# Twitter Crawler

Objective

To crawl tweets posted in Singapore contains POI keyword.

Dependence

Python libraries

selenium

scrapy

pandas

datetime

csv

twint

chromedriver: this has to be the same version with your Chrome.

Run twitter.py will crawl tweets posted in SG which contains POI keyword for you.

Run user.py will crawl the twitter user profile for you.

Current code requires header column names to append the data

Notes for important code snippets

1. Reads in the name column (from ta\_poi.csv file) in list using poi from tripadvisor to construct the url. In tweets.csv, poi name from ta\_poi.csv is reflected in keyword column.

file = pd.read\_csv('./twitter/ta\_poi.csv')

2. Getting most recent tweets (over past 3 days)

start\_date = (datetime.now() - timedelta(days=4)).strftime('%Y-%m-%d')

3. Code requires existing csv files (tweets.csv and users\_profile.csv) with correct column headers

Command line arguments (working directory is social\_media\_crawlers)

./twitter/twitter.py

./twitter/user.py