Twitter Scraping

Bashair Altalhi

Text as Data

April 4, 2022

Introduction

Scraping Tweets with API

Type of APIs:

- API Comes when you sign to Twitter developer account **but has limitations.**
- API for Academic research **access to ten millions tweets monthly and significant parameters**.

Requirements:

1. Python
2. Tweepy: Python library
3. Luck!

Scraping Tweets without API-Twint

Using Python tools:

- Twint: Twitter Intelligence Tool. It allows scrape tweets relating to certain topics, hashtags, or sort out sensitive information from Tweets like e-mail and phone numbers **but**
    1. For Windows needs manual set up.
    2. Some parameters not working.
    3. Not supporting Arabic very well.

Scraping Tweets without API- Snscrape

Using Python tools:

- Snscrape: is a scraper for social networking services. It extract user profiles, hashtags, or searches. It has no limit to max number of tweets.

Requirements:

1. Python 3.8 or higher.
2. Google Colab **or** Windows Command and Visual Studio Code.
3. Be careful with " " and ' ', and dates.

Scraping Tweets without API- Snscrape

- **Some significant parameters**
  1. jsonl : get the data into jsonl format.
  2. max-results : limit the number of tweets to collect.
  3. likecount :number of like.
  4. User: user name , id,etc.
  5. the date: date tweet was created.

Scraping Tweets without API- Snscrape

- Snscrape

Example 1: Find the total number of tweets for search term.

Example 2: Extracting tweets by search term.

Example 3: Extracting tweets from users.

Example 4: Extracting tweets by search term and users.