PROJECT 4
COMPLETE SEARCH AND ANALYTICS SOLUTION BASED ON DISSECTING TWITTER DATA

Synopsis:

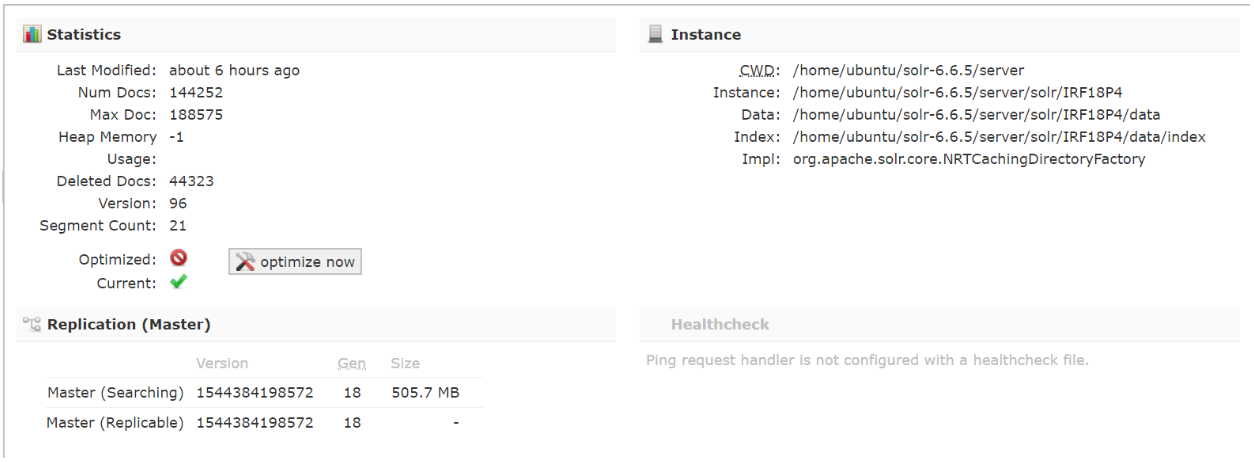
The main goal of this project is to build an end to end IR system which enables the user to get the knowledge about the particular topic, relevant tweets and analysis results. In the first step, we collected the twitter data on few particular topics and the distribution of tweets was among the selected five cities and five languages. In the next step, we implemented the BM25 on Solr-6.6.5 schema and indexed the tweets. In the third step, we created a Web stack which would display the query results along with the analysis for the selected query and faceted search to build the IR system on these tweet topics.

Twitter data:

Collected data of over 2 lakh tweets using twitter API. Collection of tweets is based on five tweet topics namely social unrest, politics, environment, crime and infrastructure and these tweet topics are collected in five languages namely English, Hindi, French, Spanish and Thai and in five cities namely New Delhi, New York, Paris, Mexico, Bangkok.

IR Model:

We created a BM25 model schema on Solr and indexed around 1.5 lakh tweet



Statistics

Last Modified: about 6 hours ago
Num Docs: 144252
Max Doc: 188575
Heap Memory Usage: -1
Deleted Docs: 44323
Version: 96
Segment Count: 21
Optimized: ⊘
Current: ✓

Instance

CWD: /home/ubuntu/solr-6.6.5/server
Instance: /home/ubuntu/solr-6.6.5/server/solr/IRF18P4
Data: /home/ubuntu/solr-6.6.5/server/solr/IRF18P4/data
Index: /home/ubuntu/solr-6.6.5/server/solr/IRF18P4/data/index
Impl: org.apache.solr.core.NRTCachingDirectoryFactory

Replication (Master)

	Version	Gen	Size
Master (Searching)	1544384198572	18	505.7 MB
Master (Replicable)	1544384198572	18	-

Healthcheck

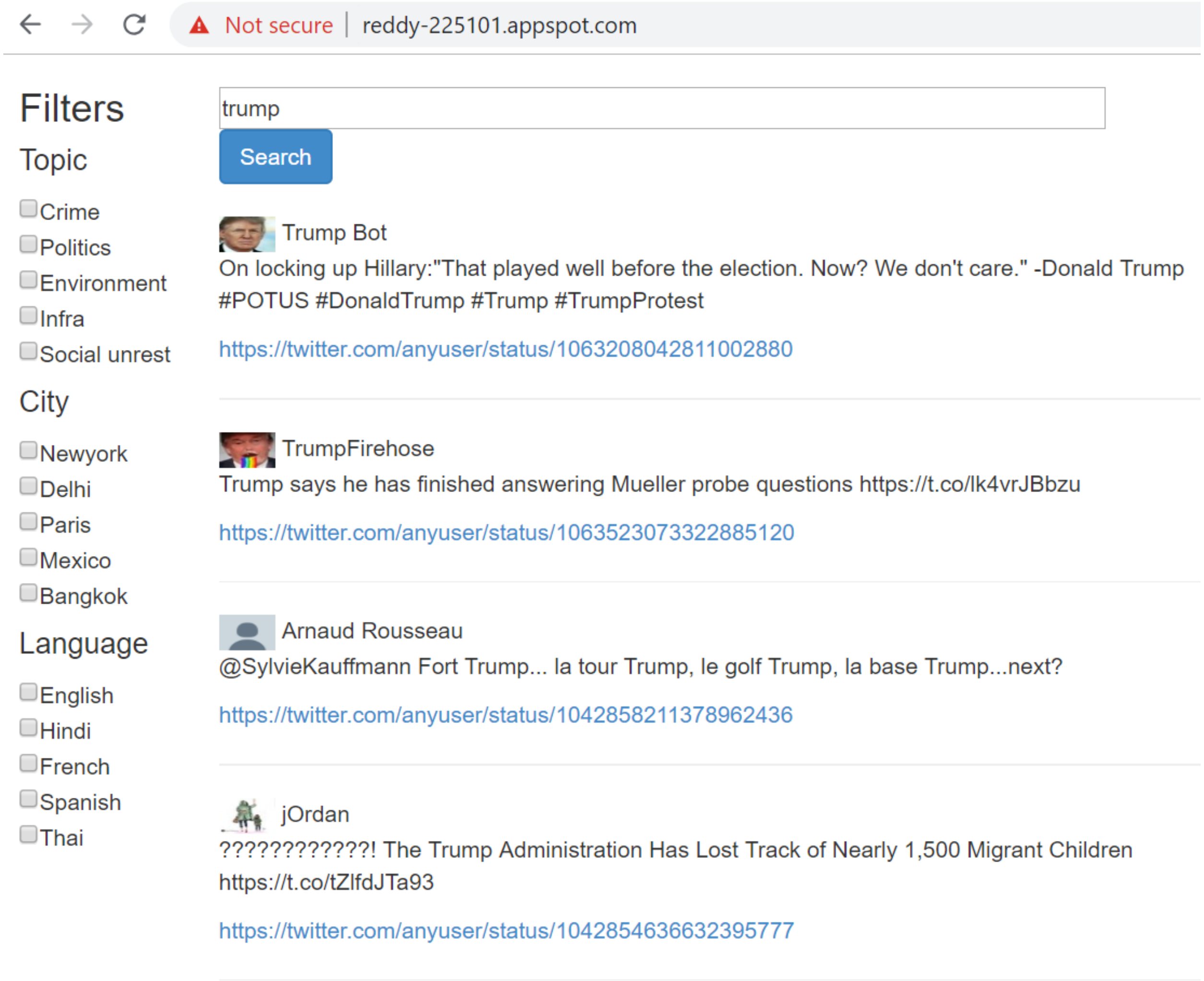
Ping request handler is not configured with a healthcheck file.

Technology Stack:

We used the java for the backend, the front end technology is JavaScript and javaservletpages to render the UI and server is Tomcat and Google cloud to host the IR system.

UI->SERVER->SOLR->SERVER(PROCESSING)->UI(DISPLAY)

UI Display:



Trump Bot

On locking up Hillary:"That played well before the election. Now? We don't care." -Donald Trump #POTUS #DonaldTrump #Trump #TrumpProtest

https://twitter.com/anyuser/status/1063208042811002880

TrumpFirehose

Trump says he has finished answering Mueller probe questions https://t.co/lk4vrJBbzu

https://twitter.com/anyuser/status/1063523073322885120

Arnaud Rousseau

@SylvieKauffmann Fort Trump... la tour Trump, le golf Trump, la base Trump...next?

https://twitter.com/anyuser/status/1042858211378962436

jOrdan

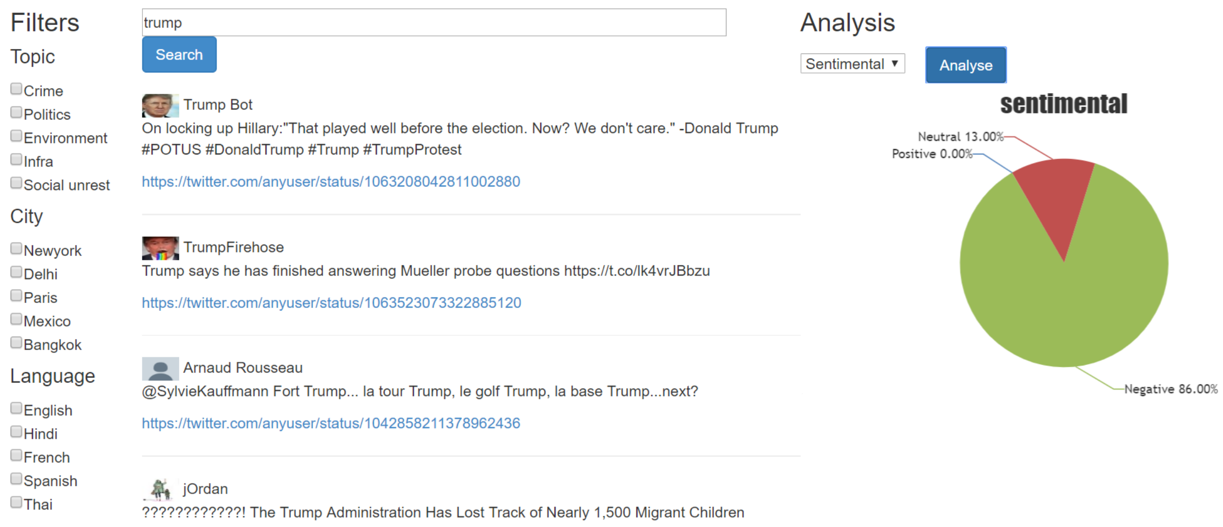
????????????! The Trump Administration Has Lost Track of Nearly 1,500 Migrant Children https://t.co/tZlfdJTa93

https://twitter.com/anyuser/status/1042854636632395777

The above image shows the retrieved tweets display. The middle part displays the tweets and the user who tweeted it and the text, the url directs to the twitter page where the tweet is present.

Sentimental Analysis:

Sentiment Analysis determines whether the writing is positive, negative or neutral. We have used Stanford.nlp package. Here in the below we have searched for trump and it shows the following sentimental analysis chart.



Filters

Topic

□Crime
□Politics
□Environment
□Infra
□Social unrest

City

□Newyork
□Delhi
□Paris
□Mexico
□Bangkok

Language

□English
□Hindi
□French
□Spanish
□Thai

trump

Search

Trump Bot
On locking up Hillary:"That played well before the election. Now? We don't care." -Donald Trump #POTUS #DonaldTrump #Trump #TrumpProtest
https://twitter.com/anyuser/status/1063208042811002880

TrumpFirehose
Trump says he has finished answering Mueller probe questions https://t.co/lk4vrJBbzu
https://twitter.com/anyuser/status/1063523073322885120

Arnaud Rousseau
@SylvieKauffmann Fort Trump... la tour Trump, le golf Trump, la base Trump...next?
https://twitter.com/anyuser/status/1042858211378962436

jOrdan
????????????! The Trump Administration Has Lost Track of Nearly 1,500 Migrant Children

Analysis

Sentimental ▾

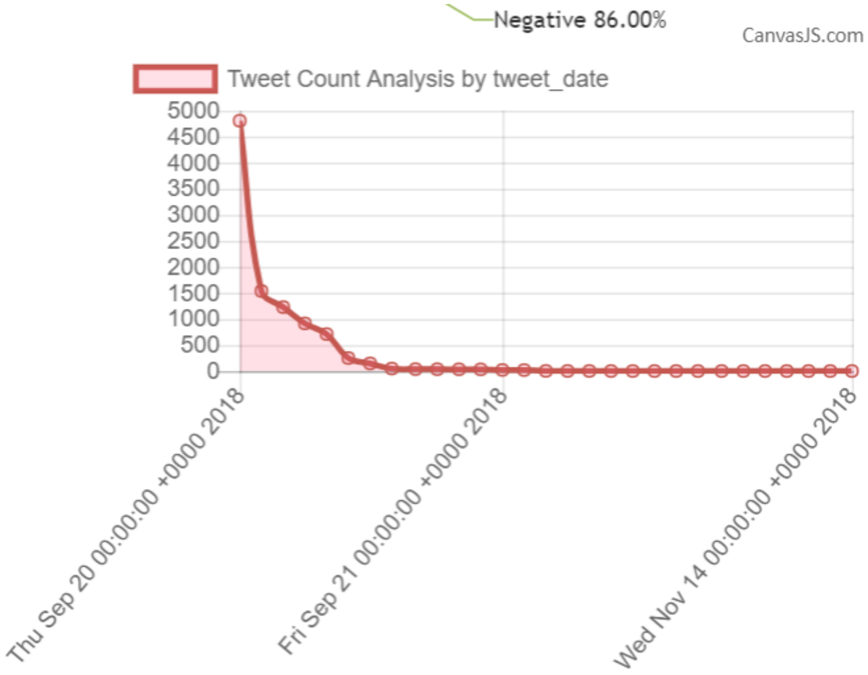
Analyse

Facet Analysis:

Below is the graph showing the distribution of tweets on various dates.

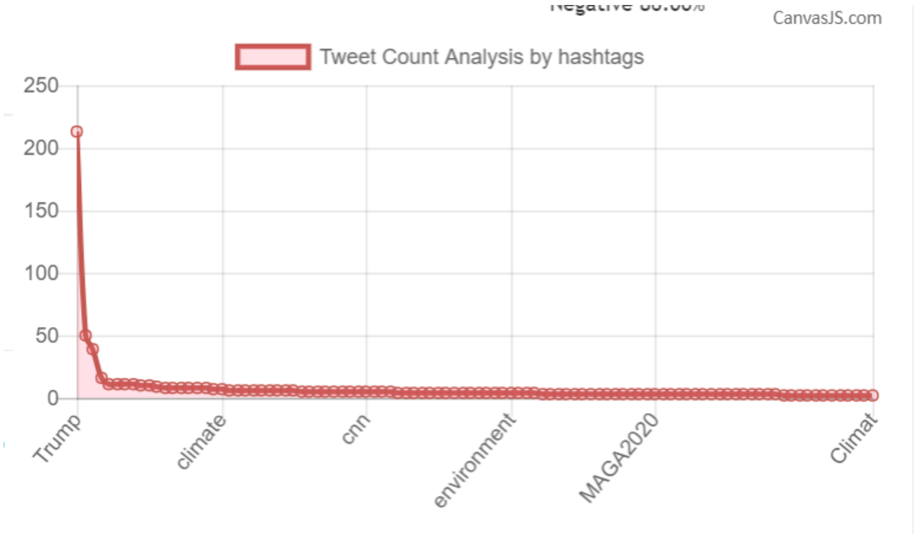
Using the below, following url we have done the facet search which is inbuilt in solr

http://18.221.162.107:8984/solr/IRF18P4/select?facet.field=tweet_date&facet=on&fl=tweet_date&indent=on&wt=json&q=trump&rows=15

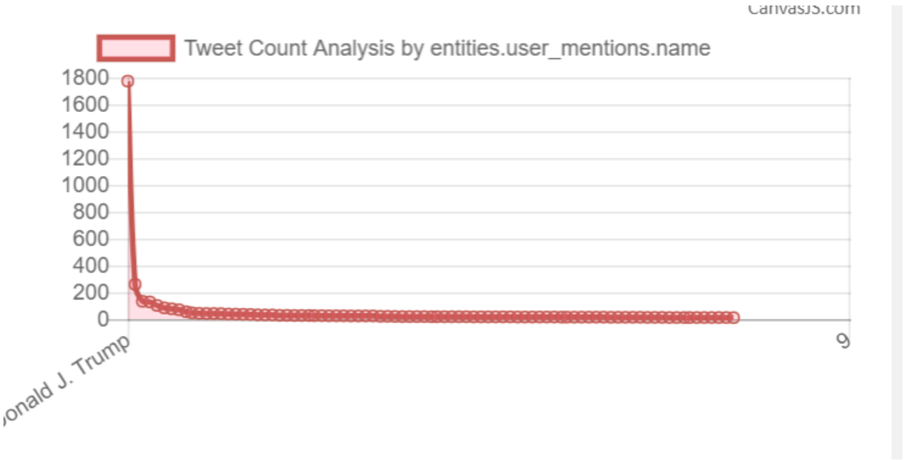


Below is the hashtag count graph. This is being accessed using the following url:
http://18.221.162.107:8984/solr/IRF18P4/select?facet.field=hashtags &facet=on&fl=hashtags&indent=on&wt=json&q=trump&rows=15

This is query dependent and we can use filters to get hashtags across the cities, languages and topics.

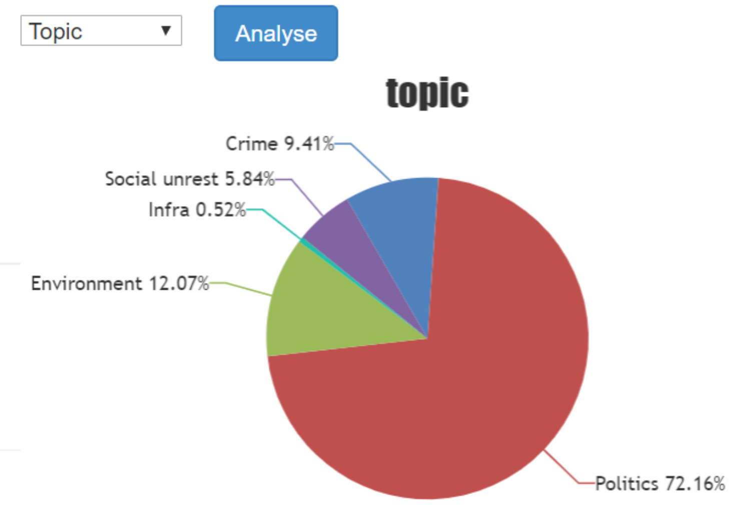


This is for the mentions, url being used is:
http://18.221.162.107:8984/solr/IRF18P4/select?facet.field=entities.user_mentions.name&facet=on&fl= entities.user_mentions.name&indent=on&wt=json&q=trump&rows=15

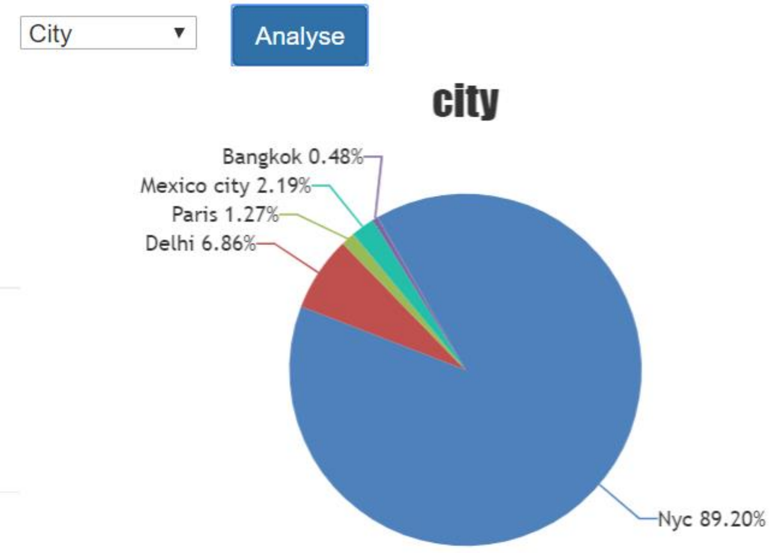


The below is for the topic and the url is

http://18.221.162.107:8984/solr/IRF18P4/select?facet.field=topic&facet=on&fl=topic&indent=on&wt=json&q=trump&rows=15

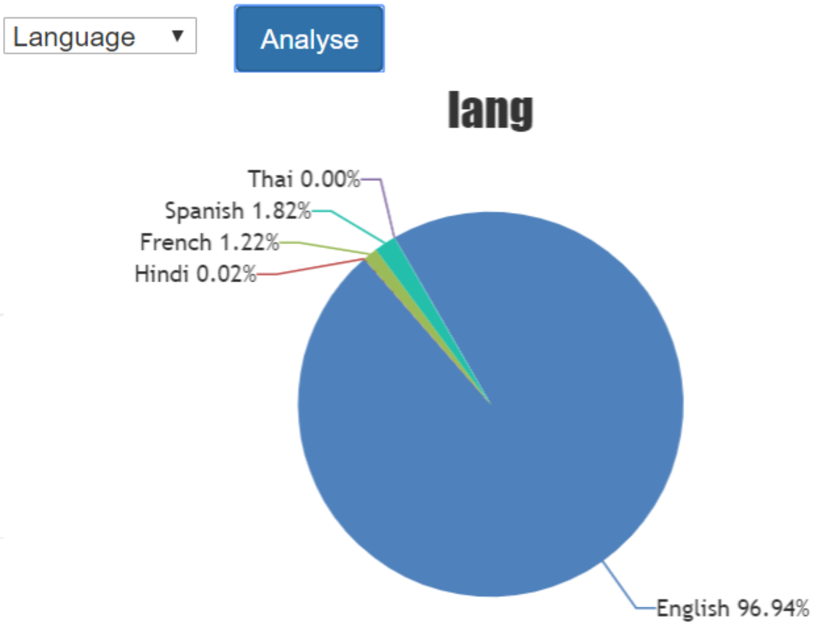


The below is for the city, the url used is http://18.221.162.107:8984/solr/IRF18P4/select?facet.field=city&facet=on&fl=city&indent=on&wt=json&q=trump&rows=15



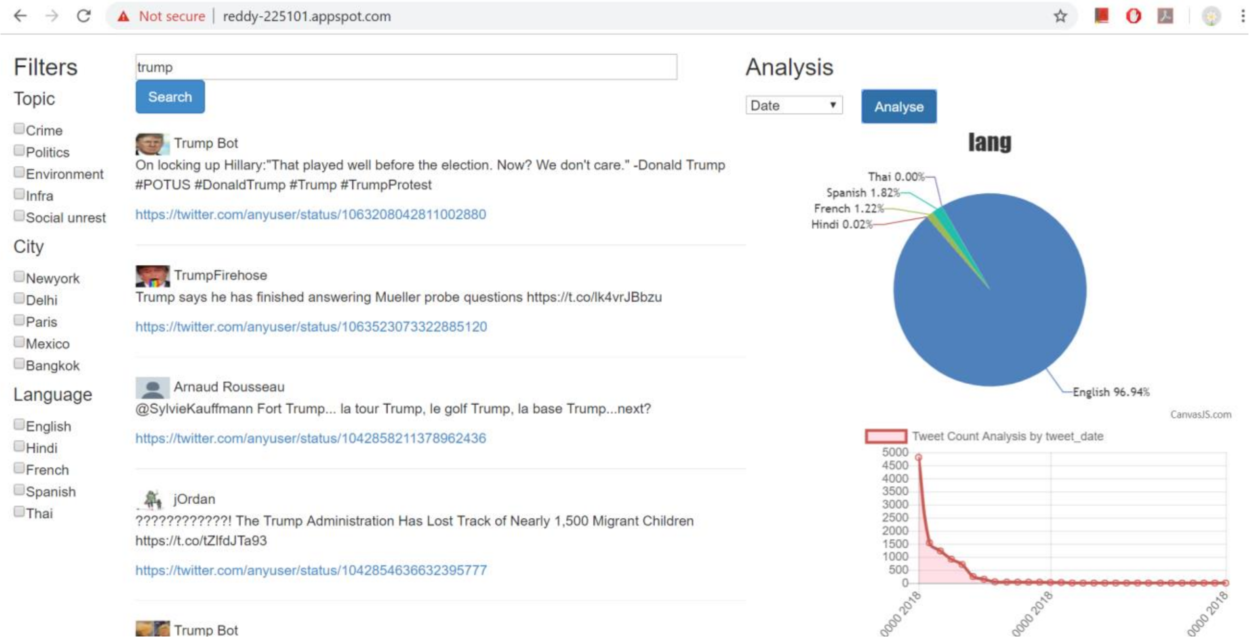
The below is for the language, the url used is:

http://18.221.162.107:8984/solr/IRF18P4/select?facet.field=lang&facet=on&fl=lang&indent=on&wt=json&q=trump&rows=15



Filters:

The below image shows the Complete UI display.



Trump Bot
On locking up Hillary:"That played well before the election. Now? We don't care." -Donald Trump #POTUS #DonaldTrump #Trump #TrumpProtest
https://twitter.com/anyuser/status/1063208042811002880

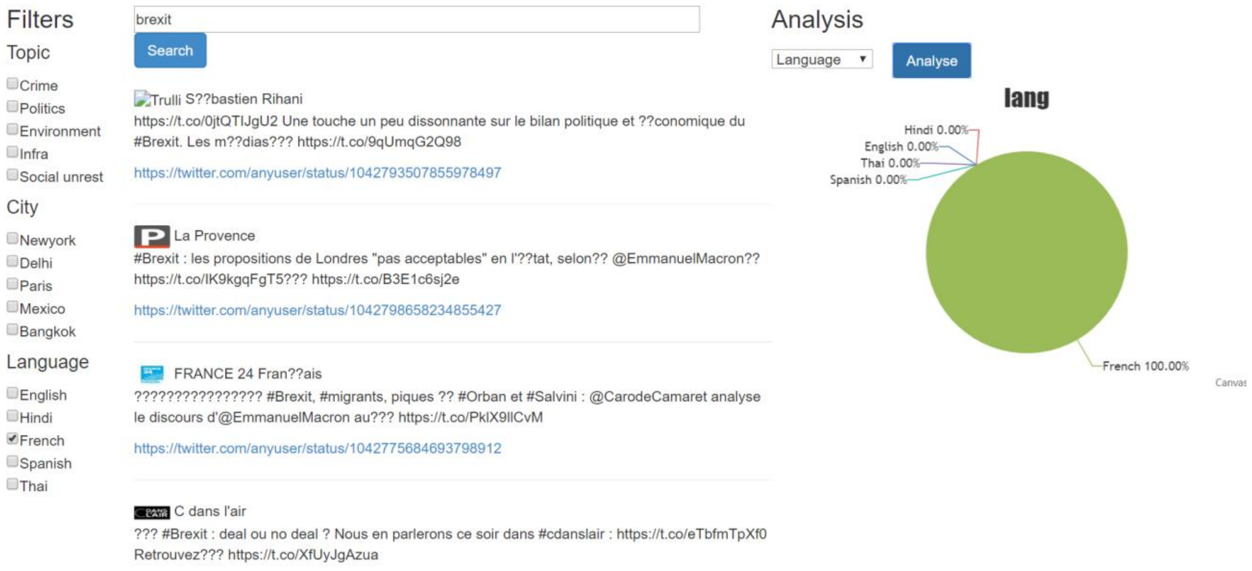
TrumpFirehose
Trump says he has finished answering Mueller probe questions https://t.co/lk4vrJBbzu
https://twitter.com/anyuser/status/1063523073322885120

Arnaud Rousseau
@SylvieKauffmann Fort Trump... la tour Trump, le golf Trump, la base Trump...next?
https://twitter.com/anyuser/status/1042858211378962436

jOrdan
????????????! The Trump Administration Has Lost Track of Nearly 1,500 Migrant Children https://t.co/tZlfdJTa93
https://twitter.com/anyuser/status/1042854636632395777

Trump Bot

After adding the language filter as French it displays the French language tweets as 100% in the chart:



Filters

Topic

☐Crime
☐Politics
☐Environment
☐Infra
☐Social unrest

City

☐Newyork
☐Delhi
☐Paris
☐Mexico
☐Bangkok

Language

☐English
☐Hindi
☑French
☐Spanish
☐Thai

brexit

Search

Trulli S??bastien Rihani
https://t.co/0jtQTIJgU2 Une touche un peu dissonnante sur le bilan politique et ??conomique du #Brexit. Les m??dias??? https://t.co/9qUmqG2Q98
https://twitter.com/anyuser/status/1042793507855978497

La Provence
#Brexit : les propositions de Londres "pas acceptables" en l'??tat, selon?? @EmmanuelMacron?? https://t.co/lK9kgqFgT5??? https://t.co/B3E1c6sj2e
https://twitter.com/anyuser/status/1042798658234855427

FRANCE 24 Fran??ais
???????????????? #Brexit, #migrants, piques ?? #Orban et #Salvini : @CarodeCamaret analyse le discours d'@EmmanuelMacron au??? https://t.co/PkIX9lICvM
https://twitter.com/anyuser/status/1042775684693798912

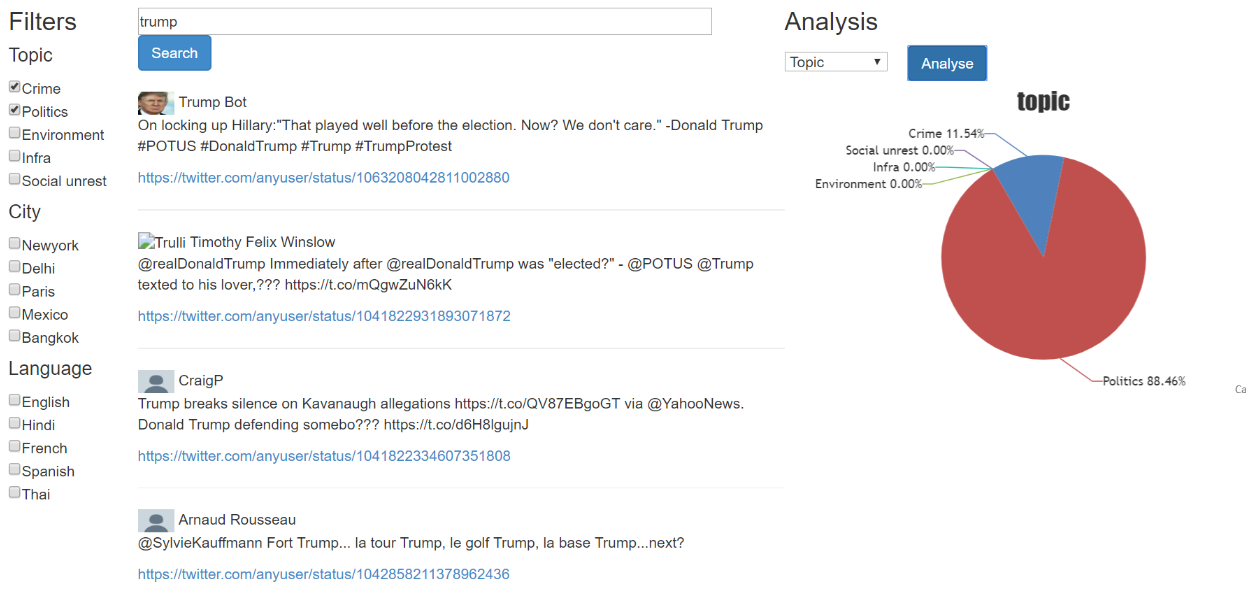
C dans l'air
??? #Brexit : deal ou no deal ? Nous en parlerons ce soir dans #cdanslair : https://t.co/eTbfmTpXf0 Retrouvez??? https://t.co/XfUyJgAzua

Analysis

Language ▾

Analyse

Here after adding the filter in Topic as Crime and Politics, it retrieves all the tweets and to the right us shows the number of tweets related to each field.



Filters

Topic

☑Crime
☑Politics
☐Environment
☐Infra
☐Social unrest

City

☐Newyork
☐Delhi
☐Paris
☐Mexico
☐Bangkok

Language

☐English
☐Hindi
☐French
☐Spanish
☐Thai

trump

Search

Trump Bot
On locking up Hillary:"That played well before the election. Now? We don't care." -Donald Trump #POTUS #DonaldTrump #Trump #TrumpProtest
https://twitter.com/anyuser/status/1063208042811002880

Trulli Timothy Felix Winslow
@realDonaldTrump Immediately after @realDonaldTrump was "elected?" - @POTUS @Trump texted to his lover,??? https://t.co/mQgwZuN6kK
https://twitter.com/anyuser/status/1041822931893071872

CraigP
Trump breaks silence on Kavanaugh allegations https://t.co/QV87EBgoGT via @YahooNews. Donald Trump defending somebo??? https://t.co/d6H8lgujnJ
https://twitter.com/anyuser/status/1041822334607351808

Arnaud Rousseau
@SylvieKauffmann Fort Trump... la tour Trump, le golf Trump, la base Trump...next?
https://twitter.com/anyuser/status/1042858211378962436

Analysis

Topic ▾

Analyse

Video Link:

Youtube video link: https://youtu.be/aq1fVDbMGYk

Google cloud link to access the project: https://reddy-225101.appspot.com/

Contributions:

Tweets collection, Indexing, UI: Mkalamda, sanand3.

Server side hosting and backend web stack implementation: aalluri, lveerise, meharaje