

LAPORAN
PRAKTIKUM BIG DATA ANALYTIC
Pertemuan Ke - 4



Dosen :
Sri Redjeki, S.Si., M.Kom.

Disusun oleh :
RAHADIYAN BONDAN PERMADI
215411119

Universitas Teknologi Digital Indonesia
UTDI
YOGYAKARTA
2022

Dasar Teori

Data Scrapping merupakan kegiatan untuk menemukan data dan mengumpulkannya. Data diambil secara langsung dari sebuah sumber, tidak hanya dari web tetapi bisa dari berbagai sumber dimana terdapat data. Ini mencakup spreadsheets, storage devices, dll. Proses ini diperlukan untuk menyaring (filtering) dan membedakan berbagai jenis data mentah dari berbagai sumber menjadi sesuatu yang berguna dan informatif.

Data Crawling adalah cara mengumpulkan semua data dalam suatu website dengan menggunakan metode tertentu dan secara otomatis (web crawling). Web crawler akan mengunjungi setiap alamat website yang diberikan kepadanya, kemudian menyerap (scraping) dan menyimpan semua informasi yang terkandung didalam website tersebut. Proses crawling dalam suatu website dimulai dari mendata seluruh url dari website, menelusurinya satu-persatu, kemudian memasukkannya dalam daftar halaman pada indeks search engine, sehingga setiap kali ada perubahan pada website, akan terupdate secara otomatis. Web crawling adalah proses mengambil kumpulan halaman 2 dari sebuah web untuk dilakukan pengindeksan sehingga mendukung kinerja mesin pencari.

Kebutuhan Alat

1. Python (Anaconda / Miniconda)
2. Jupyter Notebook / Google Colab
3. Twint / Tweepy

Langkah – Langkah dalam praktikum

1. Melakukan crawling data twitter menggunakan twint dengan Jupyter Notebook menggunakan Keyword “Omicron di Indonesia atau vaksin booster”

Dengan mengetikan coding dibawah ini (per baris)

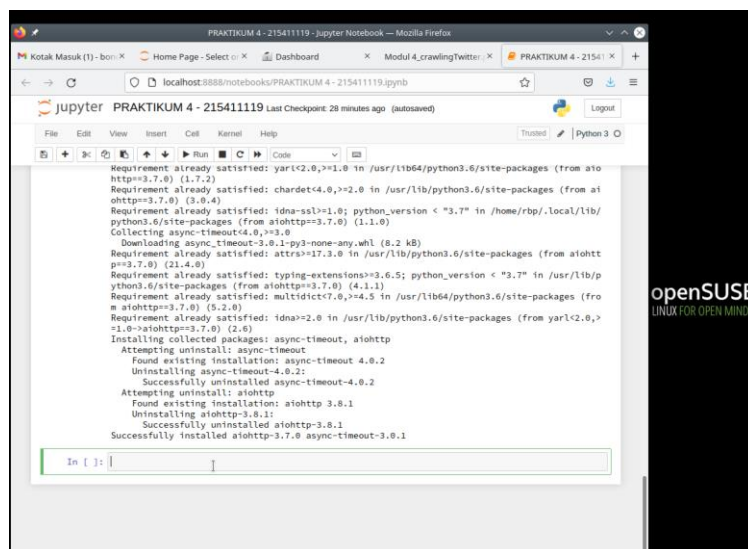
```
!git clone --depth=1
```

```
https://github.com/twintproject/twint.git
```

```
%cd twint
```

```
!pip3 install . -r requirements.txt
```

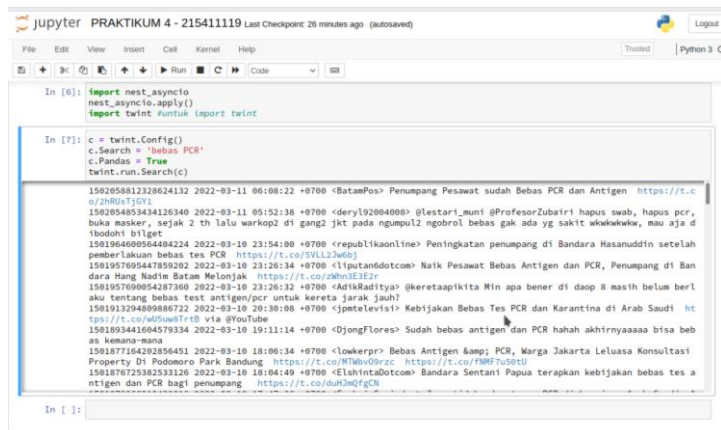
```
!pip install aiohttp==3.7.0
```



```
PRAKTIKUM 4 - 215411119 - Jupyter Notebook - Mozilla Firefox
Kotak Masuk (1) - bor... Home Page - Select... Dashboard Modul_4_crawlingTwitter... PRAKTIKUM 4 - 215411119
localhost:8888/notebooks/PRAKTIKUM 4 - 215411119.ipynb
jupyter PRAKTIKUM 4 - 215411119 Last Checkpoint: 28 minutes ago (autosaved) Logout
File Edit View Insert Cell Kernel Help Trusted Python 3
In [ ]: !git clone --depth=1 https://github.com/twintproject/twint.git
!cd twint
!pip3 install . -r requirements.txt
!pip install aiohttp==3.7.0
Requirement already satisfied: yarl<2.0,>=1.0 in /usr/lib64/python3.6/site-packages (from aiohttp==3.7.0) (1.7.2)
Requirement already satisfied: chardet<4.0,>=2.0 in /usr/lib/python3.6/site-packages (from aiohttp==3.7.0) (3.0.4)
Requirement already satisfied: idna-ssl<1.0,>=2.0 in /home/rbp/.local/lib/python3.6/site-packages (from aiohttp==3.7.0) (1.1.0)
Collecting async-timeout<4.0,>=3.0
  Downloading async_timeout-3.0.1-py3-none-any.whl (8.2 kB)
Requirement already satisfied: attrs>=17.3.0 in /usr/lib/python3.6/site-packages (from aiohttp==3.7.0) (21.4.0)
Requirement already satisfied: typing-extensions>=3.6.5; python_version < "3.7" in /usr/lib/python3.6/site-packages (from aiohttp==3.7.0) (4.1.1)
Requirement already satisfied: multidict<7.0,>=4.5 in /usr/lib64/python3.6/site-packages (from aiohttp==3.7.0) (5.2.0)
Requirement already satisfied: idna>=2.0 in /usr/lib/python3.6/site-packages (from yarl<2.0,>=1.0->aiohttp==3.7.0) (2.6)
Installing collected packages: async-timeout, aiohttp
  Attempting uninstall: async-timeout
    Found existing installation: async-timeout 4.0.2
    Uninstalling async-timeout-4.0.2:
      Successfully uninstalled async-timeout-4.0.2
  Attempting uninstall: aiohttp
    Found existing installation: aiohttp 3.8.1
    Uninstalling aiohttp-3.8.1:
      Successfully uninstalled aiohttp-3.8.1
  Successfully installed aiohttp-3.7.0 async-timeout-3.0.1
```


Melakukan scraping menggunakan keyword umum berbahasa Indonesia.

```
import nest_asyncio
nest_asyncio.apply()
import twint #untuk import twint
c = twint.Config()
c.Search = 'bebas PCR'
c.Pandas = True
twint.run.Search(c)
```



Hasil dengan format (.csv)

TUGAS :

Melakukan crawling data untuk keyword minyak goreng langka hanya untuk bulan Februari 2022.

Langkah Pertama kita Import

```
import nest_asyncio
```

```
nest asyncio.apply()
```

```
import twint
```

Kemudian ketika script dibawah ini :

```
c = twint.Config()
```

c.Search = 'Minyak Goreng Langka'

c.Sience = '2022-2-1'

c.Until = '2022-2-28'

```
c.Output = 'KasusMinyakGorengLangkaDitengahPandemi.csv'
```

twint.run.Search(c)

Hasilnya :

```
jupyter PRAKTIKUM 4 - 215411119 Last Checkpoint: an hour ago (autosaved)
File Edit View Insert Cell Kernel Help Trusted Python 3 C
In [4]: import nest_asyncio
nest_asyncio.apply()
import twint

In [14]: c = twint.Config()
c.Search = 'Minyak Goreng Langka'
c.Since = '2022-2-1'
c.Until = '2022-2-28'
c.Output = 'KasusMinyakGorengLangkaDitengahPandemi.csv'

twint.run.Search(c)

1497977050053095425 2022-02-27 23:48:54 +0700 <Crysantium_12> Njir jadi barang langka dah kayak minyak goreng. Kenapa lo gk sekalian beli pabriknya aja tae 🤔🤔🤔
149797539058058353 2022-02-27 23:42:18 +0700 <jkthvenus> @convomfs tau ga persamaan km sm minyak goreng? sama-sama langka 🤔
1497975371463299075 2022-02-27 23:42:14 +0700 <NovitaS08789871> @Saharud65305790 @anticebong @mohmahfudm Periode ke tiga, bisa² bukan hanya minyak goreng yg langka dan mahal... Jgn sampai beras jadi langka. Yg kreatif dikit lah, Indo nesia butuh pemtapiin yg tegas. Amallah. Cukup sudah penderitaan
1497974567720996865 2022-02-27 23:39:02 +0700 <UmiUm081170> @keuangannews_id Maju terus pak ubed... Jgn sampe terge rus dg kasus langka nya minyak goreng...
1497973742336503810 2022-02-27 23:35:45 +0700 <AHYS14P> Sindir Wacana Tunda Pemilu, AHY: 'Minyak Goreng Mahal dan Lan gka Kok Bicara Langgengkan kekuasaan!' https://t.co/zUkIPASnz1 https://t.co/SbST2PmSI
1497972665898115074 2022-02-27 23:31:29 +0700 <benjamalz> Seh tank Rusia bisa keabisan bensin. Curiga gue, jgn2 itu t ank sebenarnya pake bahan bakar minyak goreng. Langka, jadi mogok ditengah jalan.
149797127936675840 2022-02-27 23:29:20 +0700 <khofibarrazaqi> @RaniBoeginez Sebenarnya ini Hanya Pengalihan Isu Miny ak Goreng Yang Langka Dipasaran Saja? Buat Presiden Cabut Saja IZIN SAMIT PENGUSAHA2 NAKAL Itu Trek Besi Beton Pake Bambu? Pake Debang Pisang Aja Sekalian Didalam Toa Untuk Redam Tempe Dan BAMBU Puspap
1497968682689241095 2022-02-27 23:15:39 +0700 <capt_10R174> @ChusnulCh_ Kenapa sejenis @ChusnulCh_ gak pernah nyinggung minyak goreng yg mahal en langka ? Karena b gung minyak goreng yg mahal en langka ? karena bot gak butuh minyak, butuh kuota 🤔
1497967564021846024 2022-02-27 23:11:12 +0700 <afvckado> @prawira237 aku kalau cuma buat numis pake butter, jarang pake minyak goreng gara² langka :')
1497965319041667075 2022-02-27 23:02:17 +0700 <AhmadGuntum> @geloraco Pemerintah Indonesia gak usah sok² nherusin masalah rusia vs ukraina, urus aja dulu masalah dalam negeri
1497964797539012613 2022-02-27 23:00:13 +0700 <mclfluryoreo157> ini bahagia atau minyak goreng kok langka banget.
1497963753609908225 2022-02-27 22:56:04 +0700 <embunfajar123> Harga Minyak goreng, kedelai, gula dan daging turun drastis...dan tidak langka lagi...tapi di Metaverse negeri Wakanda
1497963663822843905 2022-02-27 22:55:42 +0700 <DIKEFACEWORLWIDE> @saalearn @catchmeupid Dampaknya minyak goreng makin langka
1497963567559032836 2022-02-27 22:55:19 +0700 <redpunkz> @nuhulmaani 🤔 Pen ketawa takut minyak goreng makin langka 🤔🤔
1497962895917056000 2022-02-27 22:52:39 +0700 <NewsSimada> Parliindungan Purb: Tak Wajar di Sumut Bisa Langka Minyak Goreng https://t.co/SCZKt26Nhr
1497962403275087873 2022-02-27 22:50:42 +0700 <Aziz3490Abd> @fadlizon Menangani minyak goreng langka gk bisa mau gaya2an.
1497956774019682312 2022-02-27 22:28:20 +0700 <fahru1771> @Wah76311930Asep @BandarJuaranya @ChusnulCh_ Goblok emang, minyak goreng langka yg disalahkan anis. pemer
14979487741988122627 2022-02-27 22:00:23 +0700 <azkadotid> Akibat Minyak Goreng Langka dan Harganya Mahal, Ibu Rumah Tangga Muda Mulai Beralih Gunakan Air Fryer https://t.co/b
1497948775779758083 2022-02-27 21:56:33 +0700 <DendiDendy> @budimandiatmiko Waah Indonesia minyak goreng langka dn tempe mahal 🤔
1497947882921807876 2022-02-27 21:53:00 +0700 <renny_setyo> 2020 Makser langka, 2021 Oksigen langka, 2022 Minyak goreng langka 🤔
```

Output (KasusMinyakGorengLangkaDitengahPandemi.csv)

KasusMinyakGorengLangkaDitengahPandemi.csv - LibreOffice Calc									
A20									
1	1497977050053095425	2022-02-27	23:48:54	+0700	<Crysantium_12>	Njir jadi barang langka dah kayak minyak goreng. Kenapa lo gk sekalian beli pabriknya aja tae 🤔🤔🤔			
2	149797539058058353	2022-02-27	23:42:18	+0700	<jkthvenus>	@convomfs tau ga persamaan km sm minyak goreng? sama-sama langka 🤔			
3	1497975371463299075	2022-02-27	23:42:14	+0700	<NovitaS08789871>	@Saharud65305790 @anticebong @mohmahfudm Periode ke tiga, bisa² bukan hanya minyak goreng yg langka d			
4	1497974567720996865	2022-02-27	23:39:02	+0700	<UmiUm081170>	@keuangannews_id Maju terus pak ubed... Jgn sampe terge nus dg kasus langka nya minyak goreng...			
5	1497973742336503810	2022-02-27	23:35:45	+0700	<AHYS14P>	Sindir Wacana Tunda Pemilu, AHY: 'Minyak Goreng Mahal dan Langka Kok Bicara Langgengkan kekuasaan!' https://t.co/zl			
6	1497972665898115074	2022-02-27	23:31:29	+0700	<benjamalz>	Seh tank Rusia bisa keabisan bensin. Curiga gue, jgn2 itu tank sebenarnya pake bahan bakar minyak goreng. Langka, jadi			
7	1497972127936675840	2022-02-27	23:29:20	+0700	<khofibarrazaqi>	@RaniBoeginez Sebenarnya ini Hanya Pengalihan Isu Minyak Goreng Yang Langka Dipasaran Saja? Buat Presiden Ca			
8	1497968682689241095	2022-02-27	23:15:39	+0700	<capt_10R174>	@ChusnulCh_ Kenapa sejenis @ChusnulCh_ gak pernah nyinggung minyak goreng yg mahal en langka ? Karena b			
9	1497967564021846024	2022-02-27	23:11:12	+0700	<afvckado>	@prawira237 aku kalau cuma buat numis pake butter, jarang pake minyak goreng gara² langka :')			
10	1497965319041667075	2022-02-27	23:02:17	+0700	<AhmadGuntum>	@geloraco Pemerintah Indonesia gak usah sok² nherusin masalah rusia vs ukraina, urus aja dulu masalah dalam negeri			
11	1497964797539012613	2022-02-27	23:00:13	+0700	<mclfluryoreo157>	ini bahagia atau minyak goreng kok langka banget.			
12	1497963753609908225	2022-02-27	22:56:04	+0700	<embunfajar123>	Harga Minyak goreng, kedelai, gula dan daging turun drastis...dan tidak langka lagi...tapi di Metaverse negeri Wakanda			
13	1497963663822843905	2022-02-27	22:55:42	+0700	<DIKEFACEWORLWIDE>	@saalearn @catchmeupid Dampaknya minyak goreng makin langka			
14	1497963567559032836	2022-02-27	22:55:19	+0700	<redpunkz>	@nuhulmaani 🤔 Pen ketawa takut minyak goreng makin langka 🤔🤔			
15	1497962895917056000	2022-02-27	22:52:39	+0700	<NewsSimada>	Parliindungan Purb: Tak Wajar di Sumut Bisa Langka Minyak Goreng https://t.co/SCZKt26Nhr			
16	1497962403275087873	2022-02-27	22:50:42	+0700	<Aziz3490Abd>	@fadlizon Menangani minyak goreng langka gk bisa mau gaya2an.			
17	1497956774019682312	2022-02-27	22:28:20	+0700	<fahru1771>	@Wah76311930Asep @BandarJuaranya @ChusnulCh_ Goblok emang, minyak goreng langka yg disalahkan anis. pemer			
18	14979487741988122627	2022-02-27	22:00:23	+0700	<azkadotid>	Akibat Minyak Goreng Langka dan Harganya Mahal, Ibu Rumah Tangga Muda Mulai Beralih Gunakan Air Fryer https://t.co/b			
19	1497948775779758083	2022-02-27	21:56:33	+0700	<DendiDendy>	@budimandiatmiko Waah Indonesia minyak goreng langka dn tempe mahal 🤔			
20	1497947882921807876	2022-02-27	21:53:00	+0700	<renny_setyo>	2020 Makser langka, 2021 Oksigen langka, 2022 Minyak goreng langka 🤔			
21									
22									
23									
24									

Demikian laporan dan tugas Pertemuan Ke-Empat yang dapat saya rangkum dan saya kerjakan, saya dapat mempraktekan penggunaan scraping dan crawling serta mampu memahami perbedaan keduanya dimana keduanya memiliki kemiripan namun pada outputnya berdasarkan output csv yang dihasilkan untuk crawling lebih spesifik membentuk tabel per masing masing baris kata atau kalimat atau lebih terstruktur mungkin lebih guna untuk pengelolaan database, namun pada scraping lebih fleksibel pada hasilnya dan data yg dikelola relatif lebih kecil meskipun bisa menggunakan tweepy dalam crawling dan scraping data twitter pada praktikum ini kita menggunakan twint.

=====TerimaKasih=====