

LOGO
UNIV/POLTEK



DIGITAL TALENT SCHOLARSHIP 2019



digitalent.kominfo.go.id

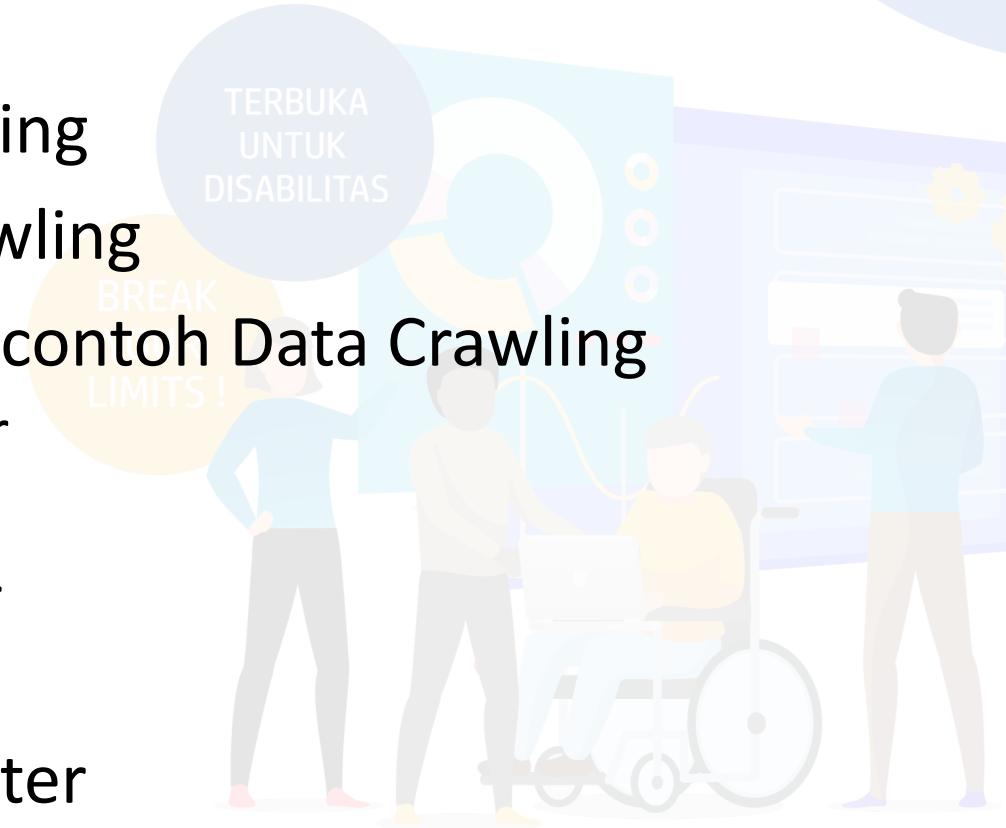
Sesi 13

Data Crawling

Big Data Analytics

Daftar Isi

- Pengenalan Data Crawling
- Konsep dasar Data Crawling
- Web Crawling, sebagai contoh Data Crawling
 - Cara Kerja Web Crawler
 - Strategi Web Crawler
 - Arsitektur Web Crawler
 - Kebijakan Web Crawler
- Data Crawling dari Twitter
 - Konfigurasi Twitter untuk Data Crawling



Pengenalan *Data Crawling* (1)

- Big Data adalah terminologi pada komputer yang merujuk kepada pengolahan data yang memiliki ukuran yang besar.
- Salah satu sumber yang dapat menyebabkan ukuran data yang besar adalah *data crawling*.
- *Data crawling* adalah teknik untuk mencari informasi dari berbagai sumber data. *Data crawling* akan melacak informasi ke setiap level sumber data yang dapat diakses.

Pengenalan *Data Crawling* (2)

- Sumber data crawling yang paling umum adalah “website”.
 - *Data crawling* yang bersumber dari website biasa disebut *web crawling*.
- *Data crawling* sangat berguna bagi perusahaan dalam mencari berbagai informasi yang tersebar di berbagai website.
- Terutama untuk perusahaan *search engine* yang menggunakan *web crawling* untuk membuat index web dari sebuah mesin pencarian di internet.

Konsep Dasar *Data Crawling* (1)

- *Data crawling* dilakukan menggunakan sebuah aplikasi *crawler* dengan konfigurasi tertentu.
- Aplikasi *crawler* adalah sebuah program yang dibuat untuk menjangkau semua sumber informasi dan melakukan aksi yang sudah ditentukan seperti memeriksa “kesegaran” dari sebuah informasi atau mengambil data dari sebuah sumber informasi.

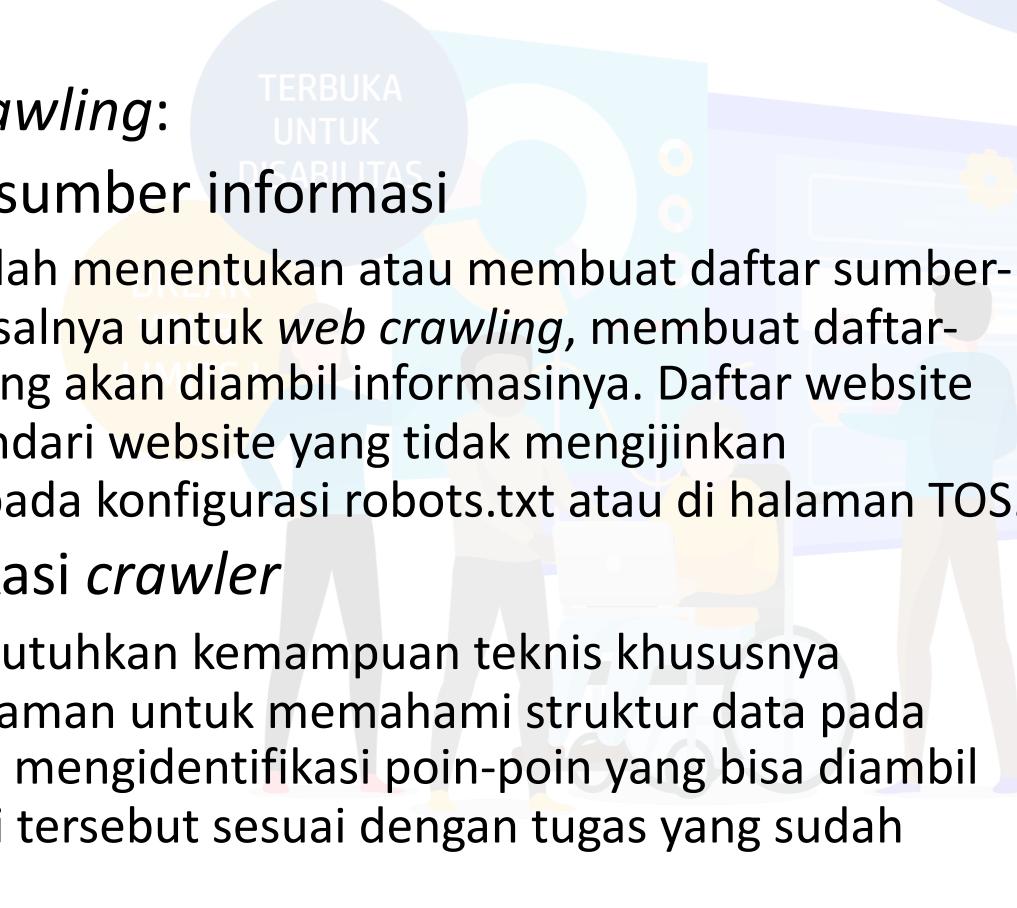
Konsep Dasar *Data Crawling* (2)

- Langkah-langkah dasar dari sebuah proses *data crawling* adalah seperti terlihat pada gambar berikut:



Konsep Dasar *Data Crawling* (3)

- Langkah-langkah *Data Crawling*:
 - Menentukan sumber-sumber informasi
 - Langkah pertama adalah menentukan atau membuat daftar sumber-sumber informasi. Misalnya untuk *web crawling*, membuat daftar-daftar URL website yang akan diambil informasinya. Daftar website harus kredibel dan hindari website yang tidak mengijinkan *automated crawling* pada konfigurasi robots.txt atau di halaman TOS.
 - Mengkonfigurasi aplikasi *crawler*
 - Langkah kedua membutuhkan kemampuan teknis khususnya kemampuan pemrograman untuk memahami struktur data pada sumber informasi dan mengidentifikasi poin-poin yang bisa diambil dari sumber informasi tersebut sesuai dengan tugas yang sudah ditentukan.



TERBUKA
UNTUK
DISABILITAS

Konsep Dasar *Data Crawling* (4)

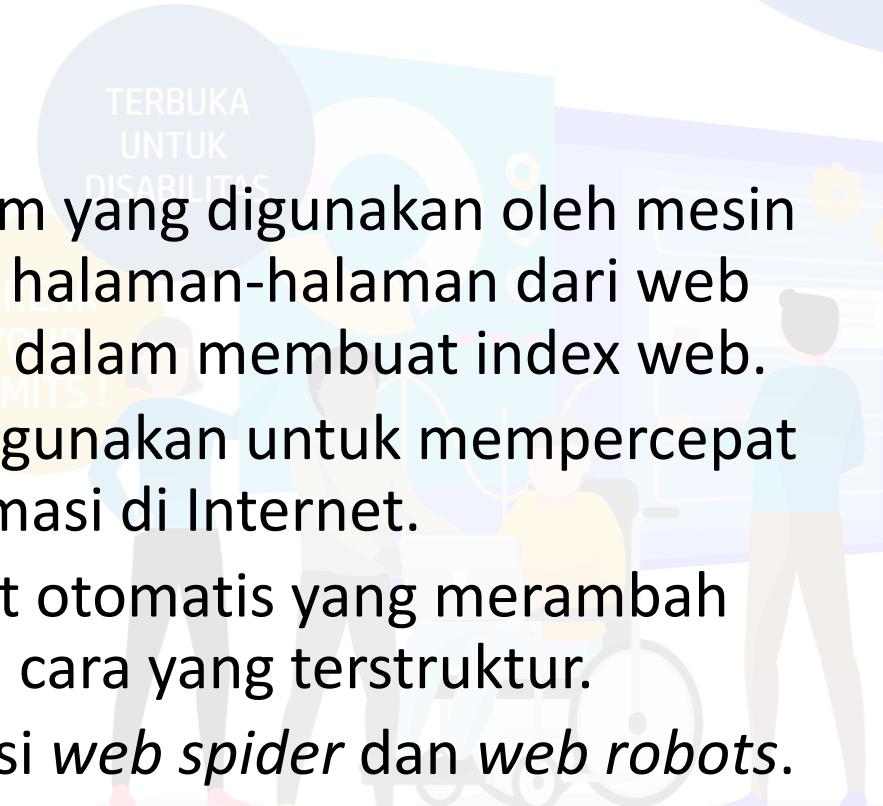
- Melakukan *cleansing* dan menghilangkan duplikasi data
 - Data awal hasil *crawler* umumnya penuh dengan data-data anomali dan mengandung duplikasi informasi. Kondisi ini dapat mempengaruhi akurasi dari proses dan analisa data. Karena itu, langkah ini menjadi penting untuk membersihkan data dari data-data anomali serta data yang terduplicasi.
- Restrukturisasi data
 - Data yang didapat dari hasil *cleansing* dan penghilangan duplikasi, perlu diubah struktur-nya ke dalam skema yang dipahami oleh komputer. Dengan data yang terstruktur, akan mempermudah pemrosesan dan analisa lebih lanjut.

Web Crawlers (1)

- Untuk memperdalam pemahaman tentang *Data Crawling*, selanjutnya akan dibahas salah satu contoh yaitu *Web crawling*.
- *Web crawling* secara spesifik mencari suatu informasi dari sekumpulan website yang ada di Internet.
- Aplikasi yang digunakan untuk melakukan *web crawling* adalah *web crawlers*.

Web Crawlers (2)

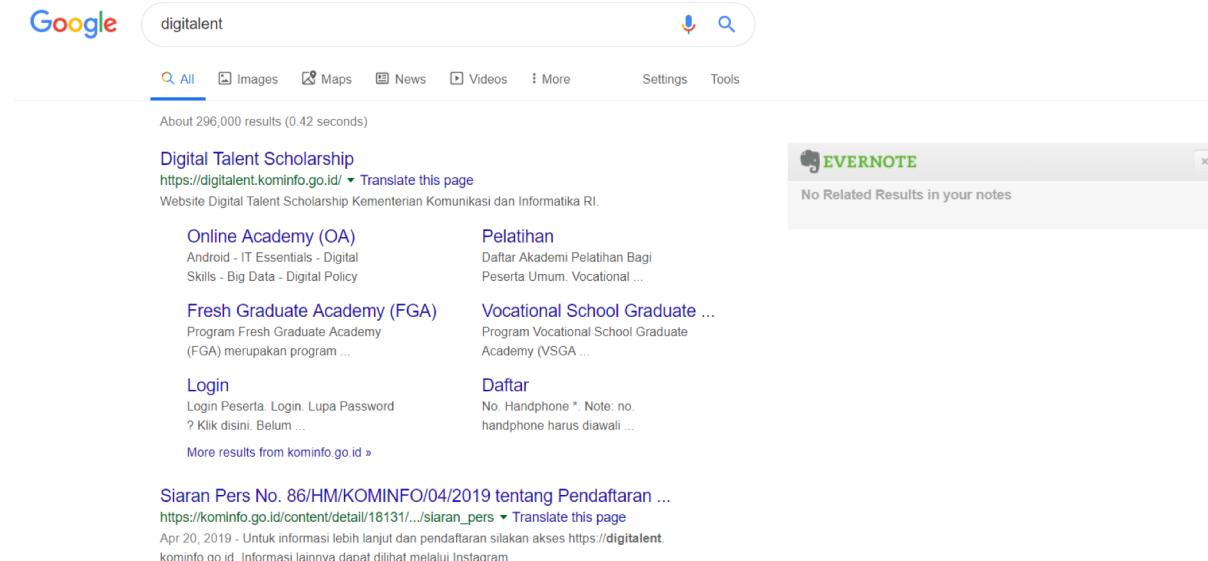
- *Web crawler* adalah
 - sebuah proses atau program yang digunakan oleh mesin pencari untuk mengunduh halaman-halaman dari web untuk diproses lebih lanjut dalam membuat index web.
 - Index web kemudian digunakan untuk mempercepat proses pencarian informasi di Internet.
 - sebuah program atau script otomatis yang merambah website di internet dengan cara yang terstruktur.
 - juga dikenal sebagai aplikasi *web spider* dan *web robots*.
 - Istilah lain yang jarang digunakan: *ants*, *bots* dan *worms*.



TERBUKA
UNTUK
DISABILITAS

Web Crawlers (3)

- *Web crawlers* sangat berguna karena:
 - Internet memiliki segudang website yang mengandung berbagai jenis informasi.
 - Mencari informasi yang relevan dari sekumpulan website di Internet membutuhkan mekanisme yang efisien.



Google digitalent

All Images Maps News Videos More Settings Tools

About 296,000 results (0.42 seconds)

Digital Talent Scholarship
<https://digitalent.kominfo.go.id/> • Translate this page

Website Digital Talent Scholarship Kementerian Komunikasi dan Informatika RI.

Online Academy (OA)
Android - IT Essentials - Digital Skills - Big Data - Digital Policy

Fresh Graduate Academy (FGA)
Program Fresh Graduate Academy (FGA) merupakan program ...

Login
Login Peserta. Login. Lupa Password ? Klik disini. Belum ...

Pelatihan
Daftar Akademi Pelatihan Bagi Peserta Umum. Vocational ...

Vocational School Graduate ...
Program Vocational School Graduate Academy (VSGA) ...

Daftar
No. Handphone *. Note: no. handphone harus diawali ...

[More results from kominfo.go.id >](#)

Siaran Pers No. 86/HM/KOMINFO/04/2019 tentang Pendaftaran ...
https://kominfo.go.id/content/detail/18131/.../siaran_pers • Translate this page

Apr 20, 2019 - Untuk informasi lebih lanjut dan pendaftaran silakan akses <https://digitalent.kominfo.go.id>. Informasi lainnya dapat dilihat melalui Instagram ...

Cara Kerja *Web Crawler* (1)

- *Web crawler* mengawali proses dari daftar URL website yang harus dikunjungi yang biasa disebut *seeds*.
 - Saat *web crawler* mengunjungi setiap website di dalam daftar tersebut, ia akan mengidentifikasi sebuah *hyperlinks* yang ada di dalam website tersebut dan menambahkannya ke dalam daftar URL yang telah dikunjungi, yang disebut *crawl frontier*.
- URL dari daftar *crawl frontier* kemudian menjadi *seeds* untuk iterasi kunjungan berikutnya berdasarkan sekumpulan konfigurasi yang sudah ditetapkan.

Cara Kerja *Web Crawling* (2)

- *Web crawlers* bekerja dengan algorima sebagai berikut:

Initialize queue (Q) with initial set of known URL's.

Until Q empty or page or time limit exhausted:

 Pop URL, L , from front of Q .

 If L is not an HTML page (.gif, .jpeg, .ps, .pdf, .ppt...)

 exit loop.

 If already visited L , continue loop (get next url).

 Download page, P , for L .

 If cannot download P (e.g. 404 error, robot excluded)

 exit loop,

 else.

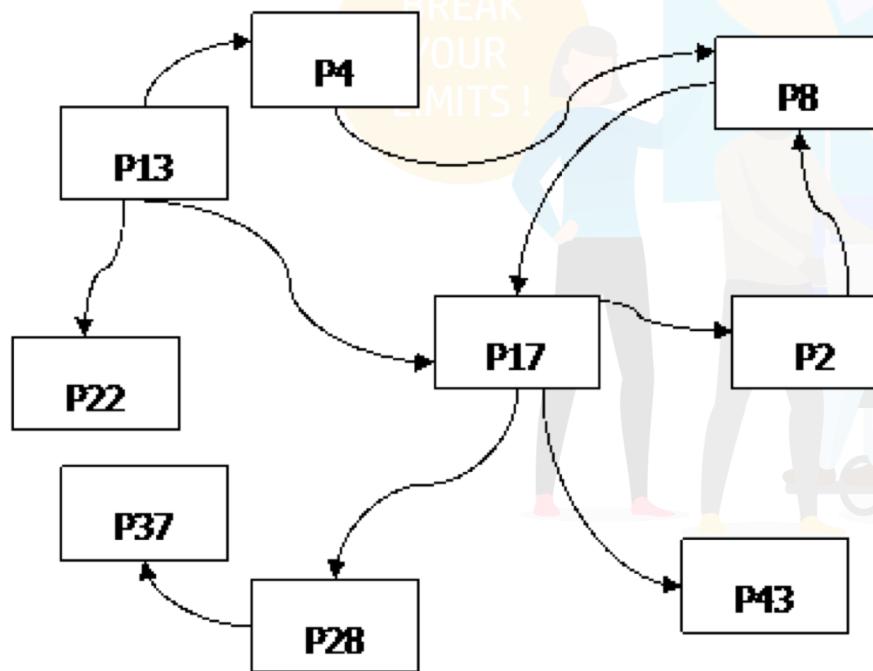
 Index P (e.g. add to inverted index or store cached copy).

 Parse P to obtain list of new links N .

 Append N to the end of Q .

Cara Kerja Web Crawling (3)

- Hasil *web crawler* biasanya berupa graph index seperti pada gambar berikut:

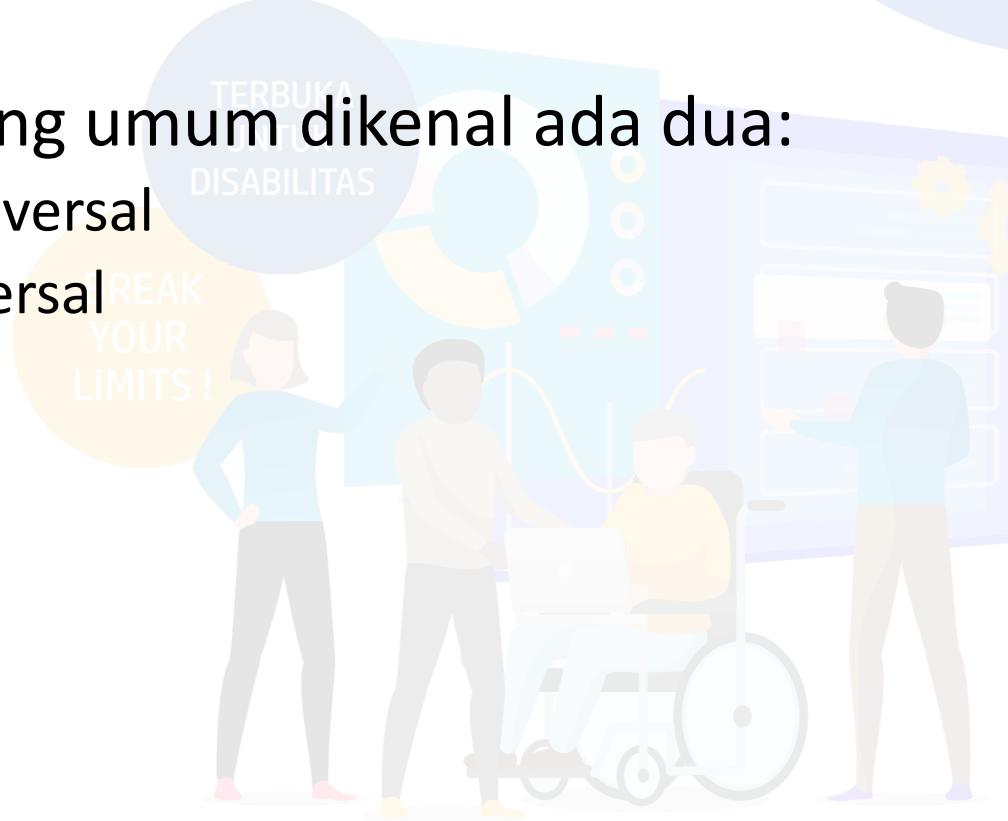


Strategi Web Crawler (1)

- Strategi *web crawler* juga bisa disebut sebagai algoritma aplikasi *web crawler*.
- Strategi *web crawler* memiliki asumsi:
 - Web adalah sebuah *directed graph* yang sangat besar, dengan dokumen sebagai *vertices* dan hyperlinks sebagai *edges*.
 - Analisa *directed graph* lebih lanjut menggunakan algoritma *graph traversal*.
 - Dalam algoritma *graph traversal*, *nodes* atau *vertices* digambarkan sebagai kotak dan *directed edges* digambarkan sebagai anak panah.

Strategi Web Crawler (2)

- Strategi *web crawler* yang umum dikenal ada dua:
 - Breadth-First Search Traversal
 - Depth-First Search Traversal

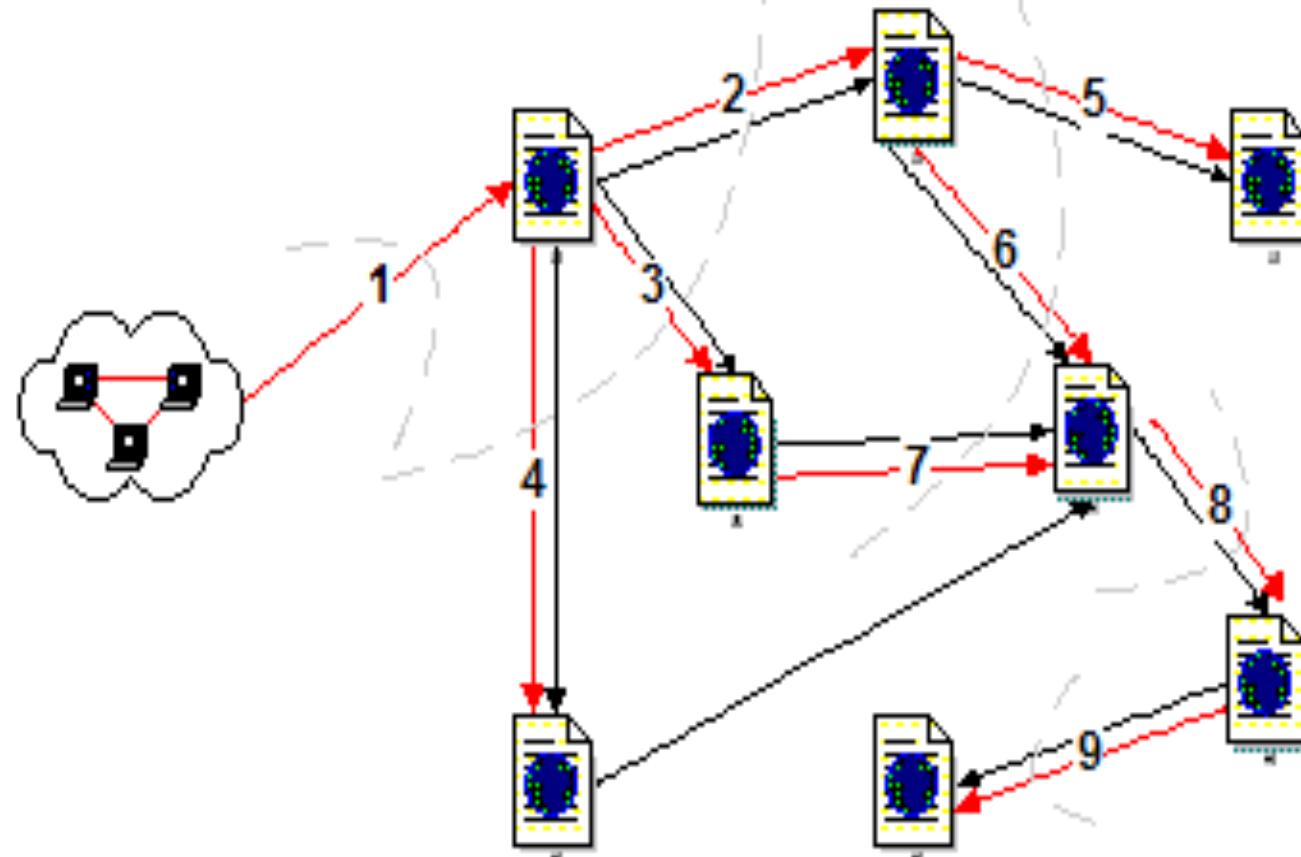


Strategi Web Crawler (3)

- Breadth-First Traversal
 - Dengan sembarang *graph* dan sekumpulan *seeds*, *graph* akan bisa dilalui dengan algoritma:
 1. Masukkan daftar URL di dalam *seeds* ke dalam antrean;
 2. Persiapkan daftar *nodes* yang telah dikunjungi (daftar ini awalnya adalah kosong);
 3. Selama antrean masih berisi data URL:
 - a. Hilangkan *node* pertama dari antrean;
 - b. Tambahkan *node* tersebut ke dalam daftar *node* yang telah dikunjungi.
 - c. Untuk setiap *edge* yang dimulai dari *node* tersebut:
 - i. Jika *node* pada akhir *edge* telah terdapat di dalam daftar *node* yang telah dikunjungi atau sudah ada di dalam antrean, jangan lakukan apa-apa terhadap *edge* tersebut;
 - ii. Sebaliknya, tambahkan *node* pada akhir *edge* ke bagian akhir dari antrean.

Strategi Web Crawler (4)

- Ilustrasi strategi breadth-first crawlers:

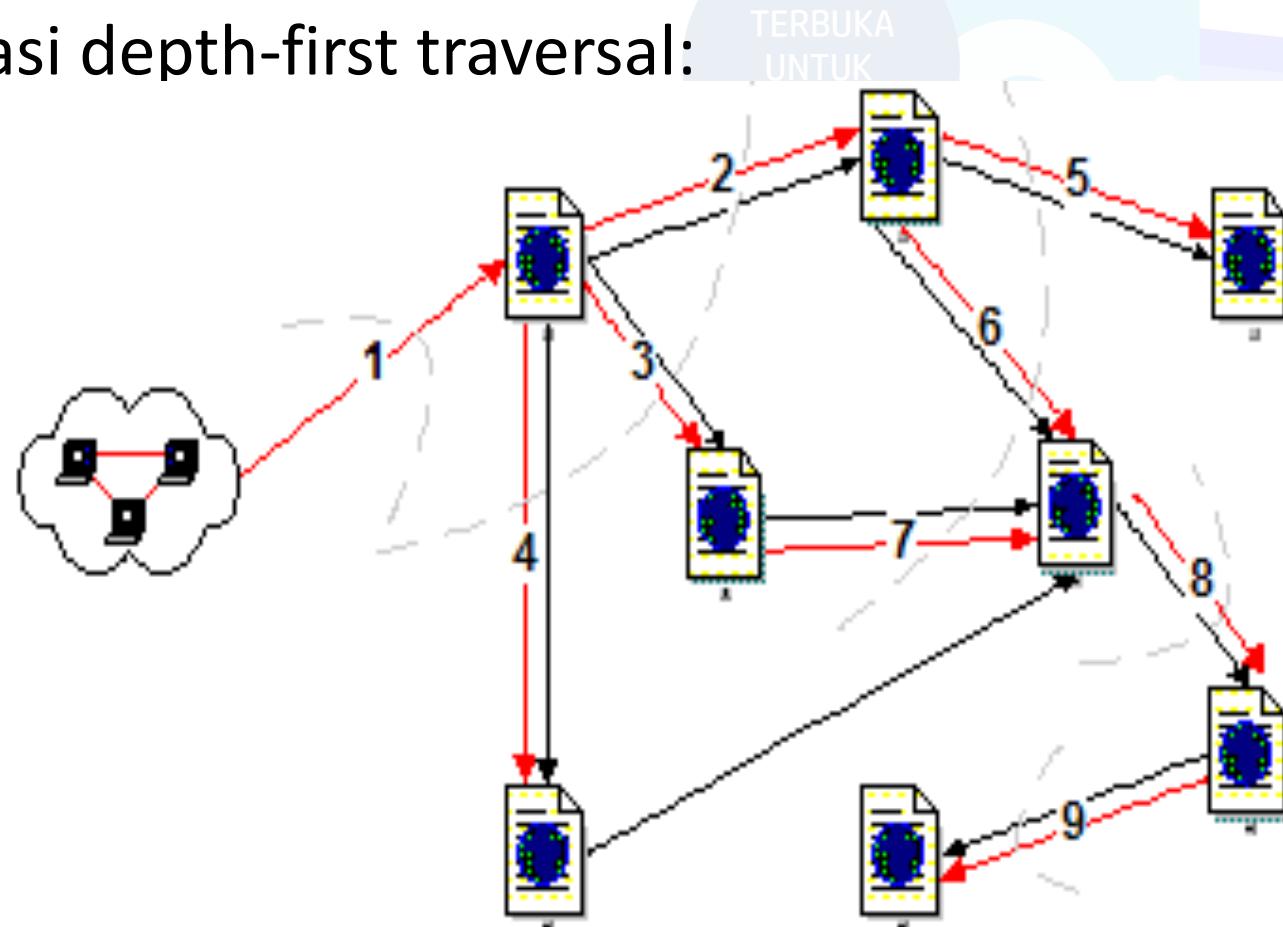


Strategi Web Crawler (5)

- Algoritma depth-first search (DFS):
 - Ambil link pertama yang belum dikunjungi dari halaman awal.
 - Kunjungi link tersebut dan ambil link pertama yang belum dikunjungi.
 - Ulangi langkah-langkah di atas hingga tidak ada lagi link yang belum dikunjungi.
 - Kunjungi link yang belum dikunjungi berikutnya yang terdapat di dalam level sebelumnya dan ulangi langkah kedua.

Strategi Web Crawler (6)

- Ilustrasi depth-first traversal:

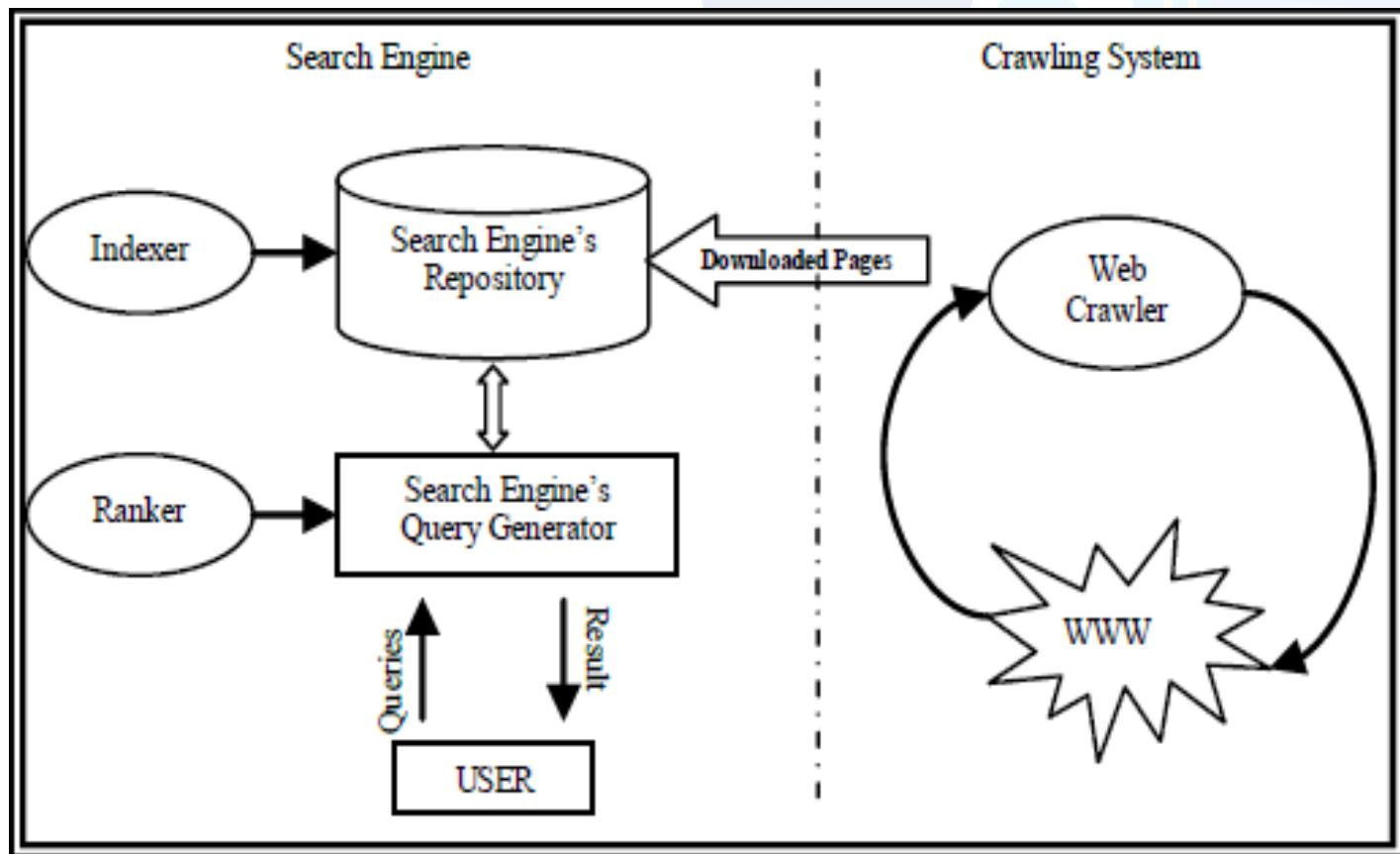


Strategi Web Crawler (7)

- Depth-First vs Breadth-First:
 - Depth-First bekerja pada satu *vertices* hingga mencapai sebuah *edge*
 - Tidak cukup bagus jika *edge* berada di *vertices* yang berbeda
 - Tidak lengkap atau optimal
 - Menggunakan ruang yang lebih sedikit di banding Breadth-First
 - Lebih sedikit *node* yang telah dikunjungi yang perlu diawasi
 - Ukuran lebih kecil
 - Breadth-First memeriksa semua alternative dengan lebih teliti
 - Lengkap dan optimal
 - Membutuhkan memory yang cukup besar

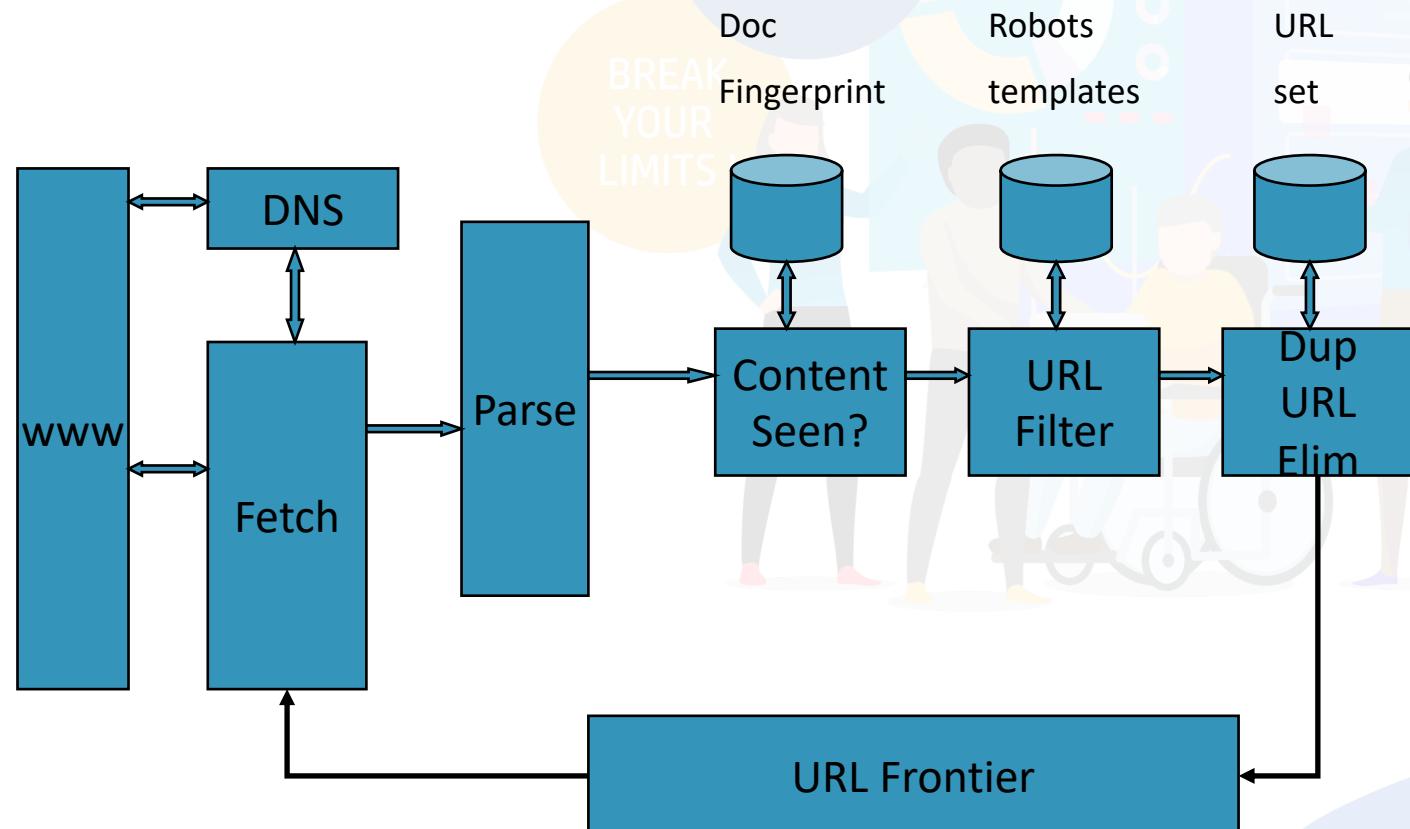
Arsitektur Web Crawler (1)

- Ilustrasi arsitektur *search engine*:



Arsitektur Web Crawler (2)

- Ilustrasi arsitektur web crawler:



Arsitektur Web Crawler (3)

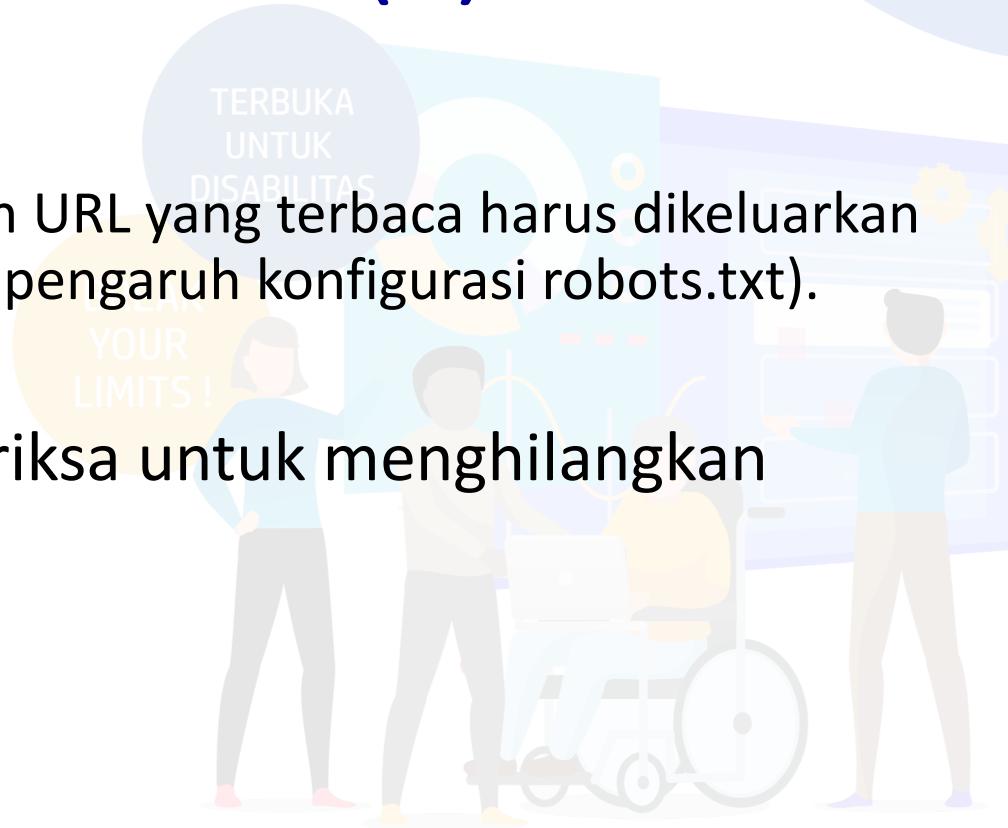
- **URL Frontier:** mengandung URL yang belum dikunjungi. Awalnya, sebuah *seeds* disimpan pada URL Frontier dan sebuah *web crawler* memulai proses dari URL yang berada di dalam URL frontier.
- **DNS:** singkatan dari Domain Name Service yang memetakan sebuah domain name dengan alamat IP.
- **Fetch:** umumnya menggunakan protokol HTTP untuk mengakses URL.
- **Parse:** halaman web di urai dan mengambil informasi text dan link.
- **Content Seen?:** menguji apakah sebuah halaman web dengan isi yang sama telah diakses melalui URL yang lain.

Arsitektur Web Crawler (4)

- **URL Filter:**

- Untuk memeriksa apakah URL yang terbaca harus dikeluarkan dari daftar URL Frontier (pengaruh konfigurasi robots.txt).
- URL harus dinormalisasi.

- **Dup URL Elim:** URL diperiksa untuk menghilangkan duplikasi.



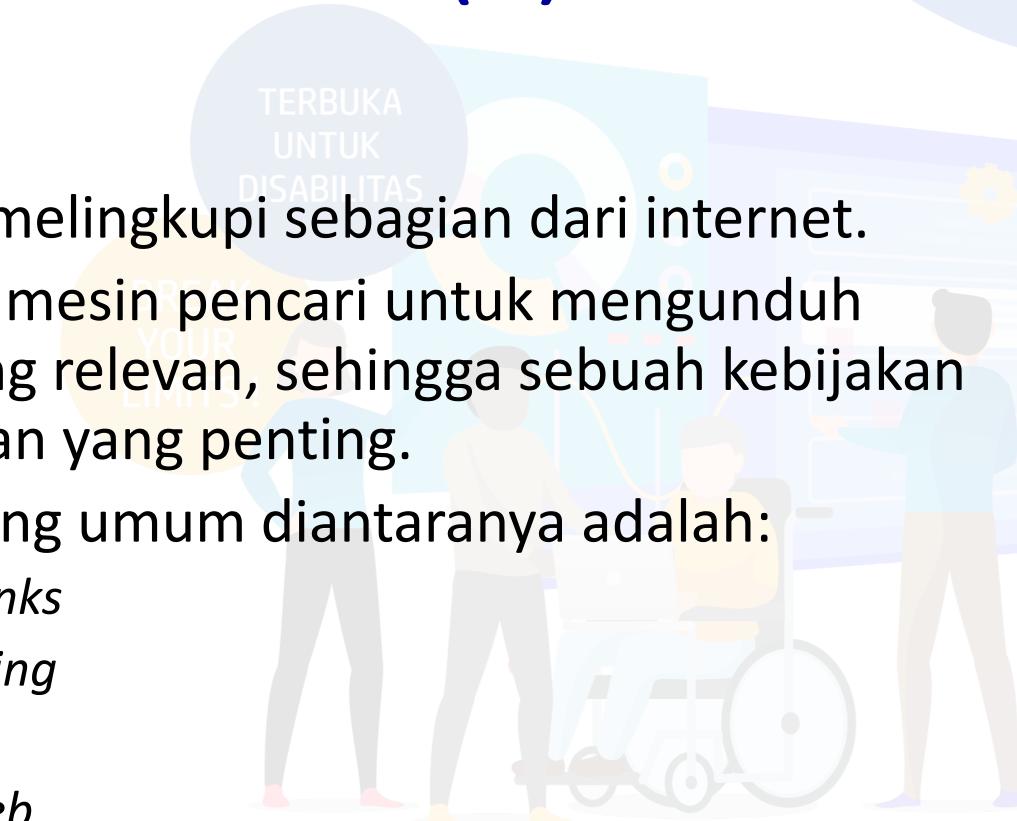
Kebijakan *Web Crawler* (1)

- Kebijakan *web crawler* terdiri dari:
 - Kebijakan *selection* yang menentukan halaman mana yang bisa diakses.
 - Kebijakan *re-visit* yang menentukan kapan sebuah halaman dikunjungi kembali untuk mendeteksi adanya pembaharuan isi.
 - Kebijakan *politeness* untuk menentukan bagaimana menghindari proses yang memberi kelebihan beban bagi website.
 - Kebijakan *parallelization* mengatur cara mengkoordinasi *web crawler* yang terdistribusi.

Kebijakan Web Crawlers (2)

- Kebijakan *Selection*:

- Mesin pencari hanya melingkupi sebagian dari internet.
- Hal ini membutuhkan mesin pencari untuk mengunduh halaman-halaman yang relevan, sehingga sebuah kebijakan *selection* memiliki peran yang penting.
- Kebijakan *selection* yang umum diantaranya adalah:
 - *Restricting followed links*
 - *Path-ascending crawling*
 - *Focused crawling*
 - *Crawling the Deep Web*

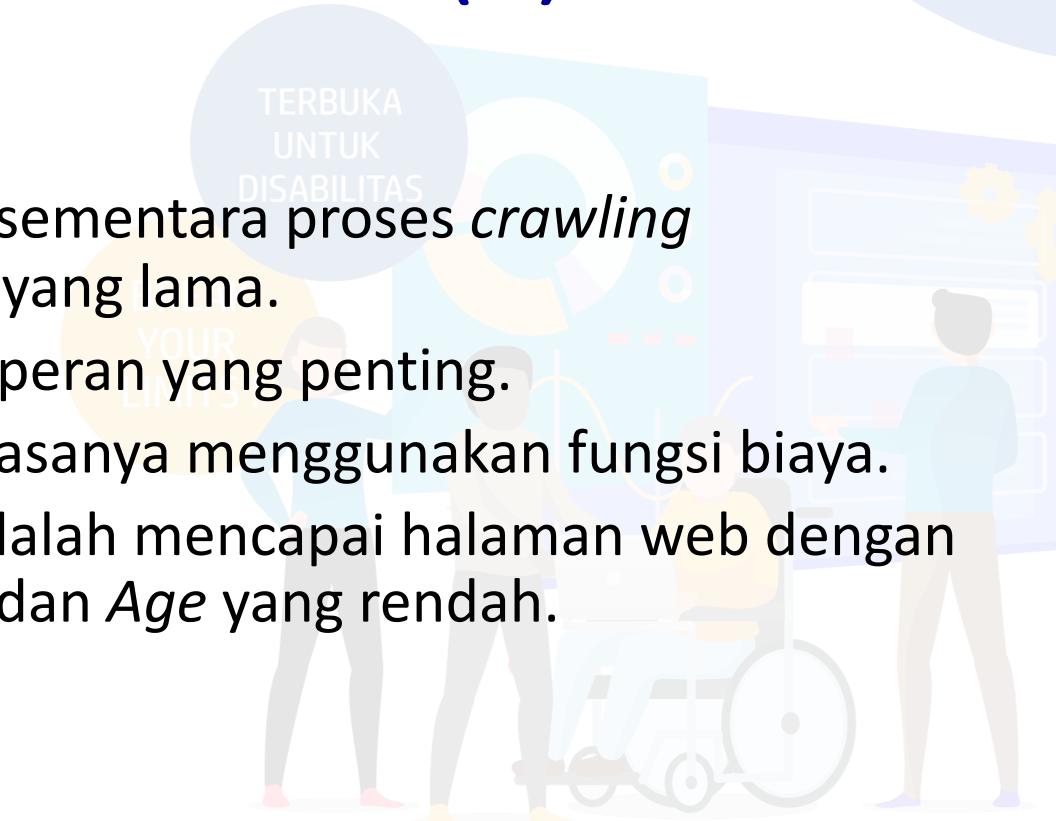


TERBUKA
UNTUK
DISABILITAS

Kebijakan Web Crawlers (3)

- Kebijakan *Re-visit*:

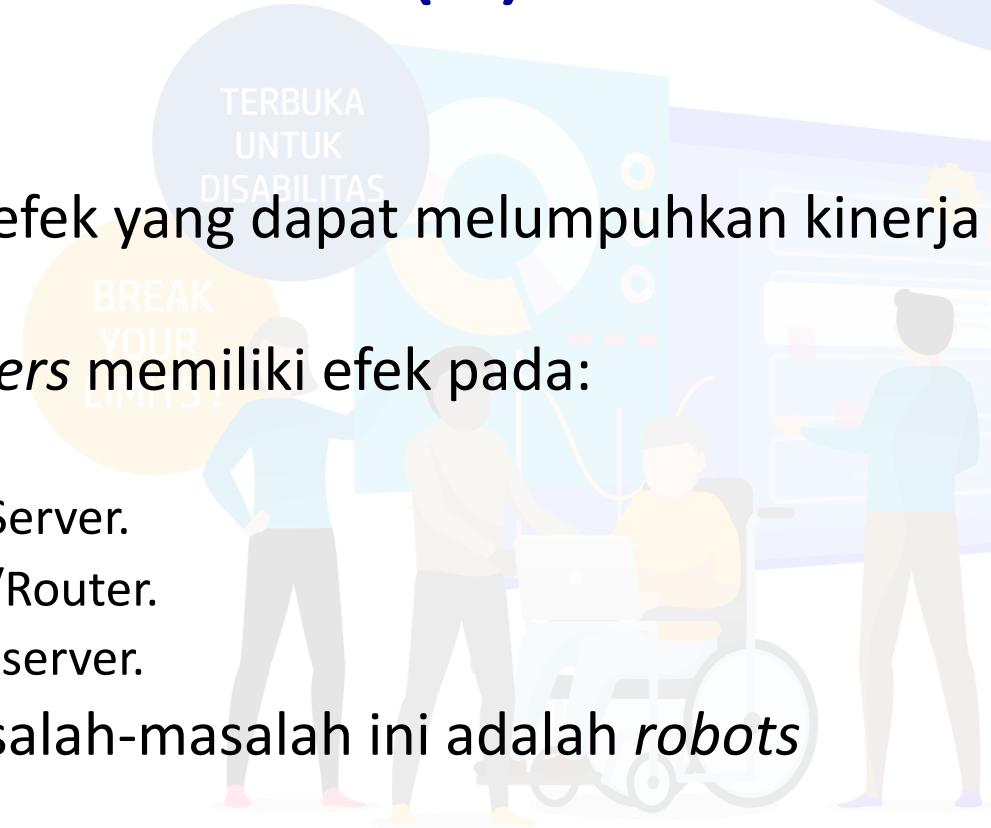
- Web bersifat dinamis sementara proses *crawling* membutuhkan waktu yang lama.
- Faktor biaya memiliki peran yang penting.
- *Freshness* dan *Age*- biasanya menggunakan fungsi biaya.
- Tujuan dari *crawler* adalah mencapai halaman web dengan *Freshness* yang tinggi dan *Age* yang rendah.
- Dua kebijakan *re-visit*:
 - *Uniform Policy*
 - *Proportional Policy*



Kebijakan *Web Crawlers* (4)

- Kebijakan *Politeness*:

- *Crawlers* bisa memiliki efek yang dapat melumpuhkan kinerja sebuah website.
- Penggunaan *web crawlers* memiliki efek pada:
 - Sumber daya jaringan.
 - Kelebihan beban pada Server.
 - Kerusakan pada Server/Router.
 - Gangguan jaringan dan server.
- Solusi terpisah dari masalah-masalah ini adalah *robots exclusion protocol*.



Kebijakan Web Crawlers (5)

- Untuk mendukung kebijakan *Politeness*, penerapan teknisnya adalah dengan menggunakan *robot exclusion*.
 - Website dan halaman bisa menspesifikasi bahwa robot tidak boleh mengunjungi atau melakukan index pada area tertentu.
 - Dua komponen dari *robot exclusion*:
 - **Robot Exclusion Protocol** (`robots.txt`): website memberi spesifikasi daftar direktori yang tidak boleh diakses.
 - **Robots META Tag**: Setiap halaman/dokumen memberi *meta tag* yang melarang robot untuk melakukan index atau mendeteksi link yang ada pada halaman tersebut.

Kebijakan Web Crawlers (6)

- Untuk mendukung kebijakan *Politeness*, penerapan teknisnya adalah dengan menggunakan *robot exclusion*.
 - Website dan halaman bisa menspesifikasi bahwa robot tidak boleh mengunjungi atau melakukan index pada area tertentu.
 - Dua komponen dari *robot exclusion*:
 - **Robot Exclusion Protocol** (`robots.txt`): website memberi spesifikasi daftar direktori yang tidak boleh diakses.
 - **Robots META Tag**: Setiap halaman/dokumen memberi *meta tag* yang melarang robot untuk melakukan index atau mendeteksi link yang ada pada halaman tersebut.

Kebijakan Web Crawlers (7)

- Admin website menaruh sebuah file “robots.txt” pada direktori utama (root) dari direktori web. Contohnya:
 - <http://www.ebay.com/robots.txt>
 - <http://www.Wikipedia.com/robots.txt>
- File “robots.txt” berisi daftar direktori yang tidak boleh diakses oleh robot tertentu (user-agent).
 - Contoh kode “robots.txt” yang melarang semua robot untuk keseluruhan website:

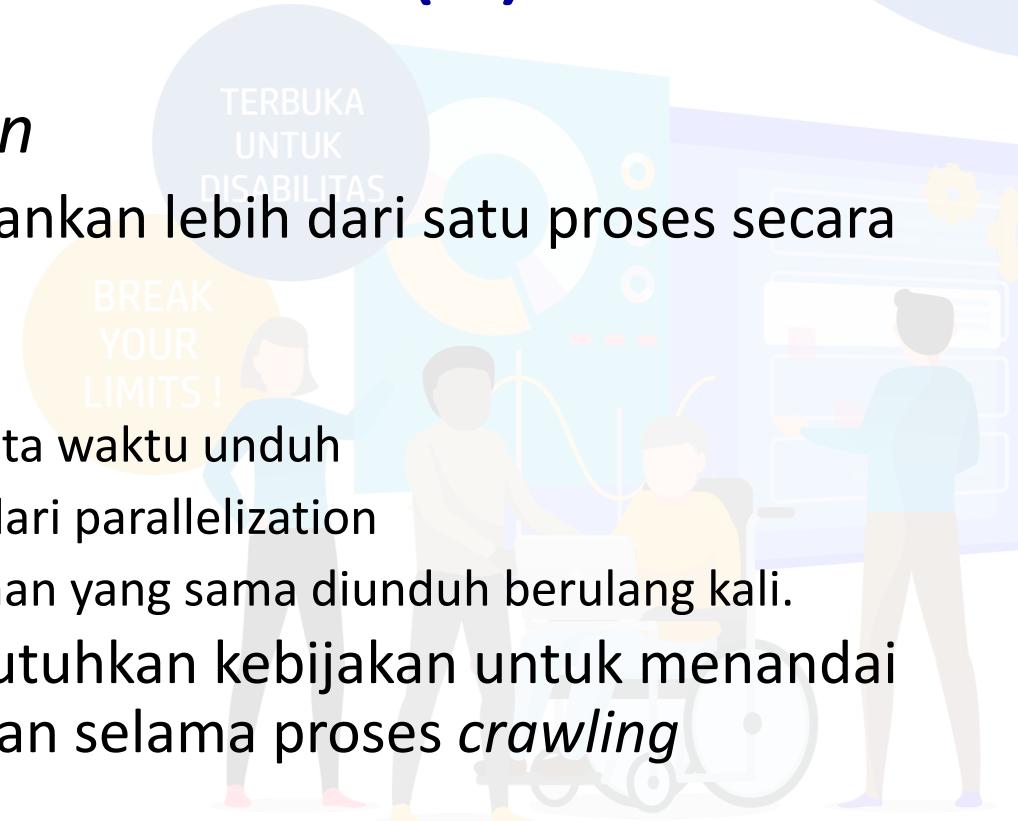
User-agent: *

Disallow: /

Kebijakan Web Crawlers (8)

- Kebijakan *Parallelization*

- Aplikasi *crawler* menjalankan lebih dari satu proses secara parallel.
- Tujuannya:
 - Memaksimalkan rata-rata waktu unduh
 - Mengurangi *overhead* dari parallelization
 - Untuk mencegah halaman yang sama diunduh berulang kali.
- Sistem *crawling* membutuhkan kebijakan untuk menandai URL baru yang ditemukan selama proses *crawling* berlangsung.



Data Crawling dari Twitter (1)

- Contoh berikutnya mengenai *Data Crawling* yang sekarang paling banyak digunakan dalam algoritma *Machine Learning* adalah *Data Crawling* dari Twitter.
- Twitter adalah aplikasi jaringan social yang popular dimana penggunanya dapat membagi pesan-pesan pendek yang disebut *tweets*.
- Penggunaan Twitter sangat beragam dari sekedar saling berbagi ide, link, gambar dan video, komentar jurnalis pada event *live* hingga promosi produk.
- Twitter dapat mencapai 500 juta tweets dalam satu hari, yang merupakan data yang sangat besar untuk diolah.

Data Crawling dari Twitter (2)

- Cara mengolah data dari twitter tidak berbeda dari tahapan pada *data crawler*.
- Tahap pertama, seorang data analis harus memiliki beberapa *key* yang memungkinkan koneksi ke API yang disediakan oleh Twitter.
- Data analis menggunakan *key* tersebut pada bahasa pemrograman yang dipilih untuk menganalisa data-data twitter.

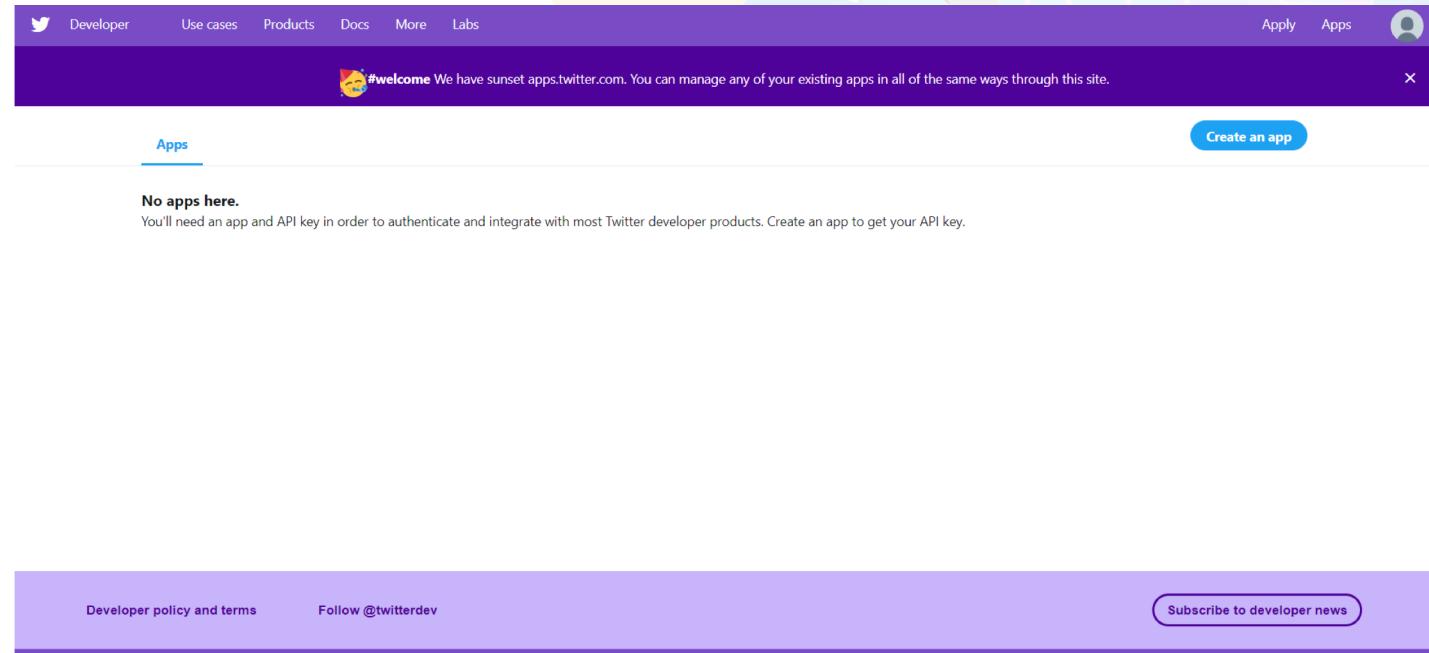
Data Crawling dari Twitter (3)

- Untuk mendapatkan *access key* ke twitter, data analis harus membuat aplikasi yang berinteraksi dengan Twitter API.
- Langkah pertama adalah mendaftarkan aplikasi ke Twitter. Linknya adalah:
<https://developer.twitter.com/en/apps>
- Login menggunakan akun twitter atau buat akun twitter jika belum memilikinya.

Python – Twitter Step-by-Step (1)

- Registrasi aplikasi di twitter.

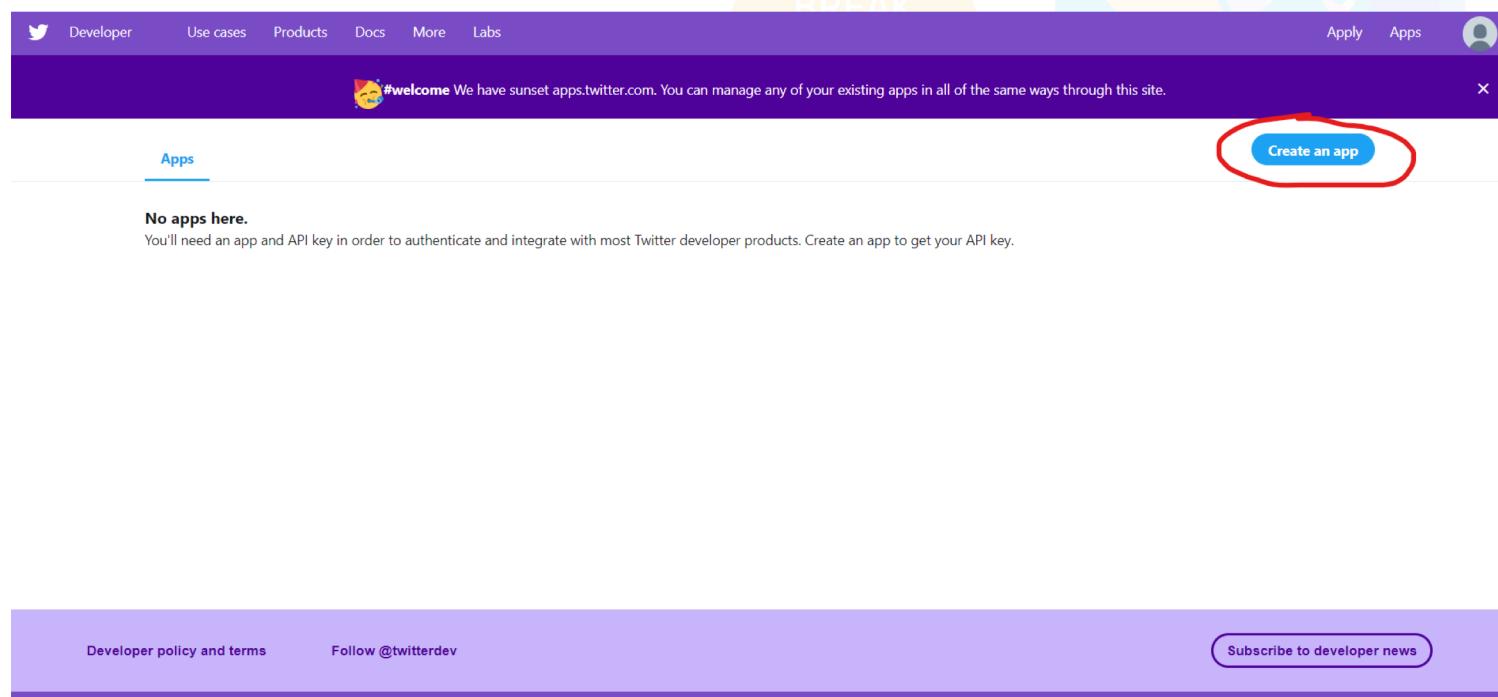
- Di browser, masukkan url:
<https://developer.twitter.com/en/apps>
- Tampilannya:



The screenshot shows the Twitter Developer website's 'Create an app' page. At the top, there is a purple header with navigation links: 'Developer', 'Use cases', 'Products', 'Docs', 'More', 'Labs', 'Apply', 'Apps', and a user icon. A welcome message from the Twitter logo says: '#welcome We have sunset apps.twitter.com. You can manage any of your existing apps in all of the same ways through this site.' Below this, there is a section titled 'No apps here.' with a sub-instruction: 'You'll need an app and API key in order to authenticate and integrate with most Twitter developer products. Create an app to get your API key.' At the bottom of the page, there are links for 'Developer policy and terms', 'Follow @twitterdev', and 'Subscribe to developer news'.

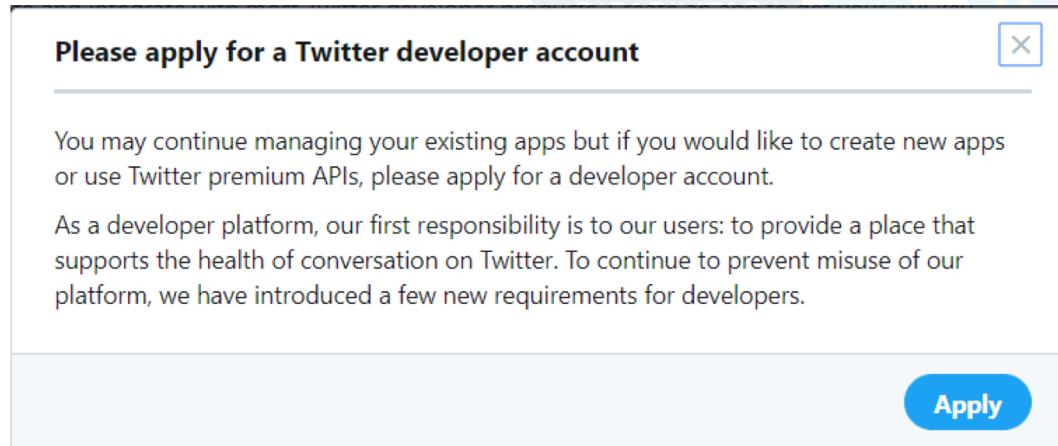
Python – Twitter Step-by-Step (2)

- Klik tombol Create an app yang ada di bagian kanan atas.



Python – Twitter Step-by-Step (3)

- Saat di klik, akan muncul sebuah *window* seperti berikut:



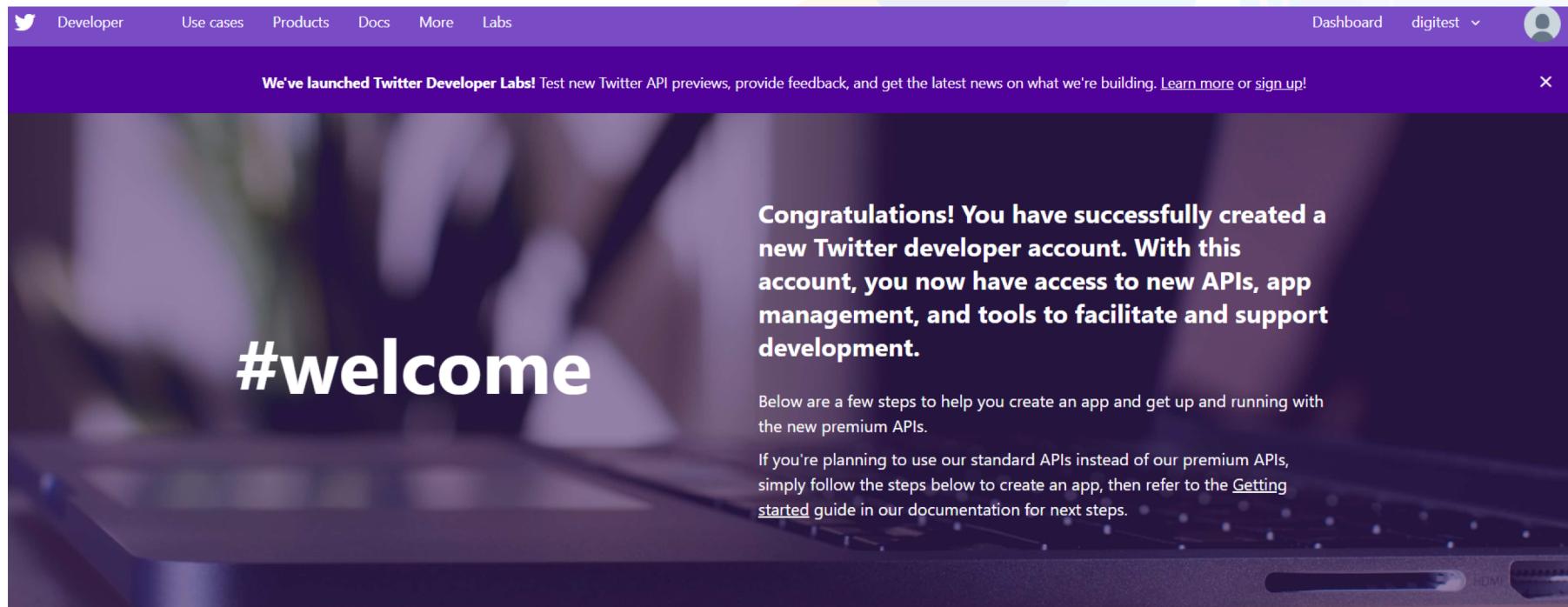
- Window* ini akan mengkonfirmasi akun twitter yang baru pertama kali ingin membuat aplikasi.
- Klik tombol Apply untuk mengkonfirmasi pembuatan aplikasi yang pertama kali.

Python – Twitter Step-by-Step (4)

- Ikuti instruksi pada langkah-langkah berikutnya untuk mendapatkan akses developer di twitter.
- Bila pengisian survey untuk mendapatkan akses developer berhasil, Twitter akan mengirimkan email konfirmasi ke email yang terdaftar di twitter.
- Klik link yang terdapat di dalam email tersebut untuk menyelesaikan proses pendaftaran sebagai akun developer di twitter.

Python – Twitter Step-by-Step (5)

- Bila berhasil, akan tampil halaman sebagai berikut:



The screenshot shows the Twitter Developer Labs landing page. At the top, there is a purple header bar with links for Developer, Use cases, Products, Docs, More, and Labs. On the right side of the header are links for Dashboard and a user profile icon. A message banner at the top states: "We've launched Twitter Developer Labs! Test new Twitter API previews, provide feedback, and get the latest news on what we're building. [Learn more or sign up!](#)" Below the banner, the main content area features a large image of hands typing on a keyboard. Overlaid on this image is the text "#welcome" in large white letters. To the right of the image, a congratulatory message reads: "Congratulations! You have successfully created a new Twitter developer account. With this account, you now have access to new APIs, app management, and tools to facilitate and support development." Below this message, smaller text provides instructions: "Below are a few steps to help you create an app and get up and running with the new premium APIs. If you're planning to use our standard APIs instead of our premium APIs, simply follow the steps below to create an app, then refer to the [Getting started](#) guide in our documentation for next steps." At the bottom left, there is a "Helpful tools" section with a "Get started" button.

Python – Twitter Step-by-Step (6)

- *Scroll down* pada halaman tersebut untuk menemukan link pembuatan aplikasi.
- Klik link pembuatan aplikasi untuk mendapat twitter key yang akan digunakan dalam pemrograman twitter.

Get started

 [Create an app](#)

To use an API, we require you create an app as part of our OAuth authorization scheme. Visit the [Apps](#) page of this developer portal to create one. Then, return to this page to complete the next step.

Python – Twitter Step-by-Step (7)

- Klik link *Create an app* untuk memulai pembuatan aplikasi di twitter.
- Isi form untuk menyelesaikan pembuatan aplikasi di twitter.

Apps > [Create an app](#)

 Understanding apps

What is an app? 

Why register an app? 

Which products require an API key? 

App details

The following app details will be visible to app users and are required to generate the API keys needed to authenticate Twitter developer products.

App name (required) 

Maximum characters: 32

Application description (required)
Share a description of your app. This description will be visible to users so this is a good place to tell them what your app does.

Please be detailed.

Python – Twitter Step-by-Step (8)

- Jika berhasil mengisi form pembuatan aplikasi di twitter, akan muncul halaman berikut:

Apps > Digitest2019

App details Keys and tokens Permissions

App details

Details and URLs

App icon
App icon is default, click edit to upload.

App Name
Digitest2019

Description

You are using the default icon now, change it in app editing mode. the example of text mining application

Website URL
<https://digitalent.kominfo.go.id/>

Sign in with Twitter
Disabled

Python – Twitter Step-by-Step (8)

- Klik link Keys and Tokens untuk melihat key dan token aplikasi twitter yang dapat dimanfaatkan pada pemrograman python.
- Tampilannya seperti berikut:

The screenshot shows a web interface for managing API keys and tokens. At the top, there's a navigation bar with 'Apps' and 'Digitest2019'. Below it, there are three tabs: 'App details', 'Keys and tokens' (which is currently selected), and 'Permissions'. The main content area is titled 'Keys and tokens' and contains a sub-section for 'Consumer API keys'. It displays two keys: 'ulpjy4nA5y6JqkuqqR78pTdXC' (API key) and 'Y2rrlc4H4qmYfi6XkyuTTubSheA0W7U8Szw5ZfUE4KrfzYXb8' (API secret key). There's a 'Regenerate' button next to these keys. Below this, there's another section for 'Access token & access token secret', which lists one token: '1145151860644311041-aqQN0wP3dzMrc9hxgqgPwU6HmOzDdW' (Access token) and its secret: 'bTGHSMS4WBAI59Jx7ZgTT4Zu2lahxTE6Fo1g0VG3cFTN2M' (Access token secret). A 'Read and write' button is shown under this section. At the bottom of the page are 'Revoke' and 'Regenerate' buttons.

Python – Twitter Step-by-Step (9)

- Setelah mendapatkan key dan token dari twitter, langkah selanjutnya adalah menginstall library tweepy untuk python.
- Instruksinya:
 - sudo pip install tweepy



Python – Twitter Step-by-Step (10)

- Setelah berhasil meng-*install* library tweepy, gunakan kode berikut untuk mengatur koneksi program python ke aplikasi twitter.

```
1 import tweepy
2
3 # Variables that contains the credentials to access Twitter API
4 ACCESS_TOKEN = 'your_access_token'
5 ACCESS_SECRET = 'your_access_secret'
6 CONSUMER_KEY = 'your_consumer_key'
7 CONSUMER_SECRET = 'your_consumer_secret'
8
9
10 # Setup access to API
11 def connect_to_twitter_OAuth():
12     auth = tweepy.OAuthHandler(CONSUMER_KEY, CONSUMER_SECRET)
13     auth.set_access_token(ACCESS_TOKEN, ACCESS_SECRET)
14
15     api = tweepy.API(auth)
16     return api
17
18
19 # Create API object
20 api = connect_to_twitter_OAuth()
```

Gunakan access token dan consumer key yang didapat dari aplikasi twitter sebelumnya.

Python – Twitter Step-by-Step (11)

- Gunakan kode berikut untuk mendapatkan *tweets* dari aplikasi yang sudah dibuat.
- Bila berhasil, keluaran kode python akan menampilkan semua *tweets* yang bisa dibaca oleh akun twitter milik kita sendiri.

```
1 # tweets from my stream
2 public_tweets = api.home_timeline()
3 for tweet in public_tweets:
4     print(tweet.text)
```

Python – Twitter Step-by-Step (12)

- Keluarannya seperti berikut:

TERBUKA
UNTUK
DISABILITAS

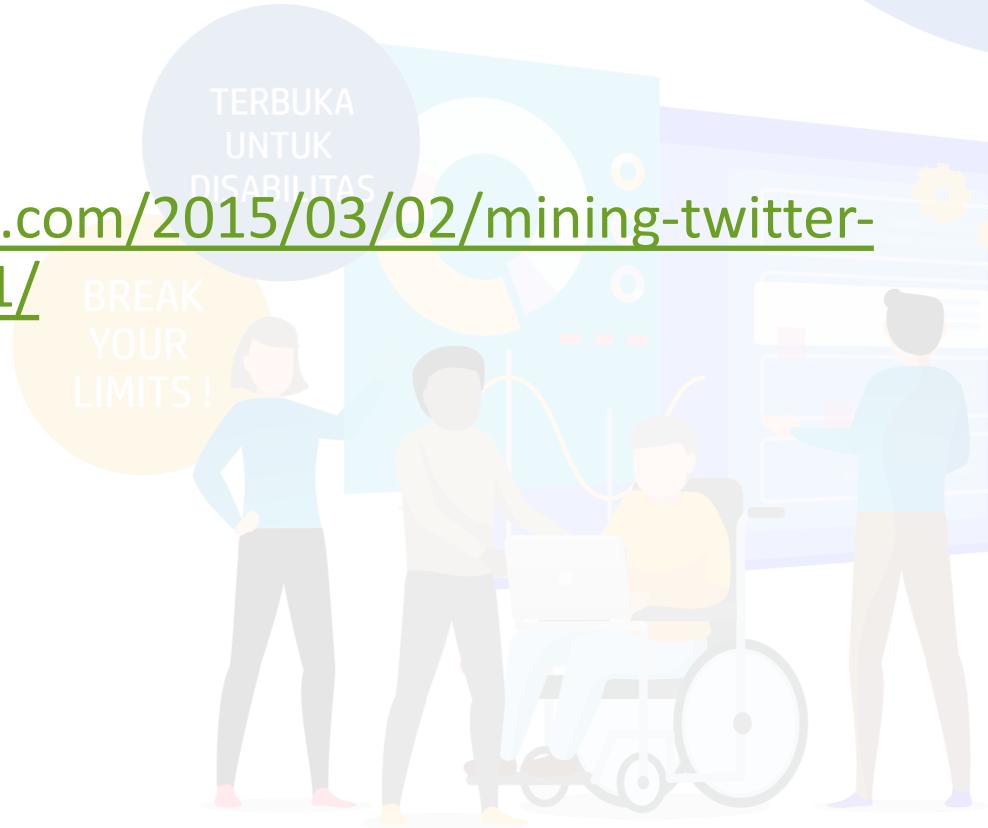
Sekjen Kominfo Paparkan Pelindungan Data Pribadi dan Beasiswa Talenta Digital <https://t.co/raGLLS9Nld>
Forum APT Sepakati 5 Strategi Baru Kebijakan Digital Asia Pasifik <https://t.co/gwi3PH7Loq>
KTT G-20, Presiden Usulkan Perlunya IDEA Hub untuk Kurangi Kesenjangan <https://t.co/9RoxIvLdxm>
12.Selamat Hari Keluarga Nasional ke-26 untuk semua Keluarga Indonesia! Jadilah keluarga yang sejahtera dan akar tu... <https://t.co/BATgJpqbla>
11.Dalam upaya meningkatkan komitmen keberhasilan program tersebut, BKKBN juga bersinergi dengan Kementerian/Lembag... <https://t.co/7zsbc7IaR8>
10.HARGANAS XXVI sekaligus menjadi ajang menggelorakan & mensosialisasikan Program Kependudukan, Keluarga Berencan... <https://t.co/hTc00i6zs>
9.Konsep ini sangat bermanfaat dalam menghadapi Era 4.0 agar anggota keluarga tetap memiliki ketahanan dan tetap be... <https://t.co/GTd8PL1jFT>
8.... termasuk menguatkan kembali Gerakan Kembali ke Meja Makan dan Delapan Fungsi Keluarga (agama, cinta kasih, per... <https://t.co/tr043E1qz1>
7.Peringatan HARGANAS XXVI merefleksikan 4 konsep pendekatan ketahanan keluarga yaitu Keluarga Berkumpul, Keluarga... <https://t.co/qPAfSq2SpE>
6.Puncak peringatan HARGANAS XXVI dimeriahkan dengan serangkaian kegiatan seru, seperti: Seminar, Pelayanan KB dan... <https://t.co/bsnmW2vSbP>
5.Puncak Peringatan HARGANAS XXVI akan berlangsung di Kota Banjarbaru, Kalimantan Selatan pada 6 Juli 2019 mendatang... <https://t.co/5inl9ILgSa>
4.Nah, #SobatKom, tahun 2019 menjadi peringatan HARGANAS ke-26. Mengambil tajuk ‘Hari Keluarga, Hari Kita Semua’ da... <https://t.co/0YYyARs6Wq>
3.Presiden Joko Widodo sering menyampaikan, bangsa Indonesia akan menjadi bangsa kuat jika keluarga-keluarga Indone... <https://t.co/wBPsWEThwl>
2.Inilah yang menjadi dasar pembentukan Keputusan Presiden Republik Indonesia No.39 Tahun 2014 tentang Hari Keluarg... <https://t.co/WMUGye2Ldg>
1.Keluarga merupakan pondasi terpenting dalam pembentukan karakter manusia, bahkan menjadi titik awal penentu kual... <https://t.co/oVk5k5B00z>
RT @KSPgoid: Twitter Moment @KSPgoid: Pujian Menteri Saudi kepada Presiden @jokowi: Indonesia Punya Dua Menteri Perempuan Terbaik di Dunia...

Keluaran bisa berbeda tergantung akun twitter masing-masing.

- Referensi latihan:

- <https://marcobonzanini.com/2015/03/02/mining-twitter-data-with-python-part-1/>

- Pertanyaan?



IKUTI KAMI



- [digitalent.kominfo](https://www.facebook.com/digitalentscholarship)
- [digitalent.kominfo](https://www.instagram.com/digitalentscholarship/)
- [DTS_kominfo](https://twitter.com/DTS_kominfo)
- [Digital Talent Scholarship 2019](https://www.telegram.org/channels/DigitalTalentScholarship2019)

Pusat Pengembangan Profesi dan Sertifikasi
Badan Penelitian dan Pengembangan SDM
Kementerian Komunikasi dan Informatika
Jl. Medan Merdeka Barat No. 9
(Gd. Belakang Lt. 4 - 5)
Jakarta Pusat, 10110

