

# CSE 467 Project

## Tokenization and Sentiment Analysis of Trump’s Tweets

Author(s)   Aniruddha Nandi,Matthew Kirshy



### Abstract

**Key words**   ·   **tokenization, corpus generation and parsing, visualization, sentiment analysis**

Data and codes will be available at <https://github.com/anandimous/467-tokenization-sentiment-analysis.git>

Our weekly blog is also avaiable at the same url as the github repository under the 'Wiki' section

### Proposed Methods

We would like to download and analyze all of Trumps tweets (as sourced from <http://trumptwitterarchive.com/archive> which contains over 35,000 entries for us to analyze as of 10/23/2018). We will pre-process (tokenize and create a corpus) of all of the tweets so that we can further analyze the language he uses. We will use the corpus to analyze the sentiment of individual tweets against buzz words such as "guns", "women", "immigrants", etc to see how his sentiment changes as newsworthy events have happened in the past. We will produce a couple of interactive graphs based on the results sourced from the above. In addition, we can add visualizations in our graphs by using brat (<http://brat.nlplab.org>) to include structured annotations on Trump’s tweets.

...

### Results

As mentioned under proposed mentods, we intend to have some interactive graphs - one will be a heat map of his frequency of tweets per day that can be clicked to display the the unprocessed tweet, processed version of the tweet (tokenized and pre-processed), buzz words contained in the tweet, and the overall sentiment of the tweet. The second will be a graph that will display his tweets on a timeline but will be a in the form of a line graph with multiple lines in respect to the above mentioned buzz words. We intend to see how much his opinion on certain topics change with respect to the news during that time. Finally, we can use brat to provide structured annotation on some of his tweets to further understand the most prevalent sequence of linguistic attributes in the same

...