

Working with twitter in Python

Aashish Pandey

2018-01-09

Contents

| | | |
|---|---|----|
| 1 | tweepy - A Python library for accessing the Twitter API | 5 |
| 2 | - Creating the app | 7 |
| 3 | - Getting Started | 13 |
| 4 | - Getting all tweets from a user | 15 |
| 5 | - Searching Tweets with a hashtag or Keywords | 19 |

Chapter 1

tweepy - A Python library for accessing the Twitter API

Twitter is one of the most popular social networks where users tweet their opinion about all sorts of things. All Firms/Brands now have twitter accounts and they use twitter platform to engage with their customers. Everyday users on twitter generate more than 500 million tweets. Often we want to download and analyze tweets data.

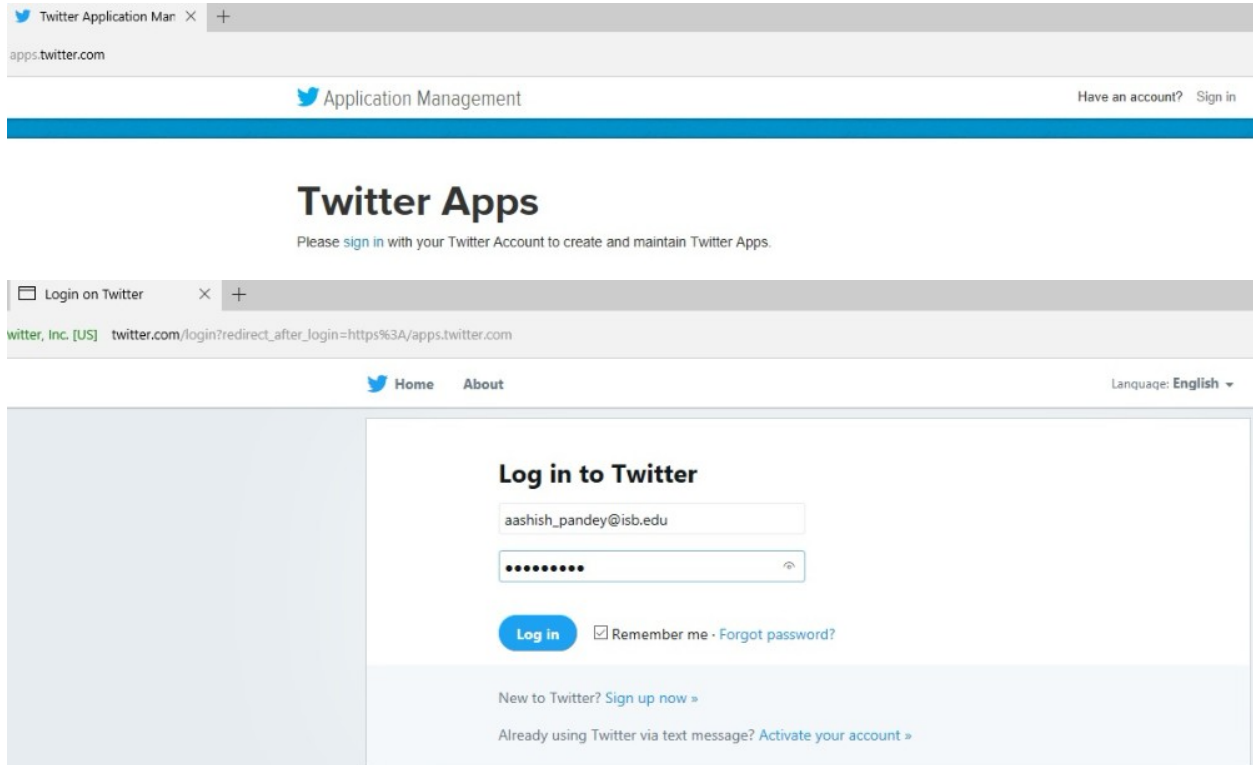
Twitter provides free API access to its tweets data-base. So a user can access and manage his or her twitter account with help of twitter API. In this article I will discuss a python library “tweepy” for accessing twitter data with twitter API. In order to access twitter data in python, a user must create an app on Twitter Developer Platform. Please see the instructions in next section for creating the app and generating the credentials for accessing twitter API.

Chapter 2

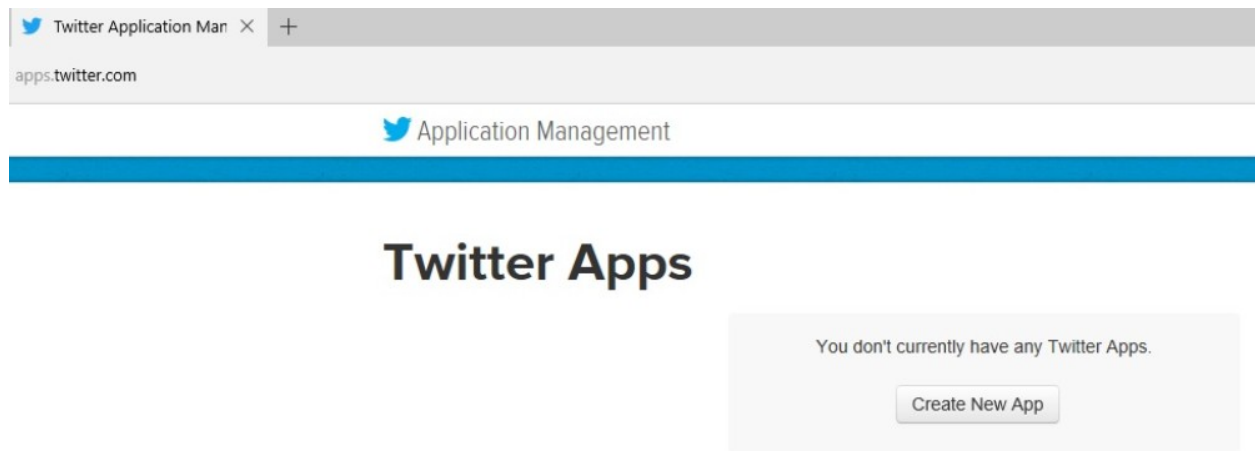
- Creating the app

Please note that you should have an active twitter account to access twitter data with twitter API. If you don't have a twitter account, then first register on twitter.

Step 1 - Open Twitter Application Management and log in with your Twitter Account.



Step 2 - Now click on Create new application.



Step 3 - In this page enter all the details. In website you can enter any valid url for page redirect. (for example <https://twitter.com/> in this case). Leave the Callback URL blank. Scroll down , accept the license agreement and click on create your twitter application.

Create an application

Application Details

Name *

Your application name. This is used to attribute the source of a tweet and in user-facing authorization screens. 32 characters max.

Description *

Your application description, which will be shown in user-facing authorization screens. Between 10 and 200 characters max.

Website *

Your application's publicly accessible home page, where users can go to download, make use of, or find out more information about your application. This fully-qualified URL is used in the source attribution for tweets created by your application and will be shown in user-facing authorization screens. (If you don't have a URL yet, just put a placeholder here but remember to change it later.)

Callback URL

Where should we return after successfully authenticating? [OAuth 1.0a](#) applications should explicitly specify their `oauth_callback` URL on the request token step, regardless of the value given here. To restrict your application from using callbacks, leave this field blank.

Developer Agreement

☒ Yes, I have read and agree to the [Twitter Developer Agreement](#).


Create your Twitter application

Your application has been created. Please take a moment to review and adjust your application's settings.

tweepy_demo_app

Test OAuth

Details Settings Keys and Access Tokens Permissions

 App for academic demonstration
<https://twitter.com/>



Organization

Information about the organization or company associated with your application. This information is optional.

| | |
|----------------------|------|
| Organization | None |
| Organization website | None |

< >

Step 4 - Now click on Keys and Access Tokens tab to see the Consumer Key and Consumer Secret. After this scroll down and click on Create my access token.

 Application Management
 

tweepy_demo_app

Test OAuth

[Details](#)
[Settings](#)
[Keys and Access Tokens](#)
[Permissions](#)

Application Settings

Keep the "Consumer Secret" a secret. This key should never be human-readable in your application.

| | |
|------------------------------|---|
| Consumer Key (API Key) | LT68-2M4K-Mg7L8-4H2B-2 |
| Consumer Secret (API Secret) | h |
| Access Level | Read and write (modify app permissions) |
| Owner | aashishpandey_i |
| Owner ID | 2889837547 |

Application Actions

Regenerate Consumer Key and Secret

Change App Permissions

Your Access Token



You haven't authorized this application for your own account yet.

By creating your access token here, you will have everything you need to make API calls right away. The access token generated will be assigned your application's current permission level.

Token Actions

Create my access token

Step 5 - In this page you can see all the keys and token details. Copy Consumer Key, Consumer Secret, Access Token and Access Secret for your python code.

 Application Management 

tweepy_demo_app

Test OAuth

Details

Settings

Keys and Access Tokens

Permissions

Application Settings

Keep the "Consumer Secret" a secret. This key should never be human-readable in your application.

| | |
|------------------------------|---|
| Consumer Key (API Key) | L [REDACTED] |
| Consumer Secret (API Secret) | [REDACTED] |
| Access Level | Read and write (modify app permissions) |
| Owner | aashishpandey_i |
| Owner ID | 2889837547 |

Application Actions

Regenerate Consumer Key and Secret

Change App Permissions

Your Access Token

This access token can be used to make API requests on your own account's behalf. Do not share your access token secret with anyone.

| | |
|---------------------|-----------------|
| Access Token | [REDACTED] |
| Access Token Secret | [REDACTED] |
| Access Level | Read and write |
| Owner | aashishpandey_i |
| Owner ID | 2889837547 |

Chapter 3

- Getting Started

After creating the app, we will copy Consumer Key, Consumer Secret, Access Token and Access Token Secret, and paste in python code as shown below. We need these credentials to authenticate our twitter account with twitter API in python. In the example below I am authenticating python session with twitter API and then downloading home timeline tweets from my twitter account. After that I am printing first 5 tweets for demonstration purpose.

```
# Define API Details -
consumer_key = '9nfJfIfxef36oDhsxODIKdHDx'
consumer_secret = 'wxtb3Z3h59BxdK6UxKhvaiLp1N7Z8FnHQ3QTqwBoda0REPRUQf'
access_token = '2889837547-oDJqC5bT2TUwb6p9rdFZYWseCHluXNIiH1wZedv'
access_token_secret = 'KDuRUTTP7TYjVSwALEVsJWsyv7vsG0X75NngF0p0kVomj'
# Import tweepy module
import tweepy
# Create OAuthHandler instance
auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
# Set Access token
auth.set_access_token(access_token, access_token_secret)
# Create Authenticated api connection
api = tweepy.API(auth)
# Download home timeline tweets
public_tweets = api.home_timeline()
# Print first 5 tweets
for tweet in public_tweets[0:4]:
    print(tweet.text)
```

```
## Rs 6-trn direct tax collections rescues govt in fiscally difficult year https://t.co/UsYtbEbInt
## RT @Snowden: The journalists exposing the #Aadhaar breach deserve an award, not an investigation. If
## UIDAI firewalls 5,000 officials from Aadhaar portal following the 'breach' | ETtech https://t.co/zz
## 30 Khan market eateries may be sealed today | ET Retail https://t.co/W9HI0InAdo
```


Chapter 4

- Getting all tweets from a user

Often we want to download recent tweets from a twitter account. Let's say we want to download the tweets for screen name 'CocaCola', 'pepsi' and 'redbull'. In the code below, we import 'tweepy' and 'csv' modules in python and declare API credentials. Then we create a list of three screen names and define a for loop to download tweets for these three screen names and write tweets to respective csv files.

Here we are using `user__timeline` function from tweepy module to get the users tweets. In one query twitter API provides 200 tweets so we put the condition on maxid to get all the tweets. Please note that Twitter API provides upto 3200 tweets for a screen name/ user.

```
import tweepy
import csv
consumer_key = '9nfJfIfxef36oDhsxODIKdHDx'
consumer_secret = 'wxtb3Z3h59BxdK6UxKhvaiLp1N7Z8FnhQ3QTqwBoda0REPRUQf'
access_token = '2889837547-oDJqC5bT2TUwb6p9rdFZYWseCHluXNIiH1wZedv'
access_token_secret = 'KDuRUTTP7TYjVSwALEVsJWsyv7vsG0X75NngF0p0kVomj'

#Twitter only allows access to a users most recent 3200 tweets with this method
#authorize twitter, initialize tweepy
auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_token_secret)
api = tweepy.API(auth)
screen_names = ['@CocaCola', '@pepsi', '@redbull']
for screen_name in screen_names:
    #initialize a list to hold all the tweepy Tweets
    alltweets = []

    #make initial request for most recent tweets (200 is the maximum allowed count)
    new_tweets = api.user_timeline(screen_name = screen_name,count=200)

    #save most recent tweets
    alltweets.extend(new_tweets)
    #.encode("utf-8")
    #outtweets0 = [[tweet.id_str, tweet.created_at, tweet.text] for tweet in new_tweets]

    #save the id of the oldest tweet less one
    oldest = alltweets[-1].id - 1
    #keep grabbing tweets until there are no tweets left to grab

    while len(new_tweets) > 0:
```

```

print("getting tweets before %s" % (oldest))

#all subsiquent requests use the max_id param to prevent duplicates
new_tweets = api.user_timeline(screen_name = screen_name,count=200,max_id=oldest)

#save most recent tweets
alltweets.extend(new_tweets)

#update the id of the oldest tweet less one
oldest = alltweets[-1].id - 1
# print("...%s tweets downloaded so far" % (len(alltweets)))
#transform the tweepy tweets into a 2D array that will populate the csv

outtweets = [[tweet.id_str, tweet.created_at, tweet.text] for tweet in alltweets]

# print the first 5 results
for tweet in outtweets[0:4]:
    print(tweet)
# Write tweets to a csv file

with open('%s_tweets.csv' % screen_name, 'w', encoding='utf-8') as f:
    writer = csv.writer(f)
    writer.writerow(["id","created_at","text"])
    writer.writerows(outtweets)
    f.close()

```

```

## getting tweets before 946826291633950721
getting tweets before 944675229686751232
getting tweets before 942751424143904767
getting tweets before 940639891179954175
getting tweets before 938079148572016640
getting tweets before 935369897361727488
getting tweets before 932459339696230400
getting tweets before 929338060386525183
getting tweets before 927624277763018751
getting tweets before 925068538951368703
getting tweets before 922840841642790911
getting tweets before 920636836157980671
getting tweets before 918275375809138688
getting tweets before 916020340203155455
getting tweets before 912738468664668159
getting tweets before 910214554650595329
getting tweets before 910153711430852611
['950605914226216961', datetime.datetime(2018, 1, 9, 5, 51, 26), 'Cheers to @AlabamaFTBL for taking home the trophy']
['950515610655522817', datetime.datetime(2018, 1, 8, 23, 52, 35), 'Alright football fans, show your spirit!']
['950487161358430209', datetime.datetime(2018, 1, 8, 21, 59, 33), '@crjowens Her name looks good on that jersey']
['950486902821539841', datetime.datetime(2018, 1, 8, 21, 58, 31), '@crjowens While we couldn't put every']
getting tweets before 931197735881134079
getting tweets before 913480632323493887
getting tweets before 897473005701521407
getting tweets before 879755804269740031
getting tweets before 864831361206693887
getting tweets before 837688828936794111
getting tweets before 825720686878588928

```



```

getting tweets before 813426888140062719
getting tweets before 810907773756207104
getting tweets before 807423056255844351
getting tweets before 796735188248199167
getting tweets before 791307210437648383
getting tweets before 778345934124158975
getting tweets before 770644653527994367
getting tweets before 766661486513164290
getting tweets before 763451994250043391
getting tweets before 763422483655655423
['950416786830319623', datetime.datetime(2018, 1, 8, 17, 19, 54), '@preciousjudee Judith! Local available
['950416001107222529', datetime.datetime(2018, 1, 8, 17, 16, 47), "@CarrieLaird3 We're so sorry for your
['950381765037903873', datetime.datetime(2018, 1, 8, 15, 0, 44), "@Bethsdream Beth, we're sorry about the
['950064570420391936', datetime.datetime(2018, 1, 7, 18, 0, 19), 'TFW someone tells you your team is lo
getting tweets before 943251347134124031
getting tweets before 941365754318000127
getting tweets before 941018667894099967
getting tweets before 940722909290655743
getting tweets before 940665397866651647
getting tweets before 940625890467344383
getting tweets before 940606783328018432
getting tweets before 938167996371079167
getting tweets before 935580080159559679
getting tweets before 931271402795225089
getting tweets before 928375622820417535
getting tweets before 923658897382170623
getting tweets before 902319901116915711
getting tweets before 892876523572690943
getting tweets before 890278454150782975
getting tweets before 885984393793339391
getting tweets before 885917982366601216
['949410306731671552', datetime.datetime(2018, 1, 5, 22, 40, 30), '@iamcamtyus Off to a great day! When
['949409709068554241', datetime.datetime(2018, 1, 5, 22, 38, 8), '@ChaseAfterThis_ Ever had a Red Bull
['949409574188089350', datetime.datetime(2018, 1, 5, 22, 37, 36), '@ppxolv Congrats! What type of inter
['949409027007631360', datetime.datetime(2018, 1, 5, 22, 35, 25), '@lalalalindsaay Congrats! What is yo

```


Chapter 5

- Searching Tweets with a hashtag or Keywords

We can search the tweets using twitter API however standard search API has limitations. See the note below from Twitter API documentation -

Please note that Twitter's search service and, by extension, the Search API is not meant to be an exhaustive source of Tweets. Not all Tweets will be indexed or made available via the search interface.

If you need all the tweets for a particular keyword, then you have to signup for Premium search API or Enterprise search APIs. In the code below, we import 'tweepy' and 'csv' modules in python and declare API credentials. Then we create a list of three hashtags (Key words - '#ola', '#uber', '#fraud') and define a for loop to download tweets for these three hashtags and write tweets to respective csv files.

```
import tweepy
import csv
consumer_key = '9nfJfIfxef36oDhsxODIKdHDx'
consumer_secret = 'wxtb3Z3h59BxdK6UxKhvaiLp1N7Z8FnHQ3QTqwBodaOREPRUQf'
access_token = '2889837547-oDJqC5bT2TUwb6p9rdFZYWseCHluXNIiH1wZedv'
access_token_secret = 'KDuRUTTP7TYjVSwALEVsJWsyv7vsGOX75NngF0p0kVomj'
#authorize twitter, initialize tweepy
auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_token_secret)
api = tweepy.API(auth)
# Search list
hashtags = ['#ola', '#uber', '#fraud']
for hashtag in hashtags:

    #initialize a list to hold all the tweepy Tweets
    alltweets = []

    #make initial request for most recent tweets (100 is the maximum allowed count)
    new_tweets = api.search(q = hashtag, count=100, lang = 'en' )
    alltweets.extend(new_tweets)
    oldest = alltweets[-1].id - 1

#    print("...%s tweets downloaded so far" % (len(alltweets)))

    pagecnt = 2
```

```

while len(new_tweets) > 0 and pagecnt < 41:
    new_tweets = api.search(q = hashtag, count=100, lang = 'en', max_id=oldest )
    alltweets.extend(new_tweets)
    pagecnt = pagecnt+1
    # print("...%s tweets downloaded so far" % (len(alltweets)))

outtweets = [[tweet.id_str, tweet.created_at, tweet.text] for tweet in alltweets]
# print the first 5 results
for tweet in outtweets[0:4]:
    print(tweet)

# Write tweets to csv file
with open('%s_tweets.csv' % hashtag, 'w', encoding='utf-8') as f:
    writer = csv.writer(f)
    writer.writerow(["id","created_at","text"])
    writer.writerows(outtweets)
    f.close()

```

```

## ['950671031911424000', datetime.datetime(2018, 1, 9, 10, 10, 11), 'We provide best #TaxiBookingApp D
['950669897939341312', datetime.datetime(2018, 1, 9, 10, 5, 40), 'Ola, ICICI Bank offer innovative solu
['950637699215499264', datetime.datetime(2018, 1, 9, 7, 57, 44), 'RT @PerceptEMC: #Ola @SunburnFestival
['950633559731052544', datetime.datetime(2018, 1, 9, 7, 41, 17), '@Olacabs books @Uber in Rajasthan.\n\
['950672757834047489', datetime.datetime(2018, 1, 9, 10, 17, 2), 'RT @AirAsia: [Press Release] #AirAsia
['950672696135864320', datetime.datetime(2018, 1, 9, 10, 16, 48), 'RT @paultmadden: Good afternoon @TfL
['950672487729238021', datetime.datetime(2018, 1, 9, 10, 15, 58), 'RT @ubg700: #Uber #FREE Get up to 6 p
['950672343696855041', datetime.datetime(2018, 1, 9, 10, 15, 24), 'RT @AirAsia: [Press Release] #AirAsia
['950672358154551296', datetime.datetime(2018, 1, 9, 10, 15, 27), '@eBayAU So over 5 weeks ago, i bough
['950671952993234944', datetime.datetime(2018, 1, 9, 10, 13, 50), "Hey @realDonaldTrump I'd be willing
['950671744662167552', datetime.datetime(2018, 1, 9, 10, 13, 1), 'Cabinet Office puts out market feeler
['950670848750510080', datetime.datetime(2018, 1, 9, 10, 9, 27), 'Dangerous game.. #travel #fraud - Tra

```