

## **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:

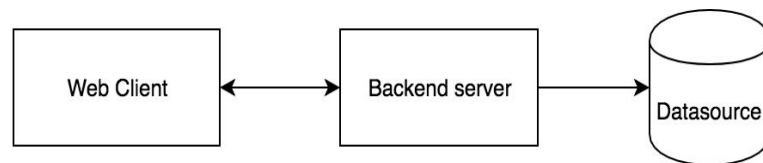
This project is an attempt to build an end to end IR system that lets the user to search and derive insights from social data collected from Twitter.

## 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

rahul ghandi

Tweets

@ArvindKejriwal Sir you should contest the election for PM You are more popular than Rahul Ghandi and an illiterate... <https://t.co/WScEaMWJTo>

Thu Nov 29 2018 UTC 15:00:00

Positive

@republic After Rahul Gandhi he is helping BJP

Wed Nov 28 2018 UTC 17:00:00

Positive

@RoHitMannn @FeignShourie Bjp will b defeated despite rahul gandhi .

Wed Nov 28 2018 UTC 18:00:00

Negative

TRS and MIM both want BJP to win - Rahul Gandhi

Wed Nov 28 2018 UTC 11:00:00

Positive

@anuragteddy @ANI But overall PM modi better thn Rahul gandhi...

Wed Nov 28 2018 UTC 11:00:00

Positive

@sherryontopp Now Rahul Gandhi ji was Pappu at one time.😂😂😂

Fri Nov 30 2018 UTC 21:00:00

Neutral

#Trending HashTags

#Rajasthan

2 Tweets

#TelanganaElections2018

2 Tweets

#RahulGandhi

2 Tweets

#Biubiu

2 Tweets

#TimesNow

1 Tweets

#Hanuman

1 Tweets

#LS2019

1 Tweets

#GOVT

1 Tweets

#AlookIfactory

1 Tweets

#Telangana

1 Tweets

## 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.

Tweetlytics

Home

Analytics

modi

Cities

☒ Delhi

☐ Bangkok

☐ Paris

☐ New York

☐ Mexico city

Topics

☐ Environment

☒ Politics

☐ Crime

☐ Social Unrest

☐ Infrastructure

Language

☐ English

☒ Hindi

Tweets

@INCIndia MP ki umeed chod do yrr... 75% me 70% modi sarkar. @BJP4MP

#MadhyaPradeshElection... <https://t.co/92YQGmfpMu> Translate

Thu Nov 29 2018 UTC 03:00:00

Neutral

NDTV --Modi ji aur BJP Ke against aur --Congress ke liye hawa banne ke liye poori Takkat laga lo --NDTV --Ravish K... <https://t.co/2G02KVX08N> Translate

Thu Nov 29 2018 UTC 03:00:00

Neutral

@aajtak @ashu3page Yogi ji ne bhi modi and Amit shah ki University mai padhai ki hai. Translate

Thu Nov 29 2018 UTC 03:00:00

Neutral

@Syedhas11537119 @RavishFanClub Jakaat kisi dusre dharm k logo ko dekar warpai kar do??..himmat hai to..MODI TO YUH... <https://t.co/kJvUfKrYDR> Translate

Thu Nov 29 2018 UTC 04:00:00

Neutral

#Janghoshnapatra\_Rajasthan kya modi ji ko apni maa ko saath Rakhna chahiye Translate

Thu Nov 29 2018 UTC 04:00:00

Neutral

#AapalAn9617017 @RavishFanClub Jakaat kisi dusre dharm k logo ko dekar warpai kar

#Trending HashTags

#Modi

30 Tweets

#ModiFor2019

8 Tweets

#ModiNationPride

5 Tweets

#modi

3 Tweets

#KisanMuktiMarch

3 Tweets

#राजस्थान\_बोले\_हर\_मोदी

3 Tweets

#NaMo2019

2 Tweets

#NewIndia

2 Tweets

#indiatrustsmodi

2 Tweets

#Trump

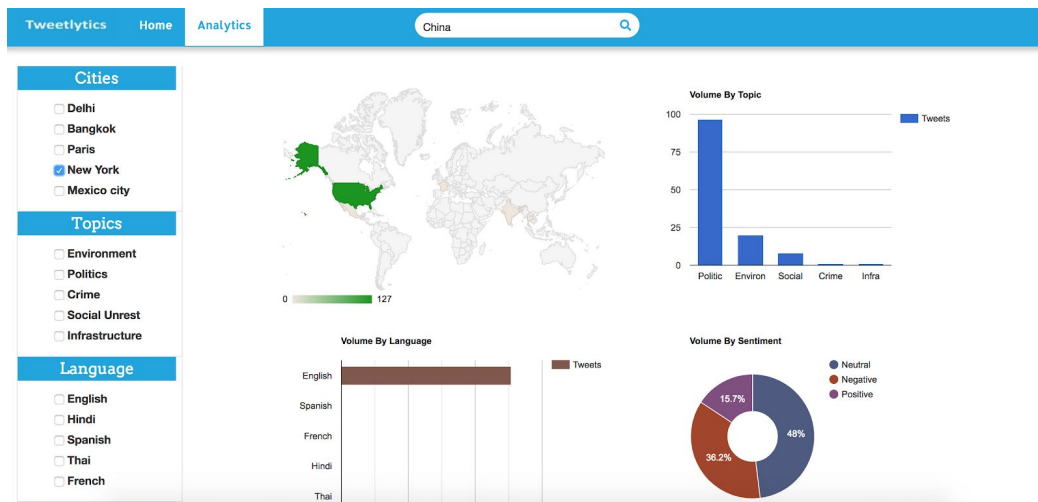
2 Tweets

## 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 are 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.

China to Set Up Evaluation System to Better Business Climate. <a href="https://t.co/En22mbvPrN">https://t.co/En22mbvPrN</a> <a href="https://t.co/q0fBbTPsxl">https://t.co/q0fBbTPsxl</a> Thu Nov 29 2018 UTC 22:00:00	Positive
New paper submission alert! "Mid-21st century ozone air quality in China under emissions scenarios and climate chan... <a href="https://t.co/1C391gbaqs">https://t.co/1C391gbaqs</a> Thu Nov 29 2018 UTC 22:00:00	Positive
I love how people think they can end climate change without destroying communism in China, as if those two issues are totally unrelated. Thu Nov 29 2018 UTC 14:00:00	Neutral
TRUMP: "WE HAVE A DEAL WITH CHINA that we are thinking about but we may or may not do it." Thu Nov 29 2018 UTC 16:00:00	Neutral
trump said we are "close" to deal w/ china 30 minutes after his lawyer implicates him in federal court.... \$FXI \$SMH 🤔 Thu Nov 29 2018 UTC 16:00:00	Positive
stocks are so stupid *TRUMP SAYS HE IS CLOSE TO DOING SOMETHING WITH CHINA <a href="https://t.co/LeAygXdmMS">https://t.co/LeAygXdmMS</a> Thu Nov 29 2018 UTC 16:00:00	Negative

## 6) Translation:

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

## Tweets

साझा चिंताओं पर एकजुट हो सकते हैं मोदी-शी <https://t.co/3wZbaujE7b> @Subhayan\_ism  
#NarendraModi #China #India #G20... <https://t.co/K26W3dG3hl> **Translate**  
Modi-Shi can be united on shared concerns <https://t.co/3wZbaujE7b> @Subhayan\_ism  
#NarendraModi #China #India #G20 ... <https://t.co/K26W3dG3hl>

Thu Nov 29 2018 UTC 12:00:00 **Neutral**

### 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/>