Twitter Sentiment Analysis of the Brexit result

Surya Vajjhala
U19590925
Email vajjhala@bu.edu

Zhi Dou
U21392913
Email zhidou@bu.edu

Progress Report 1

The data was scraped from the twitter web-page itself as there were restrictions on the data available through the API (only tweets in the past 7 days are accessible)
The total dataset would be very large several GB but for now we have provided only a sample.

The sample output data is in csv format and is only a sample. The complete data has yet to be scrapped.
This sample consists of all tweets with $\#Brexit$ posted on Twitter between 16:09 to 17:34 on 2016-06-24 . This is just after the results of the Brexit vote.(which came out on June 23rd)

Data Attributes:

The tweets were scraped and then using regular expressions unnecessary information in the tweets was removed . $TextBlob$ library was use to do search for key words(using noun_phrases method of $TextBlob$ library) and analyze the polarity of the tweets . The attributes that were saved :

- 'noun_phrases ': All the nouns present in the text

- 'geo': The location of the tweet

- 'polarity': The measure of the sentiment of the tweet. It ranges from -1 to 1 where 1 is positive sentiment and -1 is negative sentiment

- 'subjectivity': Measure of the subjectivity of the tweet, ranges form 0 to 1, where 0.0 is very objective and 1.0 is very subjective.

- 'time': The datetime object of the time when the tweet was posted

- 'hashtags': The other hashtags associated with the tweet. These will be used to do further analysis and determining the content of the tweet