/04/2017): Phase A
Week 6 (02/05/2017 - 02/12/2017): Phase A (to be completed by the end of this week) & Phase B
Week 7 (02/13/2017 - 02/19/2017): Phase B & Phase C
Week 8 (02/20/2017 - 02/26/2017): Phase B (As a conservative estimate, we foresee ourselves working on Phase B till the end of 8th week to ensure correct implementation) & Phase C
Week 9 (02/27/2017 - 03/05/2017): Phase C & Debugging
Week 10 (03/06/2017 - 03/12/2017): Debugging & Refining

On a more general level, we intend to develop a working product by 9th week and use the remaining time to determine where we need to debug and improve the program so that the Chrome Extension serves its original intended purposes and hopefully more. Our work and progress on each Phase will not necessarily be sequential but rather parallel/simultaneous. We plan to work on multiple phases at a time in order ensure timely progress and to contemplate and address the relationships between different Phases to offer a well-integrated product at the end.