Nama: Alya Fauziyah  
Content: Web Scraping/Crawling dan Google AdSennse

**Pengertian Web Scraping**

*Web scraping* adalah proses pengambilan data atau esktraksi dari sebuah website, lalu data tersebut umumnya disimpan dalam sebuah format tertentu.

*Web scraper* merupakan program dengan yang masuk ke halaman website, mengunduh kontennya, mengekstrak data dari konten, desain, dan berbagai kerumitannya , bergantung pada kebutuhan.

Lebih singkatnya, Scraping merupakan teknik mengumpulkan data pada sebuah website melalui proses ektraksi informasi menggunakan **Hypertext Transfer Protocol (HTTP).**

Secara umum, ada dua metode yang bisa kamu gunakan :

* **Manual***–* metode ini mengharuskan kamu menyalin data dengan melakukan *copy paste* dari sebuah laman.
* **Otomatis** – metode ini menggunakan *coding*, aplikasi, hingga *extension browser*. *Tool* ini sekarang lebih digemari karena dapat melakukan *scraping* secara cepat.

Penerapan scraping dapat menggunakan beberapa teknik, antara lain:

* Copy-Paste
* HTML Parsing
* DOM Parsing
* Vertical Aggregation
* Xpath
* Google Sheet
* Text Patern Machine

**Cara Kerja Web Scraping**

Cara kerja web scraper cenderung sederhana sekaligus rumit. Pertama, web scraper akan diberikan sejumlah URL untuk dimuat sebelum dilakukan proses scraping. Scraper kemudian memuat kode HTML laman yang dituju. Beberapa scraper yang lebih canggih bahkan mampu memuat seluruh isi dalam website, termasuk CSS dan elemen Javascript.

Dari data yang sudah dimuat, scraper kemudian akan mengekstrak data yang sebelumnya sudah dipilih oleh pengguna sebelum program dijalankan. Dia akan memprogram scraper dengan mengutamakan dua informasi tersebut dan mengabaikan review produk atau komentar pembeli.

Data yang sudah diekstrak kemudian akan dikumpulkan dalam satu format, yang nantinya akan digunakan oleh data scientist untuk proses analisis. Format yang digunakan dapat berupa CSV, Excel, atau bahkan format JSON yang nantinya bisa digunakan sebagai API.

**Pengertian Web Crawling**

Web crawling adalah proses di mana mesin pencari mengirimkan tim robot (crawler atau spider) dalam mencari dan memindai konten yang berada di halaman website. Di mana konten ini dapat berupa artikel, gambar, video, ataupun dokumen.

Alat yang digunakan dalam web crawling adalah web crawler atau sering juga disebut sebagai web spider. Alat ini akan menemukan konten terbaru dengan mengidentifikasi dan merekam setiap link yang ditemukannya pada halaman yang telah dipindai, lalu memasukkannya ke dalam indeks berupa database yang berisi URL. Adapun beberapa contoh tools populer untuk web crawling adalah Googlebot, HTTrack, Cyotek Webcopy, dan Webhose.

Lebih singkatnya, Crawling merupakan teknik mengumpulkan data pada sebuah website dengan memasukkan **Uniform Resource Locator (URL).** URL ini menjadi acuan untuk mencari semua hyperlink yang ada pada website. Kemudian dilakukan indexing untuk mencari kata dalam dokumen pada setiap link yang ada.

Penerapan crawling dapat menggunakan beberapa teknik, antara lain:

* Selective Crawling
* Popularity
* Focused Crawling
* Distributed Crawling
* Paralel Crawling
* Web Dynamic

**Cara Kerja Web Crawler**

Sebuah website tidak dapat ditemukan begitu saja oleh search engine. Search engine harus melakukan crawling dan indexing sebelum akhirnya menampilkan konten website pada SERP mereka. Proses ini dilakukan dengan bantuan tools yang disebut web crawler, web robot, atau web spider.

Pertama, web crawler akan mengunjungi sebuah situs dan berbagai link yang terdapat dalam laman tersebut. Namun jika situsmu terbilang baru dan belum ada link lain di dalamnya, kamu bisa meminta search engine untuk mendatangi situsmu, seperti dikutip dari WebFX.

Caranya mudah. Kamu hanya perlu memasukkan URL situsmu di Google Search Console. Kemudian, tugas tools web crawling berikutnya adalah mencatat setiap link yang mereka temukan ke indeks mereka.

Namun, perlu kamu catat bahwa web crawler hanya akan mengumpulkan informasi dari laman yang bersifat publik, ya. Web crawler tidak ikut mencatat laman private yang tidak dapat diakses.

Setelah itu, web crawler akan mengumpulkan berbagai informasi, seperti tulisan dan meta tag. Informasi tersebut akan tersimpan dalam indeks search engine sehingga dapat muncul ketika pengguna mencari konten dengan keyword yang serupa.

**Perbedaan Web Crawling dengan Web Scraping**

Sekilas, kedua istilah ini memang terdengar mirip. Namun, keduanya sebenarnya berbeda, baik dari segi definisi, fokus dan cakupan, tujuan, pengetahuan terhadap website, hingga output-nya.

1. **Definisi**

Secara definisi, web crawling adalah proses membaca dan menyimpan seluruh konten ke dalam sebuah website dengan tujuan pengarsipan atau indexing. Sementara web scraping adalah proses mengekstrasi data dari sebuah website ke format file yang baru.

1. **Fokus dan Cakupan**

Fokus dan cakupan web crawling lebih besar karena lingkupnya adalah seluruh halaman dan website yang ada di internet, sedangkan lingkup web scraping hanya pada kumpulan data spesifik dari sebuah website.

1. **Tujuan**

Web crawling bertujuan mencari dan menemukan URL atau link di internet, sedangkan web scraping bertujuan menghasilkan data sebuah website target untuk dianalisis lebih jauh.

1. **Pengetahuan Terhadap Website**

Web crawling tidak perlu tahu URL atau domain yang ingin di-crawling karena tujuannya memang untuk mencari, menemukan, dan mengindeks URL tersebut. Sementara web scraping tahu di domain mana data akan diambil dari sebuah website.

1. **Output**

Output dari web crawling adalah daftar URL yang telah di-crawl, sedangkan output web scraping adalah data field seperti nama produk, harga produk, dan ukuran.

**Kesimpulan**

Ketika kita melakukan web crawling sebenarnya kita juga melakukan web scraping. Namun sebaliknya ketika kita melakukan web scraping kita belum atau tidak melakukan web crawling. Kemudian untuk perbedaannya crawling biasanya digunakan untuk data dengan skala besar. Penerapannya menggunakan bot secara otomatis dan menggunakan Application Programming Interface (API). Sedangkan scraping biasanya digunakan untuk data yang relative tidak terlalu besar dan proses pengambilan data pada element HTML ataupun XML menggunakan protocol HTTP.

**Pengertian Google AdSense**

Seperti yang sempat dijelaskan sebelumnya, Google AdSense adalah program periklanan besutan Google. Layanan periklanan ini memiliki basis CPC atau Cost Per Click dan memungkinkan pemilik situs atau website untuk memperoleh penghasilan melalui iklan yang terpasang pada situs mereka.

Layanan dari Google AdSense menampilkan iklan di sebuah situs dengan bentuk teks, gambar, maupun video yang nantinya akan ditunjukkan kepada pengunjung situs. Iklan yang ditampilkan kepada pengunjung situs pun disesuaikan dengan minat.

Google AdSense akan melihat cache aplikasi browser untuk mengetahui histori dari aktivitas internet pengunjung situs. Nantinya, iklan yang ditampilkan kepada pengunjung tersebut akan disesuaikan dengan ketertarikan pengunjung sehingga pemasang iklan mampu mendapatkan audiens iklannya dengan lebih akurat.

Dalam kata lain, semakin banyak pengunjung yang membuka atau mengeklik iklan yang terpasang pada situsmu, artinya pendapatan dari Google AdSense akan menjadi lebih tinggi lagi. Akan tetapi, untuk bisa melakukan hal tersebut, kamu harus menguasai dulu cara membuat konten di situs sesuai dengan SEO atau Search Engine Optimization.

**Cara Kerja Google AdSense**

1. Terdapat Ruang Iklan pada Situs

Agar situs mampu mendapatkan iklan dari layanan Google AdSense, kamu harus menyediakan dulu ruang untuk meletakkan iklan tersebut. Biasanya, peletakan iklan dari layanan ini berada di bagian header, tengah artikel, sidebar, dan juga footer.

2. Iklan dengan Biaya paling Mahal Menjadi Prioritas

Jumlah pengunjung dari sebuah situs juga bervariasi dengan jenis konten yang beragam pula. Nah, untuk menyiasati hal tersebut, pihak Google AdSense biasanya akan memunculkan terlebih dahulu iklan yang memiliki harga paling tinggi. Baru iklan dengan harga lebih rendah akan dimunculkan secara bergantian di ruang iklan sebuah situs.

3. Pendapatan Tergantung dari Banyaknya Iklan Dibuka dan Persaingan Bid

Perlu diingatkan kembali, pendapatan dari Google AdSense hanya akan diberikan saat iklan diklik atau dibuka oleh pengunjung. Artinya, jika iklan hanya dilihat, pemilik situs tidak akan mendapatkan penghasilan dari layanan periklanan tersebut.

DAFTAR PUSTAKA / LINK MATERI

1. <https://accurate.id/digital-marketing/crawling-adalah/#:~:text=1.-,Definisi,ke%20format%20file%20yang%20baru.>
2. <https://medium.com/@dede.brahma2/perbedaan-antara-crawling-dan-scraping-98e64e0c6439>
3. <https://glints.com/id/lowongan/web-crawling-adalah/#.Y8TJH3ZBzak>