

205229118

Mahalakshmi S

SMA Lab1. Real time crawling of tweets from Twitter

1. Install and import twitter library

In [10]:

```
import twitter
import json
```

2. Signup your twitter developer account and obtain the following credentials

consumer_key = “...”

consumer_secret = “ ... “

access_token = “ ... “

access_secret = “ ... “

In [11]:

```
CONSUMER_KEY = 'jQkSc5TR31Vdkjn51Lgrm6Pgl'
CONSUMER_SECRET = 'eAaRfKD64xSKAFTXeZy0r7RXKWif4PjHqL9D6Wtc2nrreLgI3J'
OAUTH_TOKEN = '1426417914756825088-HVzGYrkMN5g7QBpoGdhqiHp1PB5jFm'
OAUTH_TOKEN_SECRET = '1SQ1RCKwjsjJgNT3at4gbuscUd5glw4GMKun5McsXlmWo'
```

3. Write code for crawling real time data from Twitter Streaming API.

In [12]:

```
auth = twitter.oauth.OAuth(OAUTH_TOKEN, OAUTH_TOKEN_SECRET,
                           CONSUMER_KEY, CONSUMER_SECRET)

twitter_api = twitter.Twitter(auth=auth)
```

4. Retrieve data that is trending around the world by specifying any geographic region of your choice

In [13]:

```

WORLD_WOE_ID = 1 #Where on Earth Identifier
INDIA_WOE_ID = 23424848
LUCKNOW_WOE_ID = 2295377
# Prefix ID with the underscore for query string parameterization.
# Without the underscore, the twitter package appends the ID value
# to the URL itself as a special case keyword argument.

India_trends = twitter_api.trends.place(_id=INDIA_WOE_ID)
Lucknow_trends = twitter_api.trends.place(_id=LUCKNOW_WOE_ID)

print(json.dumps(India_trends, indent=1))
print()
print(json.dumps(Lucknow_trends, indent=1))

```

```

[
  {
    "trends": [
      {
        "name": "#MUNLEE",
        "url": "http://twitter.com/search?q=%23MUNLEE",
        "promoted_content": null,
        "query": "%23MUNLEE",
        "tweet_volume": 175863
      },
      {
        "name": "Pogba",
        "url": "http://twitter.com/search?q=Pogba",
        "promoted_content": null,
        "query": "Pogba",
        "tweet_volume": 305499
      },
      {
        "name": "SHAME ON YOU TEAM SAATH",
        "url": "http://twitter.com/search?q=%23SHAME-ON-YOU-TEAM-SAATH%23",
        "promoted_content": null,
        "query": "%23SHAME-ON-YOU-TEAM-SAATH%23",
        "tweet_volume": 175863
      }
    ]
  }
]

```

5. Print the trend names.

In [14]:

```
for trend in India_trends[0]['trends']:  
    print(trend['name'])
```

```
#MUNLEE  
Pogba  
SHAME ON YOU TEAM SAATH  
#7YearsOfAnjaanSupremacy  
#MUFC  
#GGMU  
#FullPaisaVasoolSale  
Joe Root  
Sancho  
Leeds  
Bruno Fernandes  
Old Trafford  
Greenwood  
Varane  
Bruno hattrick  
Ayling  
Partition  
पूर्व संध्या  
14 Months Of Injustice 2 SSR  
Hat-trick  
Lindelof  
Bairstow  
Mason  
NOT FUNNY  
Dan James  
NAYA GHAR MUBARAK JASMIN  
Bielsa  
ONE MAN ARMY PRATIK  
Ashwin  
REMINISCING SHARAD AS SAGAR  
Pique  
McTominay  
Cube  
Jinnah  
Muslim League  
SBI Debit & Credit Cards  
McFred  
Nehru  
IT Rules  
Man United  
Steve Smith  
Ishant  
T&C Apply  
Martial  
#BheemlaNayakArrivesTMRW  
#ManUtd  
#HinaKhanOnMoj  
#IndependenceDayIndia2021  
#ManchesterUnited  
#TirangaYatra
```

6. Print common trends using set function.

In [15]:

```
India_trends_set = set([trend['name']  
                        for trend in India_trends[0]['trends']])  
  
Lucknow_trends_set = set([trend['name']  
                          for trend in Lucknow_trends[0]['trends']])  
  
common_trends = India_trends_set.intersection(Lucknow_trends_set)  
  
print(common_trends)
```

```
{'#MUNLEE', 'Joe Root', 'Bruno Fernandes', 'Greenwood', 'Lindelof', 'Mason',  
'Nehru', '#7YearsOfAnjaanSupremacy', 'SHAME ON YOU TEAM SAATH', 'REMINISCING  
SHARAD AS SAGAR', '#BheemlaNayakArrivesTMRW', 'Ayling', 'Dan James', 'NAYA G  
HAR MUBARAK JASMIN', 'Cube', 'SBI Debit & Credit Cards', 'Bairstow', 'McTomi  
nay', 'ONE MAN ARMY PRATIK', 'McFred', 'Old Trafford', 'Martial', 'IT Rule  
s', '#IndependenceDayIndia2021', '#FullPaisaVasoolSale', '14 Months Of Injus  
tice 2 SSR', '#MUFC', 'Sancho', 'Varane', 'पूर्व संध्या', 'Muslim League', 'Ashwi  
n', 'Hat-trick', 'Bruno hattrick', 'Ishant', 'NOT FUNNY', 'Man United', 'Bie  
lsa', 'Pogba', '#ManUtd', 'Pique', 'Leeds', 'T&C Apply', 'Jinnah', '#GGMU',  
'Steve Smith'}
```