

# TeamLovelace - 20/05/2016

## - Xinqi Li

### Today's work:

- Try to fetch some tweets first.
- Fetch out the user mentions of the tweets, so these user mentions may be relevant to the topic that the user is interested in.
- Try to recommend some account for the users to follow.

In [1]:

```
import tweepy
import json
import pandas as pd
from collections import Counter

consumer_key = 'Bt7Ih6Bk5Y5kyLEUfbyyEwDjm'
consumer_secret = 'x2vxVpEj0uK3ITXFgYLSncgvZEWi1nhnYbdB70GmsQMQ37Vgd6'
access_token = '711915268251983875-caR8pX1PJyW5g3aaZZXo9tNcSbXPev8'
access_secret = 'DJKjqNuBmv7fUHVKDyuMmETeZHZ55IGqAmTuG0y0UysjZ'

auth = tweepy.OAuthHandler(consumer_key=consumer_key, consumer_secret=consumer_secret)
auth.set_access_token(access_token, access_secret)

api = tweepy.API(auth)
print(api)
```

<tweepy.api.API object at 0x1118a5668>

## Data Collection

For example, a user is interested in three topics, Grammys, Manchester United and Donald Trump  
Let's collect 1000 tweets of each, so 3000 tweets in total.

In [2]:

```
grammys_results = [ status._json for status in tweepy.Cursor(api.search,
                                                             q="Grammys",
                                                             count=1000).items(1000) ]
len(grammys_results)
```

Out[2]:

1000

In [3]:

```
trump_results = [ status._json for status in tweepy.Cursor(api.search,
                                                            q="Donald Trump",
                                                            count=1000).items(1000)]
len(trump_results)
```

Out[3]:

1000

In [4]:

```
manchester_results = [ status._json for status in tweepy.Cursor(api.search,
                                                                q="Manchester United",
                                                                count=1000).items(1000)]
len(manchester_results)
```

Out[4]:

1000

So as the example shown below, we can get many useful information.

In [5]:

```
grammys_results[0]
```

Out[5]:

```
{'contributors': None,
 'coordinates': None,
 'created_at': 'Fri May 20 14:25:09 +0000 2016',
 'entities': {'hashtags': [],
 'media': [{'display_url': 'pic.twitter.com/bvBoCWYkIu',
 'expanded_url': 'http://twitter.com/ochman101/status/733438
132234903553/photo/1',
 'id': 733438131060527104,
 'id_str': '733438131060527104',
 'indices': [139, 140],
 'media_url': 'http://pbs.twimg.com/media/Ci2yGCKW0AACm30.jp
g',
 'media_url_https': 'https://pbs.twimg.com/media/Ci2yGCKW0AA
Cm30.jpg',
 'sizes': {'large': {'h': 360, 'resize': 'fit', 'w': 640},
 'medium': {'h': 338, 'resize': 'fit', 'w': 600},
 'small': {'h': 191, 'resize': 'fit', 'w': 340},
 'thumb': {'h': 150, 'resize': 'crop', 'w': 150}},
 'source_status_id': 733438132234903553,
 'source_status_id_str': '733438132234903553',
 'source_user_id': 66166543,
 'source_user_id_str': '66166543',
 'type': 'photo',
 'url': 'https://t.co/bvBoCWYkIu'}]},
 'symbols': [],
 'urls': [],
 'user_mentions': [{'id': 66166543,
 'id_str': '66166543',
 'indices': [3, 13],
 'name': 'Oche Oche-Obe',
 'screen_name': 'ochman101'}]},
 'favorite_count': 0,
 'favorited': False,
 'geo': None,
 'id': 733664866075054080,
 'id_str': '733664866075054080',
 'in_reply_to_screen_name': None,
 'in_reply_to_status_id': None,
 'in_reply_to_status_id_str': None,
 'in_reply_to_user_id': None,
 'in_reply_to_user_id_str': None,
 'is_quote_status': False,
 'lang': 'en',
 'metadata': {'iso_language_code': 'en', 'result_type': 'recen
t'},
 'place': None,
 'possibly_sensitive': False,
 'retweet_count': 468,
 'retweeted': False,
 'retweeted_status': {'contributors': None,
 'coordinates': None,
 'created_at': 'Thu May 19 23:24:12 +0000 2016',
 'entities': {'hashtags': [],
 'media': [{'display_url': 'pic.twitter.com/bvBoCWYkIu',
 'expanded_url': 'http://twitter.com/ochman101/status/73343
8132234903553/photo/1',
 'id': 733438131060527104,
```

```

'id_str': '733438131060527104',
'indices': [116, 139],
'media_url': 'http://pbs.twimg.com/media/Ci2yGCKW0AACm30.jpg',
'media_url_https': 'https://pbs.twimg.com/media/Ci2yGCKW0AACm30.jpg',
'sizes': {'large': {'h': 360, 'resize': 'fit', 'w': 640},
'medium': {'h': 338, 'resize': 'fit', 'w': 600},
'small': {'h': 191, 'resize': 'fit', 'w': 340},
'thumb': {'h': 150, 'resize': 'crop', 'w': 150}},
'type': 'photo',
'url': 'https://t.co/bvBoCWYkIu'}],
'symbols': [],
'urls': [],
'user_mentions': [],
'favorite_count': 201,
'favorited': False,
'geo': None,
'id': 733438132234903553,
'id_str': '733438132234903553',
'in_reply_to_screen_name': 'ochman101',
'in_reply_to_status_id': 733435995035729924,
'in_reply_to_status_id_str': '733435995035729924',
'in_reply_to_user_id': 66166543,
'in_reply_to_user_id_str': '66166543',
'is_quote_status': False,
'lang': 'en',
'metadata': {'iso_language_code': 'en', 'result_type': 'recent'},
'place': {'attributes': {},
'bounding_box': {'coordinates': [[[2.6654364, 4.1974055],
[14.6777951, 4.1974055],
[14.6777951, 13.8881151],
[2.6654364, 13.8881151]]],
'type': 'Polygon'},
'contained_within': [],
'country': 'Nigeria',
'country_code': 'NG',
'full_name': 'Nigeria',
'id': '59b27337f38d658b',
'name': 'Nigeria',
'place_type': 'country',
'url': 'https://api.twitter.com/1.1/geo/id/59b27337f38d658b.json'},
'possibly_sensitive': False,
'retweet_count': 468,
'retweeted': False,
'source': '<a href="http://twitter.com" rel="nofollow">Twitter Web Client</a>',
'text': 'Seal - Henry Olusegun Adeola Samuel. Born to a Yoruba mother. Sold over 30 million albums to date and won 4 Grammys https://t.co/bvBoCWYkIu',
'truncated': False,
'user': {'contributors_enabled': False,
'created_at': 'Sun Aug 16 18:42:57 +0000 2009',
'default_profile': False,
'default_profile_image': False,

```

```
'description': "Osuofia in West Virginia. @Frannie_J's side  
nigga #2. IG/Snapchat: ochman101",  
'entities': {'description': {'urls': []},  
  'url': {'urls': [{'display_url': 'ochposts.tumblr.com',  
    'expanded_url': 'http://ochposts.tumblr.com',  
    'indices': [0, 23],  
    'url': 'https://t.co/DpgcBPx77m'}]}}},  
'favourites_count': 1279,  
'follow_request_sent': False,  
'followers_count': 2338,  
'following': False,  
'friends_count': 1228,  
'geo_enabled': True,  
'has_extended_profile': True,  
'id': 66166543,  
'id_str': '66166543',  
'is_translation_enabled': False,  
'is_translator': False,  
'lang': 'en',  
'listed_count': 9,  
'location': 'Lagos, Nigeria',  
'name': 'Oche Oche-Obe',  
'notifications': False,  
'profile_background_color': '7D7A77',  
'profile_background_image_url': 'http://pbs.twimg.com/profile_  
background_images/493376804711714818/bMmObYR2.jpeg',  
'profile_background_image_url_https': 'https://pbs.twimg.co  
m/profile_background_images/493376804711714818/bMmObYR2.jpeg',  
'profile_background_tile': False,  
'profile_banner_url': 'https://pbs.twimg.com/profile_banner  
s/66166543/1458685659',  
'profile_image_url': 'http://pbs.twimg.com/profile_images/73  
2336110903435265/Ve2N2K2z_normal.jpg',  
'profile_image_url_https': 'https://pbs.twimg.com/profile_im  
ages/732336110903435265/Ve2N2K2z_normal.jpg',  
'profile_link_color': '9266CC',  
'profile_sidebar_border_color': '000000',  
'profile_sidebar_fill_color': 'A15252',  
'profile_text_color': '000000',  
'profile_use_background_image': True,  
'protected': False,  
'screen_name': 'ochman101',  
'statuses_count': 38193,  
'time_zone': 'Eastern Time (US & Canada)',  
'url': 'https://t.co/DpgcBPx77m',  
'utc_offset': -14400,  
'verified': False}},  
'source': '<a href="http://twitter.com/download/iphone" rel="n  
ofollow">Twitter for iPhone</a>',  
'text': 'RT @ochman101: Seal - Henry Olusegun Adeola Samuel. B  
orn to a Yoruba mother. Sold over 30 million albums to date and  
won 4 Grammys https://...',  
'truncated': False,  
'user': {'contributors_enabled': False,  
  'created_at': 'Mon Mar 10 04:17:23 +0000 2014',  
  'default_profile': True,  
  'default_profile_image': False,
```

```
'description': 'Vice President Operations of the York Federat
ion of Students @yfslocal68',
'entities': {'description': {'urls': []},
'url': {'urls': [{'display_url': 'yfs.ca',
'expanded_url': 'http://www.yfs.ca',
'indices': [0, 23],
'url': 'https://t.co/RcNuGvN93A'}]}},
'favourites_count': 13712,
'follow_request_sent': False,
'followers_count': 1221,
'following': False,
'friends_count': 1025,
'geo_enabled': False,
'has_extended_profile': True,
'id': 2381504131,
'id_str': '2381504131',
'is_translation_enabled': False,
'is_translator': False,
'lang': 'en',
'listed_count': 13,
'location': '416',
'name': '.',
'notifications': False,
'profile_background_color': 'C0DEED',
'profile_background_image_url': 'http://abs.twimg.com/images/
themes/them1/bg.png',
'profile_background_image_url_https': 'https://abs.twimg.com/
images/themes/them1/bg.png',
'profile_background_tile': False,
'profile_banner_url': 'https://pbs.twimg.com/profile_banners/
2381504131/1462407591',
'profile_image_url': 'http://pbs.twimg.com/profile_images/730
246598609920001/Ot-E1_Ga_normal.jpg',
'profile_image_url_https': 'https://pbs.twimg.com/profile_ima
ges/730246598609920001/Ot-E1_Ga_normal.jpg',
'profile_link_color': '0084B4',
'profile_sidebar_border_color': 'C0DEED',
'profile_sidebar_fill_color': 'DDEEF6',
'profile_text_color': '333333',
'profile_use_background_image': True,
'protected': False,
'screen_name': 'Mayyree_May',
'statuses_count': 24489,
'time_zone': 'Atlantic Time (Canada)',
'url': 'https://t.co/RcNuGvN93A',
'utc_offset': -10800,
'verified': False}}
```



In [6]:

```
#fetch all the user mentions in the tweets
grammys_user_mentions = [ user_mention['screen_name']
                           for status in grammys_results
                           for user_mention in status['entities']['user_mentions']]

#fetch all the user mentions in the tweets
trump_user_mentions = [ user_mention['screen_name']
                        for status in trump_results
                        for user_mention in status['entities']['user_mentions']]

#fetch all the user mentions in the tweets
manchester_user_mentions = [ user_mention['screen_name']
                              for status in manchester_results
                              for user_mention in status['entities']['user_mentions']]

#count the user mentions
grammys_count = Counter(grammys_user_mentions)
grammys_mentions_counts = grammys_count.most_common()

#count the user mentions
trump_count = Counter(trump_user_mentions)
trump_mentions_counts = trump_count.most_common()

#count the user mentions
manchester_count = Counter(manchester_user_mentions)
manchester_mentions_counts = manchester_count.most_common()
```

In [7]:

```
grammys_mentions_counts[:3]
```

Out[7]:

```
[('ochman101', 154), ('ArianaFactOnly', 117), ('chartflops', 106)]
```

In [8]:

```
trump_mentions_counts[:3]
```

Out[8]:

```
[('realDonaldTrump', 50), ('CNN', 39), ('ClassicPict', 26)]
```

In [9]:

```
manchester_mentions_counts[:3]
```

Out[9]:

```
[('FootballFunnys', 146), ('ESPNFC', 28), ('guardian', 18)]
```

In [10]:

```
#fetch all the user mentions in the tweets
grammys_users = [ user['user']['screen_name']
                  for user in grammys_results]

trump_users = [ user['user']['screen_name']
               for user in trump_results]

manchester_users = [ user['user']['screen_name']
                    for user in manchester_results]

grammys_user_count = Counter(grammys_users)
grammys_users_counts = grammys_user_count.most_common()

trump_user_count = Counter(trump_users)
trump_users_counts = trump_user_count.most_common()

manchester_user_count = Counter(manchester_users)
manchester_users_counts = manchester_user_count.most_common()
```

In [11]:

```
grammys_users_counts[:3]
```

Out[11]:

```
[('JoseCorea_', 16), ('__dUkEz', 14), ('fondlinhes', 7)]
```

In [12]:

```
trump_users_counts[:3]
```

Out[12]:

```
[('CBSNewsPoll', 8), ('ijungsukk', 7), ('RealTrevinShu', 6)]
```

In [13]:

```
manchester_users_counts[:3]
```

Out[13]:

```
[('manchester_app', 20), ('ManU_FCFans', 6), ('GAINSBURG59', 6)]
```

Now let's print the results:

In [14]:

```
print("Topic Grammys:\n\nYou may interested in the following accounts:\n")
for item in grammys_mentions_counts[:3]:
    print("@%s was mentioned %d times" % (item[0], item[1]))
print("\n")
for item in grammys_users_counts[:3][:3]:
    print("%d of @%s's tweets are about Grammys." % (item[1], item[0]))
print("\n")

print("Topic Donald Trump:\n\nYou may interested in the following accounts:\n")
for item in trump_mentions_counts[:3]:
    print("@%s was mentioned %d times" % (item[0], item[1]))
print("\n")
for item in trump_users_counts[:3]:
    print("%d of @%s's tweets are about Donald Trump." % (item[1], item[0]))
print("\n")

print("Topic Manchester United:\n\nYou may interested in the following accounts:\n")
for item in manchester_mentions_counts[:3]:
    print("@%s was mentioned %d times" % (item[0], item[1]))
print("\n")
for item in manchester_users_counts[:3]:
    print("%d of @%s's tweets are about Manchester United." % (item[1], item[0]))
print("\n")
```

#### Topic Grammys:

You may interested in the following accounts:

@ochman101 was mentioned 154 times  
@ArianaFactOnly was mentioned 117 times  
@chartflops was mentioned 106 times

16 of @JoseCorea\_'s tweets are about Grammys.  
14 of @\_\_dUkEz's tweets are about Grammys.  
7 of @fondlinhes's tweets are about Grammys.

#### Topic Donald Trump:

You may interested in the following accounts:

@realDonaldTrump was mentioned 50 times  
@CNN was mentioned 39 times  
@ClassicPict was mentioned 26 times

8 of @CBSNewsPoll's tweets are about Donald Trump.  
7 of @ijungsukk's tweets are about Donald Trump.  
6 of @RealTrevinShu's tweets are about Donald Trump.

#### Topic Manchester United:

You may interested in the following accounts:

@FootballFunnys was mentioned 146 times  
@ESPNFC was mentioned 28 times  
@guardian was mentioned 18 times

20 of @manchester\_app's tweets are about Manchester United.  
6 of @ManU\_FCFans's tweets are about Manchester United.  
6 of @GAINSBURG59's tweets are about Manchester United.

