Analyzing Relationship between Nations based on Global Microblogging behavior

Kulraj Singh Saini
Computer Science
State University of New York
Binghamton
ksaini1@binghamton.edu

Shiva Prasad Subramani
Computer Science
State University of New York
Binghamton
ss1@binghamton.edu

Keerthan Vudutha
Computer Science
State University of New York
Binghamton
bvuduth@binghamton.edu

1.ABSTRACT

In a time where the world is connected through internet, information sharing has become a major part of our life. News is traveling faster than ever. People are interested in perceiving this information sharing, not only because of the information but also because of the way, it is being presented. One of the major reason of why there is a surge in digital information is the emergence of Social Media platforms such as Facebook, Reddit, Twitter, etc.

The news sharing has been one of the reason why countries are connected to each other more than ever. The people of a community try to adapt and integrate the cultural of other communities. This connection is not only through culture but there is a socio – economic, food, mineral exchange happening between different nations

This project aims at developing a model that will collect data with Twitter as platform and determine the relationship between different countries regarding how the tweet originated from one country and how it affected the tweets in some other countries. This will be helpful in tackling issues related to these relations and create a world where all nations coexist peacefully.

2.KEYWORDS

Social Media, twitter, countries, tweets

3.DATA SOURCE

This project will have Twitter as the platform from which data will be collected, The Program will aim at collecting information related to tweets such as hashtags in the tweets, the location of the tweet, the location of the retweets, etc. this information will help us map the different relationships between the countries.
Since twitter allows 15 requests per rate limit window, which are divided into 15-minute intervals, and we expect to collect at least 100 tweets with each request, so total number of tweets to be recorded would be done accordingly.
The figure 1 shows the dashboard of twitter API that will help us to keep a track on the total number of tweets fetched and total number of requests made in the particular month by our proposed model.
Using this dashboard, the usage peaks can also be determined. This will help us to understand the role of media in delivering news.

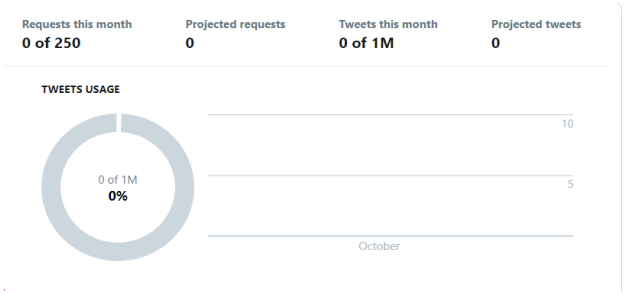

Figure 1 The twitter API dashboard to keep a record of total requests made and tweets fetched.

4.METHODOLOGY

The programming of the core data collection system will be in Java. The twitter gives developer option where user can register as developer and twitter provides the API. It has various classes and Interfaces such as:

1. Hashtag Entity: - get collections: Retrieves the identified Collection, presented as a list of tweets.(limit -1000 req /15 mins)
2. PlacesGeoResources: - GET trends/place: Returns top 50 trends of the location provided.(limit-75 req/ 15min)
3. Location: - GET geo/id/:place_id: Each Country has a specific country code, which can be obtained using this class.(limit- 75 req/ 15 min)

These can be used in our system to collect the specifics.
The other advantage of using twitter is that it provides developers with an inbuilt dashboard that can be used to track the data traffic of our program as we can see in Figure 1.
For Storage of this data we will be using Oracle and integrate it with the program. The data collected will be analyzed in a data-mining tool and the correlation coefficient will be determined.
Since the project is not specific to any particular fields, it will deal with multiple genre of trending topics and try to determine the relation of the two countries based on the tweet and its content. For complexity issues in dealing with data, the application will be developed to focus on mainly six different fields: Sports, Commerce, Politics, Environment, Entertainment and Technology.

5.ACKNOWLEDGMENT

We would like to thank our professor Jeremey H Blackburn for letting us work on this project and guiding us through and helping us with any problems faced.