

IST687 – Text Mining Lab

Step 1: Read in the positive and negative word files

- 1) Create two vectors of words, one for the positive words and one for the negative words
- 2) Note that when reading in the word files, there might be lines at the start and/or the end that will need to be removed (i.e. you should clean your dataset).

Step 2: Process in the MLK speech

- 3) Read the text file
- 4) Create a term matrix
- 5) Create a list of counts for each word

Step 3: Determine how many positive words were in the speech

- 6) Scale the number based on the total number of words in the speech
- 7) Hint: one way to do this is to use the 'match' function and then 'which' function

Step 4: Determine how many negative words were in the speech

- 8) Scale the number based on the total number of words in the speech
- 9) Hint: this is basically the same as step (3)

Step 5: Redo the 'positive' and 'negative' calculations for each 25% of the speech

- 10) Compare the results (ex. a simple bar chart of the 4 numbers).