

Code used when collecting tweets.

Use of this code is March 31, 2024, the last day to collect tweets. Operation confirmation is the same.

```
#import the library
import tweepy
from datetime import datetime, timezone
import pytz
import pandas as pd
import os
import schedule
from time import sleep

#authentication to twitter
api_key = "*****"
api_secret = "*****"
access_key = "*****"
access_secret = "*****"

auth = tweepy.OAuthHandler(api_key, api_secret)
auth.set_access_token(access_key, access_secret)
api = tweepy.API(auth)

# Setting search conditions
searchkey = "Japanese search word"
item_num = 800

#collect tweet
tweets = tweepy.Cursor(api.search,q=searchkey,lang='ja').items(item_num)

# Function: convert Greenwich Mean Time to Japan Time
def change_time_JST(u_time):
    utc_time = datetime(u_time.year, u_time.month,u_time.day, u_time.hour,u_time.minute,u_time.second,
tzinfo=timezone.utc)
    jst_time = utc_time.astimezone(pytz.timezone("Asia/Tokyo"))
    str_time = jst_time.strftime("%Y-%m-%d %H:%M:%S")
    return str_time
```

```

# extract information from collected data
tweet_date = []
for tweet in tweets:
    tweet_time = change_time_JST(tweet.created_at)
    tweet_date.append([
        tweet.id,
        tweet_time,
        tweet.text,
        tweet.user.id,
        tweet.user.screen_name,
        tweet.user.description])

#output to csv
labels=[ Japanese label name ]
df = pd.DataFrame(tweet_date,columns=labels)
file_name='*****'
file_path = os.path.join(current_path,file_name)
file_check = os.path.isfile(file_path)
else:
    df_csv = pd.read_csv(file_name)
    df_merge = pd.concat([df_csv,df])
    df_merge.to_csv(file_name,encoding='Shift jis', index=False)

# repeat execution
if __name__=='__main__':
    schedule.every(1).hours.do(main)
    While True:
        schedule.run_pending()
        sleep(120)

```

Data collection requires API key, API key secret, access key, and access secret. The data items collected as part of this were: tweet_ID, tweet_time, tweet_text, tweet_user_ID, tweet_user_screen_name, tweet_user_name, tweet_user_description. The data was gathered in Greenwich Mean Time, but was converted to Japan Standard Time.