

Sentiment Analysis - Tweet Scraping

1. Import modul yang diperlukan, apabila belum diinstall silahkan install terlebih dahulu melalui command line dengan perintah : Python -m pip install [nama modul]

```
import tweepy
import sys
import jsonpickle
import time
from datetime import datetime
from datetime import timedelta
```

2. Melakukan otentikasi untuk dapat mengakses Twitter API, apabila belum memiliki key dan token silahkan daftar atau login ke Teitter Develover Account dan buatlah key dan token

```
consumer_key = 'XX'
consumer_secret = 'XX'
access_token = 'XX'
access_token_secret = 'XX'

auth = tweepy.AppAuthHandler(consumer_key,consumer_secret)
```

3. Menyusun pencarian berdasarkan kata kunci dan filter tertentu

```
since =(datetime.now() + timedelta(days=-1)).strftime("%Y-%m-%d")
qry='kebakaran hutan -filter:links -filter:images -filter:videos '+'since:'+since
maxTweets = 600 # Isi sesuai dengan kebutuhan anda
tweetsPerQry = 100 # dibatasi Twitter maksimum 100
fName=since+'_Tweets.json' # Nama file hasil scraping

api = tweepy.API(auth, wait_on_rate_limit=True,wait_on_rate_limit_notify=True)
```

4. Menjalankan pencarian tweet dan menyimpannya dalam file

```

if (not api):
    sys.exit('Autentikasi gagal, mohon cek "Consumer Key" & "Consumer Secret" Twitter anda')

sinceld, max_id, tweetCount = None, -1, 0

print("Mulai mengunduh maksimum {0} tweets".format(maxTweets))
with open(fName,'w') as f:
    while tweetCount < maxTweets:
        try:
            if (max_id <= 0):
                if (not sinceld):
                    new_tweets=api.search(q=qry,count=tweetsPerQry, tweet_mode='extended')
                else:
                    new_tweets=api.search(q=qry,count=tweetsPerQry,since_id=sinceld,
tweet_mode='extended')
            else:
                if (not sinceld):
                    new_tweets=api.search(q=qry,count=tweetsPerQry,max_id=str(max_id - 1),
tweet_mode='extended')
                else:
                    new_tweets=api.search(q=qry,count=tweetsPerQry,max_id=str(max_id -
1),since_id=sinceld, tweet_mode='extended')
            if not new_tweets:
                print('Tidak ada lagi Tweet ditemukan dengan Query="{0}"'.format(qry));break
            for tweet in new_tweets:
                if(tweet.created_at.date() < datetime.now().date()) :
                    f.write(jsonpickle.encode(tweet._json,unpicklable=False)+'\n')
            tweetCount+=len(new_tweets)
            sys.stdout.write("\r");sys.stdout.write("Jumlah Tweets telah tersimpan: %.0f"
%tweetCount);sys.stdout.flush()
            max_id=new_tweets[-1].id
        except tweepy.TweepError as e:
            print("some error : " + str(e));break # Aya error, keluar
    print ("\nSelesai! {0} tweets tersimpan di "{1}"".format(tweetCount,fName))

```

5. Menampilkan anatomi dari objek tweet

```

with open('Hasil_Tweets.json', 'r') as f:
    #mengambil tweet pada baris pertama
    text = f.readline()
    tweet = json.loads(text)
    #menampilkan twit dalam format json pada console
    print(json.dumps(tweet, indent=4))

```

TUGAS

Buatlah file csv yang berisi id, created_at, dan full_text dari seluruh tweet hasil pencarian yang telah tersimpan dalam format json.

Hint:

```
#Silahkan lengkapi code dibawah ini
import json
import csv

csvFile = open('.....', 'a')
csvWriter = csv.writer(.....)

for l in open(.....):
    tw = json.loads(.....)
    csvWriter. ....

csvFile.close()
```