

LAMPIRAN

CRAWLING DATA TWEET (X)

```

# Import required Python package
!pip install pandas

# Install Node.js (because tweet-harvest built using Node.js)
!sudo apt-get update
!sudo apt-get install -y ca-certificates curl gnupg
!sudo mkdir -p /etc/apt/keyrings
!curl -fsSL https://deb.nodesource.com/gpgkey/nodesource-repo.gpg.key | sudo gpg --dearmor -o /etc/apt/keyrings/nodesource.gpg

!NODE_MAJOR=20 && echo "deb [signed-by=/etc/apt/keyrings/nodesource.gpg] https://deb.nodesource.com/node_${NODE_MAJOR}.x nodistro main" | s

!sudo apt-get update
!sudo apt-get install nodejs -y

!node -v
-----
Get:12 http://security.ubuntu.com/ubuntu jammy-security/universe amd64 Packages [1,047 kB]
Hit:13 https://ppa.launchpadcontent.net/ubuntugis/ppa/ubuntu jammy InRelease
Get:14 http://archive.ubuntu.com/ubuntu jammy-updates/universe amd64 Packages [1,307 kB]
Get:15 https://ppa.launchpadcontent.net/c2d4u.team/c2d4u4.0+/ubuntu jammy/main Sources [2,257 kB]
Get:16 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 Packages [1,606 kB]
Get:17 http://archive.ubuntu.com/ubuntu jammy-updates/restricted amd64 Packages [1,611 kB]
Get:18 https://ppa.launchpadcontent.net/c2d4u.team/c2d4u4.0+/ubuntu jammy/main amd64 Packages [1,158 kB]
Fetched 12.2 MB in 2s (6,013 kB/s)
Reading package lists... Done
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
ca-certificates is already the newest version (20230311ubuntu0.22.04.1).
curl is already the newest version (7.81.0-1ubuntu1.15).
gnupg is already the newest version (2.2.27-3ubuntu2.1).
gnupg set to manually installed.
0 upgraded, 0 newly installed, 0 to remove and 27 not upgraded.
deb [signed-by=/etc/apt/keyrings/nodesource.gpg] https://deb.nodesource.com/node_20.x nodistro main
Hit:1 https://cloud.r-project.org/bin/linux/ubuntu jammy-cran40/ InRelease
Hit:2 http://security.ubuntu.com/ubuntu jammy-security InRelease
Hit:3 http://archive.ubuntu.com/ubuntu jammy InRelease
Hit:4 https://developer.download.nvidia.com/compute/cuda/repos/ubuntu2204/x86_64 InRelease
Get:5 https://deb.nodesource.com/node_20.x nodistro InRelease [12.1 kB]
Hit:6 http://archive.ubuntu.com/ubuntu jammy-updates InRelease
Hit:7 http://archive.ubuntu.com/ubuntu jammy-backports InRelease
Hit:8 https://ppa.launchpadcontent.net/c2d4u.team/c2d4u4.0+/ubuntu jammy InRelease
Hit:9 https://ppa.launchpadcontent.net/deadsnakes/ppa/ubuntu jammy InRelease
Get:10 https://deb.nodesource.com/node_20.x nodistro/main amd64 Packages [4,270 B]
Hit:11 https://ppa.launchpadcontent.net/graphics-drivers/ppa/ubuntu jammy InRelease
Hit:12 https://ppa.launchpadcontent.net/ubuntugis/ppa/ubuntu jammy InRelease
Fetched 16.4 kB in 1s (12.2 kB/s)
Reading package lists... Done
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
The following NEW packages will be installed:
  nodejs
0 upgraded, 1 newly installed, 0 to remove and 27 not upgraded.
Need to get 31.3 MB of archives.
After this operation, 195 MB of additional disk space will be used.
Get:1 https://deb.nodesource.com/node_20.x nodistro/main amd64 nodejs amd64 20.10.0-1nodesource1 [31.3 MB]
Fetched 31.3 MB in 0s (82.6 MB/s)
debconf: unable to initialize frontend: Dialog
debconf: (No usable dialog-like program is installed, so the dialog based frontend cannot be used. at /usr/share/perl5/Debconf/F
debconf: falling back to frontend: Readline
debconf: unable to initialize frontend: Readline
debconf: (This frontend requires a controlling tty.)
debconf: falling back to frontend: Teletype
dpkg-preconfigure: unable to re-open stdin:
Selecting previously unselected package nodejs.
(Reading database ... 121654 files and directories currently installed.)
Preparing to unpack .../nodejs_20.10.0-1nodesource1_amd64.deb ...
Unpacking nodejs (20.10.0-1nodesource1) ...
Setting up nodejs (20.10.0-1nodesource1) ...
Processing triggers for man-db (2.10.2-1) ...
v20.10.0

```

▼ Import Library

```

import time
import datetime
import pandas as pd
import numpy as np

```

✓ Fungsi list tanggal per bulan

```
def get_dates_in_month(year, month):
    # Create the first day of the specified month
    first_day = datetime.date(year, month, 1)

    # Find the number of days in the month
    if month == 12:
        next_month = datetime.date(year + 1, 1, 1)
    else:
        next_month = datetime.date(year, month + 1, 1)
    last_day = next_month - datetime.timedelta(days=1)

    # Generate a list of dates for the month
    date_list = [(first_day + datetime.timedelta(days=i)).strftime("%Y-%m-%d") for i in range((last_day - first_day).days + 1)]

    return date_list
```

✓ Ambil Tweet per bulan dalam setahun

```
tahun = 2022

for bulans in range(1, 13):
    tanggal = get_dates_in_month (tahun, bulans)

    split_tanggal = np.array_split(tanggal, 10)

    for st in split_tanggal:
        try:
            since = st[0]
            until = st[-1]

            filename = since+'-'+until+'.csv'
            search_keyword = 'undang undang ibu kota nusantara -is:retweet lang:id since:'+since+' until:'+until
            limit = 5000

            !npx --yes tweet-harvest@latest -o "{filename}" -s "{search_keyword}" -l {limit} --token "a5ea55dbb96792d4cafbed28fccbc3db8329ccc"

            # # Specify the path to your CSV file
            file_path = f"tweets-data/{filename}"

            # # Read the CSV file into a pandas DataFrame
            df = pd.read_csv(file_path, delimiter=";")

            # Cek jumlah data yang didapatkan
            num_tweets = len(df)

            if num_tweets <= 0 :
                with open('errors.txt', 'a') as f:
                    f.writelines("['+since+', '"+until+"']"+'\n')

            print(f"Jumlah tweet dalam dataframe adalah {num_tweets}.")
            print("\n\n")

            print('Sleep time, prossess will be auto start after 10 minute \n\n')
            time.sleep(600)
        except:
            with open('errors.txt', 'a') as f:
                f.writelines("['+since+', '"+until+"']"+'\n')

            print('Sleep time, prossess will be auto start after 15 minute \n\n')
            time.sleep(900)
            pass
```



```
libpango-1.0-0 is already the newest version (1.50.0-1ubuntu1).
libwayland-client0 is already the newest version (1.20.0-1ubuntu0.1).
libx11-6 is already the newest version (2:1.7.5-1ubuntu0.3).
xvfb is already the newest version (2:21.1.4-2ubuntu1.7~22.04.5).
0 upgraded, 0 newly installed, 0 to remove and 27 not upgraded.
```

Opening twitter search page...

Filling in keywords: undang undang ibu kota nusantara -is:retweet lang:id since:2022-01-05 until:2022-01-07

Got some tweets, saving to file...

Your tweets saved to: /content/tweets-data/2022-01-05-2022-01-07.csv

Total tweets saved: 1

Got some tweets, saving to file...

No more tweets found, please check your search criteria and csv file result

Already got 1 tweets, done scrolling...

Jumlah tweet dalam dataframe adalah 1.

Sleep time, proses will be auto start after 10 minute

Welcome to the Twitter Crawler 🕷

This script uses Chromium Browser to crawl data from Twitter with *your* Twitter auth token.
Please enter your Twitter auth token when prompted.

Note: Keep your access token secret! Don't share it with anyone else.

Note: This script only runs on your local device.

up to date, audited 4 packages in 1s

found 0 vulnerabilities

Installing dependencies...

Hit:1 <https://cloud.r-project.org/bin/linux/ubuntu/jammy-cran40/> InRelease

Hit:2 <http://security.ubuntu.com/ubuntu/jammy-security> InRelease

Hit:3 https://developer.download.nvidia.com/compute/cuda/repos/ubuntu2204/x86_64/ InRelease

```
from google.colab import drive
drive.mount('/content/drive')
```

↪ Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).

✓ Ambil Tweet Manual

```
since = '2022-01-01'
until = '2022-06-31'
```

```
filename = since+'-'+until+'.csv'
search_keyword = 'undang undang ibu kota nusantara -is:retweet lang:id since:'+since+' until:'+until
limit = 1000
```

```
!npx --yes tweet-harvest@latest -o "{filename}" -s "{search_keyword}" -l {limit} --token "a5ea55dbb96792d4cafbed28fccbc3db8329ccc"
```

```
from google.colab import drive
drive.mount('/content/drive')
```

↪ Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).

