

CS410 - Text Information Systems - Progress Report

Sentiment Analysis of Tweets.

Task Status

Task Name	Hours	Status
Data Curation and Pre Processing	20	Complete
Build User Interface for Sentiment Analysis	30	In Progress
Sentiment Analysis Methods	30	In Progress
Integration with UI	20	In Progress
Preparing the presentation and documentation	20	In Progress

Current Challenges

For data curation we are using Twitter API version 1.0 to gather data so that we can train the different methods for sentiment analysis. At this time we are being limited by the throttling that Twitter applies while calling the API. We will have to create multiple files with 1000 records and build a bigger file with concatenated data.

Detailed Progress Update

Curate Dataset

We are planning to run the dataset acquisition step multiple times to gather data from twitter and then apply methods to process the dataset and make it suitable for further processing. The script to connect to twitter and fetch data has been completed and we are working on adding functions to process the data. The functions will specifically include steps as below:

1. Apply Stemmer
2. Apply Lemmatizer
3. Split data into training and test subsets