

## BAB II

### LANDASAN TEORI

#### A. *Web Scraping*

Menurut Ayani, Pratiwi dan Muhandi (2019:258) *Web scraping* adalah teknik untuk mendapatkan informasi dari *website* secara otomatis tanpa harus menyalinnya secara manual. Tujuan dari *web scraper* adalah untuk mencari informasi tertentu dan kemudian mengumpulkannya dalam *web* yang baru. *Web scraping* berfokus dalam mendapatkan data dengan cara pengambilan dan ekstraksi. Manfaat dari *web scraping* ialah agar informasi yang dikeruk lebih terfokus sehingga memudahkan dalam melakukan pencarian sesuatu.

Maka dari itu dengan menggunakan teknik *web scraping* proses mendapatkan informasi pada sebuah *website* menjadi lebih mudah dan cepat yang tentunya menggunakan sebuah *library* yang tersedia sesuai dengan bahasa pemrograman yang dipilih.

Menurut Flores, Permatasari dan Jasa (2020: 159) ada beberapa pedoman umum yang harus dipatuhi oleh para *web scrapper*.

1. Tidak ada kata sandi atau hambatan yang dilanggar untuk mengambil konten.
2. Konten yang diambil tidak boleh dilindungi hak cipta.
3. Jika dilindungi hak cipta, konten yang diambil harus mematuhi standar yang adil.
4. Tindakan *scraping* tidak boleh membebani layanan situs *webnya*.

5. *Scraping* tidak boleh melanggar ketentuan penggunaan situs yang sedang diambil.
6. *Scraping* tidak boleh mengumpulkan informasi pengguna yang sensitif.

## **B. Detik.com**

Detik.com adalah salah satu situs berita atau media online yang menawarkan semua jenis kategori berita mulai dari berita terbaru, berita otomotif sampai berita olahraga, detik.com juga mengkatagorikan berita sesuai dengan provinsi masing-masing sesuai berita yang ada diwilayah tersebut jadi informasi lebih spesifik dan jelas.

## **C. Bencana Banjir**

Bencana banjir adalah bencana alam yang sering terjadi di negara indonesia apalagi pada saat musim hujan potensi terjadi banjir akan semakin meningkat hal tersebut yang patut diwaspadai karena menimbulkan masalah seperti gatal-gatal, kekurangan air bersih bahkan merusak rumah-rumah warga yang terendam banjir.

Menurut Suprpto (2011:86) bahwa Potensi bencana banjir di Indonesia sangat besar dilihat dari topografi dataran rendah, cekungan dan sebagian besar wilayahnya adalah lautan. Curah hujan di daerah hulu dapat menyebabkan banjir di daerah hilir. Apalagi untuk daerah-daerah yang tinggi permukaan tanahnya lebih rendah atau hanya beberapa meter di atas permukaan air laut. Berdasarkan data dan informasi bencana indonesia yang dikelola Badan Nasional

Penanggulangan Bencana (BNPB) menunjukkan bahwa bencana banjir merupakan kejadian terbanyak.

#### D. Simple HTML DOM

Simple HTML DOM adalah sebuah *library* dari sekian banyak *library* bahasa pemrograman PHP yang memiliki fungsi untuk melakukan *web scraping* jadi dengan menggunakan *library* tersebut proses *scraping* menjadi lebih efektif.

Menurut Mulyani, Sutanto dan Yulianton (2016 :60) bahwa Simple HTML DOM Parser pada dasarnya akan mencari struktur HTML tertentu pada suatu halaman *web*. Penggunaannya dimulai dengan menentukan DOM. *Document Object Model* (DOM) adalah *object* model standar untuk HTML dan XML yang bersifat platform independent. Menentukan DOM dapat dilakukan dengan cara mudah, yaitu alamat URL dari *website* yang akan diambil sebagian isinya. Setelah menentukan DOM, hal yang dilakukan adalah menemukan tag HTML yang akan diambil isinya. Hal ini dilakukan dengan cara menganalisa kode program atau HTML *website*.

```
1. // panggil library simple dom parser
2. include('lib/simple_html_dom.php');
3.
4. // URL yang akan kita ambil kontennya
5. $html = file_get_html('https://www.google.com/');
6.
7. //find semua gambar dalam sebuah URL
8. foreach($html->find('img') as $element)
9.     echo $element->src . '<br>';
10.
11. //find semua link dalam sebuah URL
12. foreach($html->find('a') as $element)
13.     echo $element->href . '<br>';
```

Gambar 2.1 Contoh Source Code Library Simple HTML DOM

### **E. *Web Browser***

*Web browser* adalah sebuah perangkat lunak yang berguna untuk pencarian *website* yang tersebar, dengan *web browser* kita juga bisa mengakses untuk menampilkan sumber informasi yang ada di internet, maka dari itu *web browser* adalah perangkat lunak yang sangat penting di era saat ini, salah satu aplikasi *web browser* yang populer adalah google Chrome.

### **F. *Website***

*Website* atau situs dapat diartikan sebagai kumpulan halaman-halaman yang digunakan untuk menampilkan informasi teks, gambar diam atau gerak, animasi, suara, dan atau gabungan dari semuanya, baik yang bersifat statis maupun dinamis yang membentuk satu rangkaian bangunan yang saling terkait, yang masing-masing dihubungkan dengan jaringan-jaringan halaman. Hubungan antara satu halaman *web* dengan halaman *web* yang lainnya disebut *Hyperlink*, sedangkan teks yang dijadikan media penghubung disebut *Hypertext* (Hidayat, 2010:2).

*Website* ini adalah sebuah situs yang memiliki banyak sekali kegunaan mulai dari media promosi, sistem informasi, media penjualan serta banyak sekali kegunaan lainnya dan *website* ini sangatlah fleksibel karena bisa diakses dengan semua perangkat dan semua jenis sistem operasi karena hanya dengan menggunakan koneksi internet *user* bisa bebas mengakses halaman *web* yang ingin dibuka atau diketahui.

## G. PHP

Yudhanto dan Prasetio (2019:9) menyatakan bahwa PHP atau *Hypertext Preprocessor* adalah bahasa pemrograman *script server side* yang sengaja dirancang lebih cenderung untuk membuat dan mengembangkan *web*. Bahasa pemrograman ini memang dirancang untuk para pengembang *web* agar dapat menciptakan suatu halaman *web* yang bersifat dinamis. Sekilas tentang PHP, PHP pertama kali dibuat oleh Rasmus Lerdorf pada tahun 1995 dan terus dikembangkan hingga saat ini. Ada banyak sekali *web* termasuk CMS yang dibuat menggunakan bahasa PHP, seperti WordPress dan lain-lain.

Bahasa pemrograman PHP ini memang sangat cocok untuk pembuatan *web* karena bahasa pemrograman ini memiliki banyak sekali *tool* dan *library* yang sangat berguna untuk *user* dalam pengembangan *web* dengan menggunakan bahasa pemrograman PHP ini.

## H. MySQL

MySQL ini sistem manajemen *database* yang sangat berperan untuk pembuatan *website* dinamis sebab *website* dinamis memerlukan sebuah data yang perlu dikelola agar *website* tampak selalu *update*, MySQL ini sangat cocok bila dipadukan dengan bahasa pemrograman PHP sebab keduanya sangat membantu untuk pembuatan *website* yang lebih interaktif dan produktif lagi.

Yudhanto dan Prasetio (2019:17-18) menyatakan bahwa MySQL yang merupakan singkatan dari “*My Structured Query Language*” adalah *database* yang paling favorit saat ini. Program ini berjalan sebagai server yang

menyediakan *multi-user* mengakses ke sejumlah database baik *multithread* ataupun *multi-user* dan telah diinstall oleh sekitar 6 juta kali diseluruh dunia.

### **I. Database**

Yudhanto dan Prasetio (2019:16) menyatakan bahwa basis data (*database*) adalah kumpulan data yang disimpan secara sistematis didalam komputer yang dapat diolah atau dimanipulasi menggunakan perangkat lunak (program aplikasi) untuk menghasilkan informasi. Definisi basis data meliputi spesifikasi berupa tipe data, struktur data dan juga batasan-batasan pada data yang akan disimpan.

Dengan adanya *database* semua data bisa kita simpan dan kumpulkan sesuai dengan jenis dan bagian masing-masing, data yang ada dalam *database* bisa dikelola baik dengan melakukan penambahan data, edit data, hapus data hingga menampilkan data.

### **J. Visual Studio Code**

Visual Studio Code ini adalah sebuah *software* teks editor yang berguna untuk menulis baris-baris *coding*, visual studio code ini juga bisa digunakan untuk menulis *coding* bahasa pemrograman PHP.

Menurut Salamah (2021:1) Visual Studio Code (VS Code) ini adalah sebuah teks editor ringan dan handal yang dibuat oleh Microsoft untuk sistem operasi *multiplatform*, artinya tersedia juga untuk versi Linux, Mac, dan Windows. Teks editor ini secara langsung mendukung bahasa pemrograman JavaScript, Typescript, dan Node.js, dengan bahasa pemrograman lainnya.