

BAB 2. LANDASAN TEORI

2.1. *Web Scraping*

Web Scraping adalah teknik untuk mendapatkan informasi dari *website* secara otomatis tanpa harus menyalinnya secara manual. Tujuan dari web scaper adalah untuk mencari informasi tertentu dan kemudian mengumpulkannya dalam web yang baru.

Web scraping berfokus dalam mendapatkan data dengan cara pengambilan dan ekstraksi. (Josi, A., Abdillah, L. A., & Suryayusra., 2014).

Web scraping (panen web) adalah pengambilan sebuah dokumen semi-terstruktur dari internet, umumnya berupa halaman-halaman web dalam bahasa markup seperti HTML atau XHTML, dan menganalisis dokumen tersebut untuk diambil data tertentu dari halaman tersebut. Istilah gampangnya yaitu pengambilan konten atau sebagian data dari suatu situs web. Web scraping mempunyai banyak kegunaan dan sangat membantu masalah pengambilan data. Salah satunya yaitu untuk konten artikel dimana isi kontennya langsung diambil dari situs lain misal konten dari wikipedia. Contohnya pada pesonainformatika.com misalkan kata kunci yang dicari tidak ditemukan, maka otomatis akan melakukan scraping dari situs wikipedia berdasarkan kata kunci atau keyword yang dicari. (Akhi Syabab, 2018).

2.2. *E-Marketplace*

E-Marketplace merupakan pasar *virtual* dimana penjual dan pembeli bertemu dan melakukan berbagai jenis transaksi. Di sini orang melakukan proses transaksi dengan pertukaran barang maupun jasa untuk menghasilkan uang. Setiap individu juga dapat membuka pasar yang menjual produk atau layanan secara *online*. Pasar elektronik terhubung ke penjual dan pembeli melalui Internet atau ke mitra dalam organisasi, intranet. (Turban, E. 2012. *Electronic Commerce* (8th ed). Pearson).

2.3. PHP

PHP adalah bahasa script dan interpreter yang tersedia secara gratis dan digunakan terutama di server Web Linux. PHP, yang awalnya berasal dari Personal Home Page Tools, sekarang merupakan singkatan dari PHP: Hypertext Preprocessor, yang digambarkan sebagai "akronim rekursif". PHP dijalankan di server, sementara alternatif yang sebanding, JavaScript, dijalankan pada klien. PHP adalah alternatif untuk teknologi Active Server Page (ASP) Microsoft. (Rouse, M., 2006)

2.4. HTML

HTML adalah singkatan dari Hypertext Markup Language. Disebut hypertext karena di dalam HTML sebuah text biasa dapat berfungsi lain, sehingga dapat dibuat menjadi link yang dapat berpindah dari satu halaman ke halaman lainnya hanya dengan meng-klik text tersebut. Kemampuan text inilah yang dinamakan *hypertext*, walaupun pada implementasinya nanti tidak hanya text yang dapat dijadikan link. HTML disebut sebagai Markup Language karena bahasa HTML menggunakan tanda (mark), untuk menandai bagian-bagian dari text. Misalnya, text yang berada di antara tanda tertentu akan menjadi tebal, dan jika berada di antara tanda lainnya akan tampak besar. Tanda ini di kenal sebagai HTML tag. (Andre, 2017).

2.5. JavaScript

JavaScript adalah sebuah bahasa script pemrograman berbasis objek. JavaScript digunakan pada pembuatan aplikasi client server berbasis internet. Program JavaScript diletakkan pada halaman HTML. JavaScript menyediakan script writer dengan compact pre-built tools yang meningkatkan interaksi antara pengguna dengan halaman HTML. *Tools* tersebut dapat merespon klik dari mouse, input dari form, navigasi halaman, dan event lainnya. Respon dari *event* yang user lakukan dapat dipanggil oleh JavaScript tanpa *network transmissions*. Hal tersebut merupakan kegunaan utama dari JavaScript pada pembuatan script seperti CGI (Common Gateway Interface). JavaScript berbeda dengan Java. Java adalah pemrograman berbasis objek secara penuh oleh karena itu, bahasa Java mampu membaca program code yang berada pada text file yang kemudian di-compile dan

link untuk membuat executable. JavaScript merupakan scripting language yang tidak membuat executable file. JavaScript biasanya dipanggil oleh software yang sedang berjalan dan disesuaikan untuk dibaca oleh software tersebut. Software yang sedang berjalan memiliki modul yang mengenali script statement tersebut dan “tahu” bagaimana cara execute script yang dibuat (Cohen, 1997).

2.6. Bootstrap

Bootstrap diciptakan oleh dua orang programmer Twitter, yaitu Mark Otto dan Jacob Thornton pada tahun 2011. Bootstrap awalnya diciptakan untuk membuat standarisasi pekerjaan tool dan library untuk pekerjaan programmer Twitter. Kemudian Bootstrap berkembang dan populer pada kalangan programmer. Bootstrap adalah framework atau tools yang digunakan untuk membuat aplikasi web ataupun situs web responsive secara cepat, mudah, dan gratis. Bootstrap terdiri dari CSS dan HTML untuk menghasilkan grid, layout, typography, table, form, navigation, dan lain-lain. Di dalam Bootstrap juga sudah terdapat jQuery plugins untuk menghasilkan komponen user interface yang cantik seperti transitions, modal, dropdown, scrollspy, tooltip, tab, popover, alert, button, carousel, dan lain-lain. Kegunaan Bootstrap adalah membuat responsive website dengan cepat dan mudah dan dapat berjalan pada web browser umum seperti Chrome, Firefox, Safari, Opera, dan Internet Explorer (Alatas, 2013, p. 2).

2.7. PHP Simple HTML DOM Parser

Ide asli didapat dari HTML Parser for PHP 4 yang dibuat oleh Jose Solorzano. Dimana pada halaman php-html.sourceforge.net dijelaskan bahwa HTML Parser for PHP 4 adalah *open source HTML parser* yang ditulis menggunakan bahasa PHP. Dimana *parser* ini dilengkapi dengan alat yang mengkonversi HTML ke teks, sebagai contohnya. *Parser* ini dirancang untuk kecepatan dan fleksibilitas. *Parser* ini tidak membuatkan model objek yang baru, tetapi tidak mencegah apabila menggunakannya untuk membuat model objek jika diperlukan. Cara kerjanya adalah dengan meminta elemen atau *node* berikutnya di dokumen sesuai dengan kebutuhan. (Chen, 2014).

2.8. Selenium Webdriver

Selenium Webdriver adalah kumpulan API open source yang digunakan untuk mengotomatisasi pengujian aplikasi web. Selenium Webdriver adalah platform independen karena kode yang sama dapat digunakan pada Sistem Operasi yang berbeda seperti Microsoft Windows, Apple OS dan Linux. Selenium Webdriver juga memiliki dukungan yang lebih baik karena halaman web dinamis seperti Ajax, di mana elemen-elemen halaman web dapat berubah tanpa halaman itu sendiri sedang dimuat ulang. (Bennett, 2018). Cara menggunakan Selenium Webdriver adalah dengan melakukan instalasi *webdriver* (misalnya *chromedriver*) terlebih dahulu. Masukkan *option* untuk penggunaan webdriver pada skrip. Gunakan *option headless* agar GUI dari chrome tidak muncul sehingga chrome dapat dijalankan pada *background*. Kemudian panggil webdriver yang telah disediakan dengan memberikan *path* untuk *file executable* dari *chromedriver*. Setelah itu maka *chromedriver* berhasil dipanggil.

2.9. ChromeDriver

WebDriver adalah perangkat *open source* untuk pengujian otomatis dari aplikasi *website* di banyak *browser*. WebDriver menyediakan kemampuan untuk menavigasi ke halaman web, input pengguna, eksekusi JavaScript, dan banyak lagi. ChromeDriver adalah server mandiri yang menerapkan protokol WebDriver untuk Chromium. ChromeDriver adalah *tool* terpisah yang digunakan WebDriver untuk mengontrol Chrome. ChromeDriver dikelola oleh tim Chromium dengan bantuan dari kontributor WebDriver. (Khachatryan, 2018).

2.10. Python

Python merupakan bahasa pemrograman tingkat tinggi yang diracik oleh Guido van Rossum. Python banyak digunakan untuk membuat berbagai macam program, seperti: program CLI, Program GUI (desktop), Aplikasi Mobile, Web, IoT, Game, Program untuk Hacking, dsb. Python juga dikenal dengan bahasa pemrograman yang mudah dipelajari, karena struktur sintaknya rapi dan mudah dipahami. (Dian, 2014).

2.11. Aplikasi Sejenis

Perbandingan dengan situs sejenis dapat dilihat pada Tabel 2.1. PricePanda adalah situs pembanding harga terkemuka pada pasar yang sedang berkembang. Situs ini memberikan gambaran akan harga untuk produk elektronik dan lainnya kepada konsumen secara transparan. Selain itu, situs PricePanda juga memberikan ulasan produk secara profesional dan rinci sama halnya memberikan dengan ulasan mengenai toko. (pricepanda.co.id, 2012).

Priceprice.com adalah situs pembanding harga untuk membantu pengguna berbelanja secara pintar dengan menyediakan informasi harga /produk/pendapat mulut ke mulut tentang berbagai macam produk, seperti Komputer, Kamera, Elektronik dan Ponsel. (priceprice.com ,2013).

Fitur user tidak perlu *login* untuk memanfaatkan situs ini diperlukan agar *user* tidak perlu *login* namun sudah dapat menggunakan situs. Dalam *website* yang dibuat, daftar produk yang ditampilkan diurutkan berdasarkan berapa jumlah produk terjual, bukan berdasarkan harga termurah agar *user* dapat melihat produk apa yang laris terjual di tokopedia, bulakapak, dan shopee. *Website* yang dibuat tidak menampilkan artikel berdasarkan kategori. *Website* yang dibuat memiliki fitur *sort* harga seperti pricepanda.co.id dan priceprice.com untuk memudahkan apabila *user* ingin melihat produk diurutkan berdasarkan harga mana yang lebih rendah maupun lebih tinggi. *Website* yang dibuat juga memiliki filter kategori untuk mempermudah pencarian daftar produk terlaris.

Tabel 2.1. Tabel Perbandingan dengan Situs Sejenis

Fitur	pricepanda.co.id	priceprice.com	Website yang dibuat
User tidak perlu login untuk memanfaatkan situs.	✓	✓	✓
Menampilkan produk dari banyaknya jumlah produk yang terjual (produk terlaris).	X	X	✓

Menampilkan tampilan awal produk berdasarkan harga termurah.	X	✓	X
Adanya informasi tambahan/artikel berdasarkan kategori.	✓	✓	X
Memiliki fitur <i>sort</i> berdasarkan harga.	✓	✓	✓
Memiliki filter kategori untuk mempermudah pencarian.	✓	✓	✓