Tweetlytics - An end to end IR System

Team Name: CSK

Members:
Baalaji Varadhan Sivakumar - 50287414
Praveenkumar Rajendran - 50289568
Sanjay Bandlamudi Shailendra - 50290963
Sriram Venkataramanan - 50289666
Vijay Jagannathan - 50290947

Abstract:

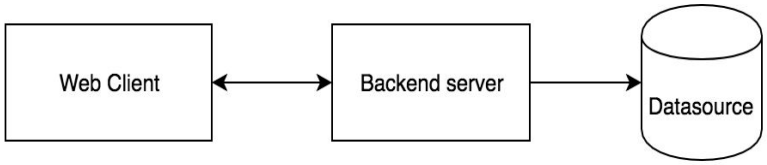
This project is an attempt to build an end to end IR system that lets the user to search and derive insights from social data collected from Twitter.

Dataset:

We have a multilingual dataset including English, Hindi, Thai, Spanish, French collected from cities New York, Delhi, Bangkok, Paris and Mexico over 5 topics Politics, Crime, Infrastructure, Social Unrest and Environment. The dataset contains over 40,000 tweets collected over a period of 2 weeks.

Architecture:

The System has three components, Client, Server, datasource

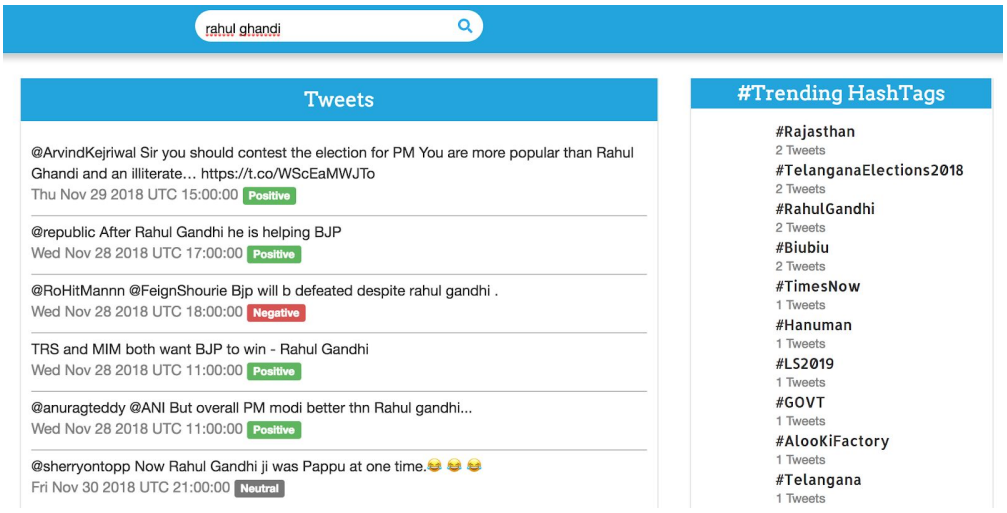


Tech Stack:

Component	**Technology**
Web Client	HTML, CSS, JavaScript, Jquery, Bootstrap, Google charts
Backend Server	Java8, Spring Boot, SolrJ, Google translate
Datasource	Solr 6.6.5

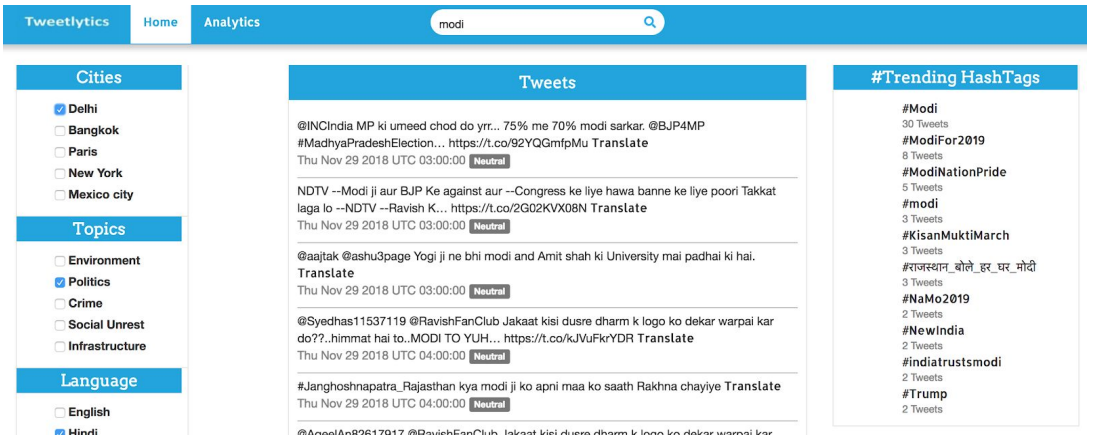
Features:

1) **Global Search:**
   Global search enables the user to search across the entire dataset. The search retrieves relevant tweets as well as the trending hashtags corresponding to the search term



2) Faceted Search:

The faceted search allows the user to further drill down the search results, the results could be filtered across Cities, Languages, Topics and Sentiments.

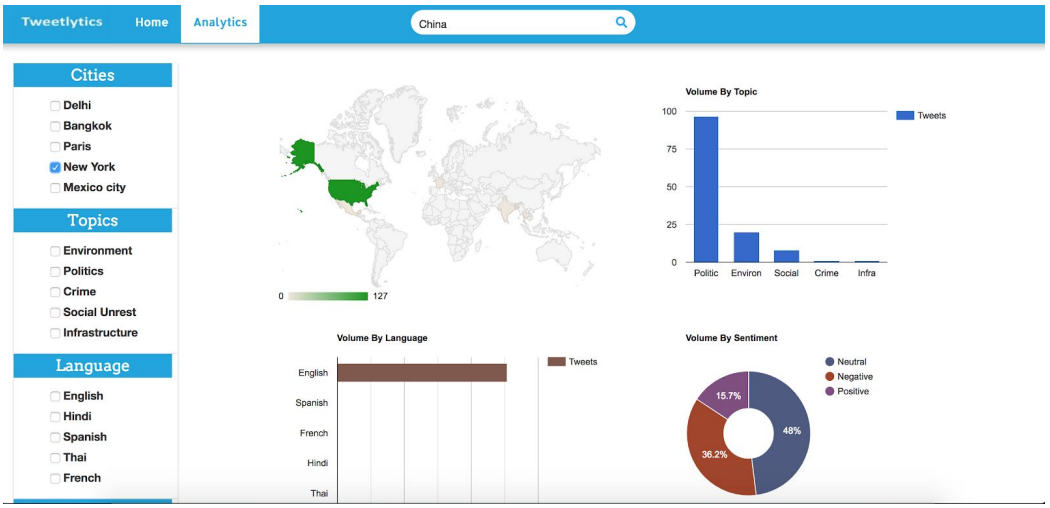


3) Trending Hashtags:

The trending hashtags for a particular search is based on the number of occurrences of the hashtag as well as the freshness of the hashtag. On clicking an hashtag, the user will be able to see the relevant tweets. To fetch the relevant tweets we used dismax parser with query boosting placing more weight on the hashtag field.

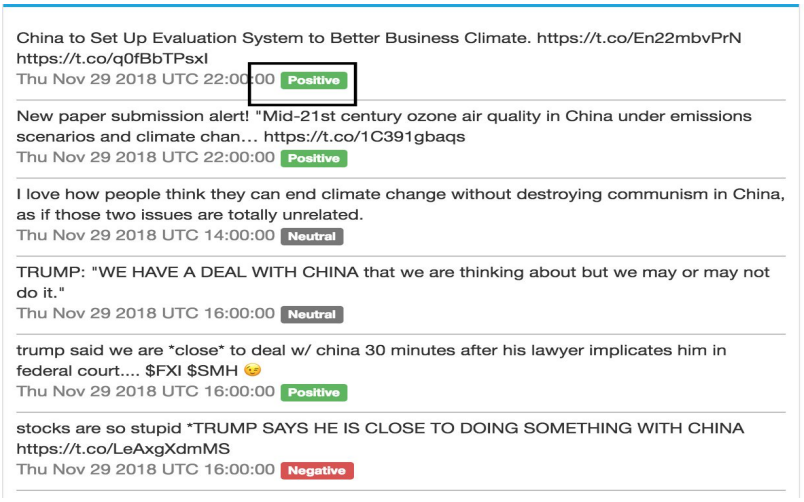
4) Analytics:

The user could analyse the data at individual query level as well as at the dataset level. There a four types of visualisation, By City, By Location, By Language, By Sentiment. We used Solr's faceted search to pull the data.



5) Sentiment Analysis:

We have preprocessed the sentiment of each tweet using a node.js module(sentiment) which uses AFINN word list as well as Emoji Sentiment Ranking to classify the tweet sentiment. Users can filter tweets by sentiment, view sentiment of each tweet, also visualise the data based on sentiments.



6) Translation:

We have added the ability to translate non english tweets using the google translate API

Hosting:

Application: http://ec2-52-14-62-68.us-east-2.compute.amazonaws.com:8080
Solr: http://ec2-52-14-62-68.us-east-2.compute.amazonaws.com:8983/solr/