

Implementation of Monitoring New Application on Permaweb Ecosystem

Dinda Anik Masruro, Rolly Maulana Awangga
Program Studi D4 Teknik Informatika
Politeknik Pos Indonesia, Jalan Sari Asih No.54,Bandung,Jawa Barat
dindaanik19@gmail.com, awangga@poltekpos.ac.id

Abstrak

Saat ini banyak sekali informasi yang dapat dengan mudah diakses melalui internet, banyak sekali data atau informasi yang menyebar ke internet dengan mudah. Namun, banyaknya data yang ada dapat menyulitkan kita untuk menganalisis suatu data karena banyaknya data yang harus diakses. Web Scraper adalah cara mengekstrak skrip yang berjalan di sana yang biasanya dipilih untuk melakukan memo web melalui parsing DOM. Node tertentu yang dikumpulkan menggunakan parser DOM dan alat seperti XPath membantu proses pengikisan halaman web. Pada penelitian ini akan menggunakan metode DOM Parsing untuk mendapatkan mengetahui data yang ada pada platform arweave. Sraeper dapat dikembangkan lebih lanjut sebagai teknik pengumpulan data di internet untuk penelitian lebih lanjut yang dapat diberikan tentang konsep big data yang akan digunakan untuk peramalan atau mendapatkan informasi yang dibutuhkan nantinya. Alone Parsing adalah cara pemecahan data atau simbol, baik dalam bahasa maupun dalam bahasa, menurut aturan tata bahasa formal. Setelah menggunakan teknik parshing nantinya akan di generate kembali menggunakan Parse Tree yaitu proses menyusun data yang berbentuk seperti pohon dan menggunakan analisis sintak untuk memecahkan masalah.

Kata Kunci : Data, Permaweb, Web Scraping, Parsing-data, Website, DOM

Abstract

Currently a lot of information that can be easily accessed via the internet, a lot of data or information that spreads to the internet easily. However, the large amount of data that exists can make it difficult for us to analyze a data because of the large amount of data that must be accessed. Web Scraper is a way of extracting scripts running there which are usually chosen to do web scraping via DOM parsing. Specific nodes collected using DOM parsers and tools like XPath help with the scraping of web pages. In this study, we will use the DOM Parsing method to find out the data on the Arweave platform. Straper can be further developed as a data collection technique on the internet for further research that can be given about the concept of big data that will be used for forecasting or to get the information needed later. Alone Parsing is a way of splitting data or symbols, both in language and in language, according to formal grammatical rules. After using the parshing technique, it will be generated again using a Parse Tree, which is the process of compiling data in the form of a tree and using syntax analysis to solve problems.

Keywords : Data, Permaweb, Web Scraping, Parsing-data, Website, DOM

1. INTRODUCTION

Web scraping kini telah menjadi trend yang banyak dilakukan oleh perusahaan atau perorangan untuk tujuan tertentu. Teknik itu sendiri adalah metode untuk mengekstrak data dari halaman web. Anda juga dapat menyalin

data detail dari halaman web ke halaman spreadsheet secara manual, tetapi biasanya data di situs web adalah data yang cukup besar sehingga membutuhkan ruang berkapasitas besar dan memakan waktu cukup lama. Oleh karena itu, salah satu cara untuk mendownload big data dari website adalah dengan

menggunakan "web scraper". Web scraper adalah program yang dapat membuka halaman web dan kemudian mengunduh data yang sudah ada di web, mengekstraknya ke dalam format terstruktur, dan menyimpannya ke dalam file atau database. Pengikis web juga dapat mengunduh konten yang biasanya berupa teks dan diformat sebagai HTML dari beberapa halaman web dan mengekstrak data darinya. Dalam penelitian ini data scraping menjadi hal yang populer digunakan untuk mengetahui data pada arweave. PermaWeb adalah jaringan operasional untuk ekosistem Arweave DApps. Dengan adanya PermaWeb ini memberikan pengembang dengan fleksibilitas yang besar dan membuat pengembangan DApp lebih mudah. Pengembang hanya perlu menggunakan API standar berdasarkan paradigma konsensus tersimpan untuk berkembang, tanpa batasan bahasa pemrograman, sedangkan Areweave (AR) adalah proyek penyimpanan terdesentralisasi yang memungkinkan penggunanya untuk mengunggah data dan menyimpannya secara permanen. Platform ini sendiri adalah dasar dari apa yang disebutkan timnya sebagai "PermaWeb" atau web terdesentralisasi permanen yang menampung dari banyak aplikasi dan platform berbasis komunikasi. Token AR sendiri digunakan untuk memberi penghargaan kepada penambang/pengembang yang membagikan dan menyimpan semua data itu dan membayar untuk menyimpan data baru apa pun. Dimana untuk mengetahui berapa data yang ada bahkan jutaan dapat dilakukan dengan menggunakan teknik scraping untuk mengetahui perkembangan informasi. Sehingga dalam penelitian ini akan menggunakan metode DOM Parsing. Secara umum juga digunakan untuk mengetahui cara kerja internal halaman web dan mengekstrak skrip yang berjalan di dalamnya, biasanya memilih untuk melakukan pengikisan web melalui parsing DOM. Node tertentu dikumpulkan menggunakan pengurai DOM dan alat seperti XPath membantu menggores halaman web.

2. STUDIES IN WEB SCRAPING

Scrape data adalah istilah yang digunakan untuk menggambarkan ekstraksi data dari file elektronik menggunakan sebuah program komputer[2]. *Web Scraping* menjelaskan penggunaan program untuk mengekstrak data dari file *HTML*[4] di internet. Biasanya data ini berupa data berpola, khususnya daftar atau

tabel. Program yang berinteraksi dengan halaman web dan mengekstrak data menggunakan set perintah yang dikenal sebagai antarmuka pemrograman aplikasi (API). API ini dapat diajarkan untuk mengekstrak data berpola dari halaman web tunggal atau dari semua halaman serupa di seluruh situs web. Kalau tidak, interaksi otomatis dengan situs web dapat dibangun ke dalam API[3], sehingga tautan di dalam halaman dapat diklik dan data diekstraksi ke halaman selanjutnya. Ini sangat berguna untuk mengekstraksi data dari beberapa halaman hasil pencarian. Selanjutnya, interaktivitas ini memungkinkan pengguna untuk mengotomatisasi penggunaan fasilitas pencarian situs web, mengekstraksi data dari beberapa halaman hasil pencarian dan hanya mengharuskan pengguna untuk memasukkan istilah pencarian daripada harus menavigasi dan mencari setiap web situs terlebih dahulu. Salah satu penggunaan utama saat ini untuk pengikisan web adalah agar bisnis melacak aktivitas penetapan harga produk mereka pesaing: penetapan harga dapat ditetapkan di seluruh situs dalam skala waktu yang relatif singkat dan dengan sedikit usaha manual. Berbagai driver komersial lainnya telah menyebabkan sejumlah besar dan berbagai program web scraping telah dikembangkan dalam beberapa tahun terakhir. Beberapa program-program ini gratis, sementara yang lain murni komersial dan mengenakan biaya satu kali atau reguler biaya berlangganan.

Web scraping merupakan suatu metode pengumpulan data melalui internet, walaupun web scraping bukanlah sesuatu yang baru, belakangan ini web scraping sangat populer digunakan untuk pemenuhan data mining[7]. Sebelumnya metode ini dikenal dalam beberapa istilah, antara lain *screen scraping*, *data mining*, *web harvesting* atau metode lain yang sejenis. Secara teori, *web scraping* adalah cara untuk mengumpulkan data menggunakan metode yang berbeda dari menggunakan *API* (*Application Programming Interface*). Cara ini biasanya dimulai dengan penulisan kode program. Yang digunakan sebagai otomatisasi *query* untuk meminta data ke server. Data dari permintaan dapat diekstraksi untuk menghasilkan informasi yang akan dicari[6].

Pada penelitian sebelumnya yang telah dilakukan oleh Shikha Mahajan dan Nikhit Kumar dalam jurnal yang berjudul "A Web scraping Approach in Node.js", yang membahas tentang web scraping dapat

diterapkan untuk mengambil gambar, tabel dan informasi lainnya dari website. Metode ini dapat ditingkatkan untuk menangani loop tak terbatas saat menggunakan tautan untuk melintasi situs web. Dalam beberapa kasus teknologi web scraping mungkin menjadi satu-satunya solusi yang dapat diterapkan ketika ingin mengumpulkan data dalam jumlah besar [8]. Sementara itu, Priyanto [6] menggunakan web scraping untuk mendapatkan informasi yang akurat tentang hidroponik.

Ada 3 metode yang biasanya digunakan untuk scraping situs web, salah satunya adalah Parsing HTML adalah metode yang paling sering digunakan dalam proses parsing data dari halaman web. Umumnya, penguraian HTML dilakukan menggunakan JavaScript dan menargetkan halaman HTML linier dan bersarang. Script ini digunakan untuk mengekstrak posting, link dan data[10]. Dan yang kedua Reguler Metode ekspresi berguna jika Anda ingin melakukannya tugas ekstraksi data sederhana. Misalnya seperti ketika Anda ingin mendapatkan daftar semua email dari halaman web. Ekspresi Reguler ini tidak cocok untuk pekerjaan ekstraksi yang kompleks, seperti mengekstraksi data dari beberapa halaman deskripsi produk di situs web E-niaga. Tapi itu akan sangat berguna untuk proses transformasi dan pembersihan data. Dan yang terakhir adalah penguraian DOM yang digunakan dalam penelitian ini adalah metode yang paling cocok untuk mengambil data di situs web karena memiliki beberapa metode utama yang dapat dukungan untuk pengumpulan data.

Dari beberapa penggunaan web scaping pada penelitian terdahulu dan penelitian yang sedang diusulkan web scraping dengan menggunakan python lebih memudahkan dan lebih efisien karena tidak begitu membutuhkan waktu lama untuk mengetahui isi keseluruhan website maupun situ-situs lainnya, dengan menggabungkan perkembangan teknologi dan perkembangan ekonomi saat ini. Perancangan web scraping dari penelitian yang diusulkan ini akan menghasilkan hasil ouputan berupa file sehingga lebih efisien untuk melihat dan membaca filenya yang dapat berupa file csv.

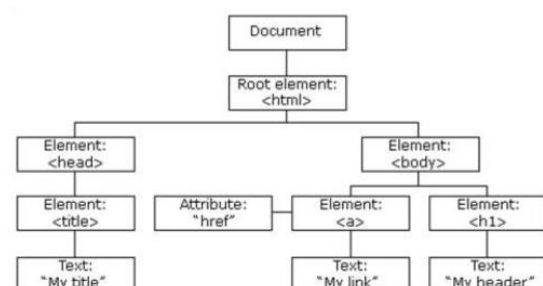
3. PROBLEM DEFINITION

Peramban web membuat tampilan dari dokumen web. Di latar belakang proses rendering ini, ada tiga operasi tambahan termasuk konstruksi Pohon DOM, pemrosesan CSS (Cascading Style Sheets) dan eksekusi

JavaScript. Gambar 1 menunjukkan halaman web sederhana dan Pohon DOM yang dibangun dari sebuah dokumen, yang telah ditentukan oleh W3C (World Wide Web Consortium) yang merupakan komunitas internasional yang bertanggung jawab untuk mengembangkan dari standar Web. Halaman web memiliki tag seperti `<html>`, `<head>`, `<title>`, `<body>`, ``, dan `<h1>` yang diberikan pada Gambar 1. Selain itu, banyak tag yang berbeda dapat digunakan di halaman web. Tag pembuka suatu elemen (OTE) mungkin memiliki atribut dan nilainya yang didefinisikan sebagai `A D {a1 D v1, a2 D v2, : : :, an D vn}`. Atribut seperti `id`, `class`, dan `itemprop` yang ditunjukkan pada Gambar 1 membantu memberikan informasi tambahan tentang elemen. Informasi tambahan ini dapat digunakan untuk memformat yang diinginkan

4. EXTRACTION METHOD

Metode ekstraksi dalam web scraping dapat diklasifikasikan sebagai: Metode berbasis DOM dan berbasis string. metode berbasis DOM didasarkan pada pembuatan pohon DOM untuk halaman web dan pencarian elemen tertentu di pohon ini. Metode berbasis string adalah berfokus pada mengekstraksi konten dari CH secara langsung. Di browser web, pengembang web dapat mengakses, menambah, mengubah dan hapus semua elemen di halaman web dengan menggunakan metode Berbasis JavaScript pada pohon DOM. Pengurai pemrograman bahasa dalam tugas ini biasanya dibangun di atas struktur ini. Untuk contoh, parser ini memiliki metode serupa yang membuat Pohon DOM dan akses elemen pohon ini. Ekstraksi metode yang paling banyak digunakan untuk melakukan pengikisan web dengan melintasi pohon DOM.



GAMBAR 1. Dom Tree objek

Dalam JavaScript, ada beberapa metode termasuk `getElementsByTagName`, `getElementById`, `getElementsByClassName`, dan `getElementsByName` untuk melintasi pohon DOM dan mengekstraksi konten (Ce)

menggunakan tag (te) atau atribut dari id, kelas, dan nama (Ae) di OT, masing-masing. Namun, metode ini tidak sesuai untuk mengakses atribut barang prop. metode `querySelector()` mengembalikan elemen pertama yang cocok dengan pemilih CSS tertentu dalam dokumen HTML. Misalnya, `querySelector("[itemprop D headline]")` metode dapat digunakan untuk mengakses elemen. Dalam penelitian ini, `MS_HTMLDocument` in .Net memiliki mirip dengan metode JavaScript karena itu dipilih untuk dibandingkan dengan metode lain. Meskipun metode tersebut tersedia untuk melintasi pohon DOM, W3C, yang menetapkan standar Web resmi, telah memperkenalkan XPath standar baru untuk menavigasi melalui elemen dan atribut di pohon DOM. XPath menggunakan ekspresi jalur untuk mengakses sebuah elemen. Ekspresi jalur ini sangat mirip ekspresi jalur dalam sistem file komputer. Untuk memilih elemen, ekspresi XPath adalah `'/html/body/div/div/h1[@itemprop D"judul"]'`. XPath ekspresi juga dapat digunakan dalam JavaScript, Java, Python, Ruby, C# dan banyak bahasa lainnya. Dalam studi ini, parser .Net HAP2 lebih disukai untuk mengimplementasikan ekspresi XPath. Beberapa bahasa pemrograman memiliki fasilitas kueri umum untuk diterapkan ke semua sumber informasi. Di .Net, LINQ (Language-Integrated Query) mendukung fasilitas ini. Dalam penelitian ini, parser3 AngleSharp yang memungkinkan penggunaan LINQ dipilih. Tiga metode yang dijelaskan di atas didasarkan pada pohon DOM yang merupakan struktur yang nyaman untuk memanipulasi semua elemen. Namun, seorang programmer ahli hanya membutuhkan sejumlah elemen untuk web scraping. Dalam situasi ini, ekspresi reguler, teknik pemrosesan teks yang terkenal, dapat dianggap sebagai solusi untuk mendapatkan konten elemen tertentu. Misalnya, ekspresi berikut dapat digunakan untuk ekstraksiV:

- `<h1.itemprop D.headline.>(.*?)</h1>`
- `<h2.itemprop D.Summary.>(.*?)</h2>`
- `<p>(.*?)</p>`

Bentuk elemen, dan cocok untuk a bahasa biasa yaitu bahasa formal [9]. Bahasa reguler yang dapat diekspresikan menggunakan ekspresi reguler. Namun, bentuk elemen tidak sesuai untuk digunakan ekspresi reguler karena elemen ini dapat diwakili 2HAP (HTML Agility Pack) adalah parser HTML tangkas yang membangun DOM pohon dan memungkinkan untuk mengurai Pohon DOM

ini. 3Perpustakaan pengurai kurung sudut pamungkas (AngleSharp) membangun a Pohon DOM berdasarkan spesifikasi resmi W3C, dalam bentuk bersarang. Artinya, bentuk elemen adalah bebas konteks karena jumlah elemen dalam adalah tidak diketahui. Jika jumlah elemen dalam diketahui, masalah dapat diekspresikan dengan ekspresi reguler. Selain itu, kompleksitas waktu dapat ditingkatkan karena bahasa reguler dapat dikenali oleh mesin status akhir dari m-status di mana m adalah bilangan terhingga. Ini berarti membutuhkan $O(m)$ ruang dan $O(n)$ waktu, di mana n adalah panjang data. Di sisi lain, bahasa bebas konteks dapat dikenali oleh seperti robot pushdown non-deterministik algoritma CYK [10]. Kompleksitas pada waktu dari algoritma ini adalah $O(n^3 G)$, di mana n adalah panjang data dan jumlah aturan. Penelitian ini bertujuan untuk mencoba meningkatkan efisiensi waktu web scraping dengan mengubahnya ke tingkat bahasa biasa.

pendekatan menggunakan informasi tambahan yang diperoleh dari proses merangkak.

5. APPROACH OVERVIEW

Langkah pertama dalam mengekstrak konten web dari halaman web menggunakan metode DOM. Modul itu sendiri adalah file yang berisi tentang kode pemrograman python. Modul digunakan untuk memecah program besar menjadi file yang lebih kecil untuk manajemen dan organisasi yang lebih mudah. Modul membuat kode dapat digunakan kembali, artinya satu modul dapat digunakan kembali di mana pun diperlukan. Langkah pertama adalah mengimport modul python ke dalam program yang dibuat. Jadi dengan mengimpor modul, definisi, variabel, fungsi dan lain-lain yang ada di dalam modul kita bisa menggunakannya. Dalam penelitian ini dapat dimulai dengan mengimport modul yang dibutuhkan terlebih dahulu (baris 1-6) sehingga lebih mudah untuk melakukan proses scraping. Lalu kita bisa memasukkan chromedriver dimana chromedriver sendiri adalah server mandiri yang menerapkan standar webdriver W3C. WebDriver sendiri merupakan tool open source untuk pengujian otomatis, dengan kemampuan seperti navigasi, input pengguna, javascript dan lain-lain (baris 8), kemudian bisa menggunakan `driver.get` untuk menuju link yang ingin digunakan dimana `get` berfungsi untuk mendapatkan (baris 10). Untuk memaksimalkan tampilan web yang sedang di

scraping di windows bisa menggunakan `maximize_window()` (baris 11) sehingga tampilan web saat di running/dijalankan akan maksimal seluruh layar. `WebDriverWait` dapat digunakan untuk menunggu waktu loading saat pemrosesan menunggu di website arweave untuk memuat semua aplikasi yang ada di dalam web (baris 13). Kemudian kita bisa menyesuaikan bagian yang ingin di scraping dengan element-element yang ada pada website arweave dengan menggunakan ekstraksi dari pohon DOM dengan mengambil dari struktur HTML. Setelah proses scraping selesai maka kita bisa mengubah outputannya menjadi file csv.

[illegible]

6. EXPERIMENTS

Di bagian ini, pertama-tama diberikan informasi tentang kumpulan data dan pola ekstrasinya yang akan dibahas pada penelitian kali ini. Dimana hasil dari scrape web arweave tersebut isinya bisa kita lihat berupa data yang berisi tentang App Name, App Link, dan App Category dengan data yang diambil website tersebut.

Kemudian kita dapat melakukan analisis data outputan hasil scrape dengan menggunakan jupyter notebook dengan pandas untuk melakukan analisis deskriptif untuk dapat membaca file temp.csv dengan melalui pandas untuk melakukan analisis data sederhana.

	App Name	App Link	App Category
0	Pomodoro Timer	https://mtfvzwnv2pwykoicvxpoe7ao5rp4nhaueueux2...	development
1	dgjit	https://mtfvzwnv2pwykoicvxpoe7ao5rp4nhaueueux2...	development
2	OpenBits	https://mtfvzwnv2pwykoicvxpoe7ao5rp4nhaueueux2...	development
3	Ar Inspector	https://mtfvzwnv2pwykoicvxpoe7ao5rp4nhaueueux2...	development
4	ArQL Explorer	https://mtfvzwnv2pwykoicvxpoe7ao5rp4nhaueueux2...	development
5	Gitstamp	https://mtfvzwnv2pwykoicvxpoe7ao5rp4nhaueueux2...	development
6	WeaveID	https://mtfvzwnv2pwykoicvxpoe7ao5rp4nhaueueux2...	development
7	ArweaveApps	https://mtfvzwnv2pwykoicvxpoe7ao5rp4nhaueueux2...	development
8	Arweave Design	https://mtfvzwnv2pwykoicvxpoe7ao5rp4nhaueueux2...	development
9	JSON Editor	https://mtfvzwnv2pwykoicvxpoe7ao5rp4nhaueueux2...	development

Dari data diatas dapat kita lihat hasil preview 10 *record* teratas dari web dengan tampilan file csv hasil scrape pada website arweave sebagai berikut.

[illegible]

```
In [5]: df.columns
```

```
Out[5]: Index(['App Name', 'App Link', 'App Category'], dtype='object')
```

Dari data diatas dapat kita lihat bahwa data hasil scrape terdapat 3 kolom dimana berisi tentang App Name, App Link, dan App Category dengan type data elements ialah dtype='object'

```
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 235 entries, 0 to 234
Data columns (total 3 columns):
#   Column          Non-Null Count  Dtype
---  -
0   App Name        235 non-null    object
1   App Link        235 non-null    object
2   App Category    235 non-null    object
dtypes: object(3)
memory usage: 5.6+ KB
```

Dari data di atas kita dapat melihat informasi lebih detail dari data berdasarkan nama kolom dan jumlah record-nya sehingga kita dapat mengetahui bahwa ada 3 object yaitu App Name, App Link, dan App Category dengan memori usage sebesar 5.6+ KB

```
df.describe()
```

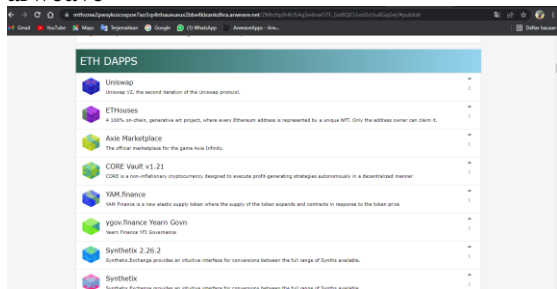
App Name		App Link	App Category
count	235	235	235
unique	230	230	14
top	Uniswap	https://mtfwzmv2pwykoicvxpoe7ao5rp4nhaueuex2...	ETH dapps
freq	3	2	54

Dari data di atas kita dapat melihat deskripsi data seperti jumlah data, data yang unik, frekuensi dan sebagainya sehingga kita bisa mengetahui secara detail jumlah data di dalam file. Dan lebih memudahkan kita untuk mengetahui bahwa ada 235 data dengan 230 merupakan data unik pada app name, dan juga ada 235 data dengan 230 data unik pada app link, begitupun dengan app category dengan 235 jumlah data seluruhnya dan 14 data unik.

```
df['App Category'].value_counts()
```

```
ETH dapps      54
social         32
media          31
storage        25
finance        23
games          20
development    12
identity        7
security        7
EOS dapps       7
stores          5
health          4
gambling        3
wallet          1
Name: App Category, dtype: int64
```

Sehingga dari data yang telah kita scrape dan lihat dari jenis categorynya ada ETH dapps dengan 54 data, social 32 data begitupun seterusnya yang dapat kita lihat pada website arweave



GAMBAR 6. Tampilan Website Arweave

7. KESIMPULAN

Sebagian besar studi sebelumnya tentang *web scraping* adalah tentang mekestrasi konten web secara otomatis dengan mengabaikan update terbaru dari website tersebut. Studi ini menggunakan metode ekstrasi tradisional berdasarkan pohon *DOM* yang memperhitungkan semua elemen dalam halaman web. Penelitian ini telah berhasil mengekstrasi dan melakukan analisis data aplikasi pada *web arweave* untuk memberikan informasi perkembangan aplikasi yang ada pada website arweave. Teknik *web scraping* dengan pustaka selenium dan csv dapat digunakan untuk mengekstrasi data aplikasi yang ada pada *web arweave*. Pada penelitian kali ini dari jenis category aplikasi yang ada pada *web arweave* dapat di bagi menjadi 14 category yang bisa bertambah sewaktu-waktu sesuai dengan publish app yang nantinya akan di tambahkan oleh penegmbang.

Referensi:

- [1] Lusiana Citra Dewi, M. C.: Social Media Web Scraping using Social Media Developers API and Regex. ELSEVIER, 445, 2019.
- [2] Sajati, G. P.: WEB SCRAPING DENGAN MENGGUNAKAN PYTHON. *ResearchGate*, 1-4, 2020.
- [3] Haddaway, N. R.: The Use of Web-scraping Software in. *ResearchGate*, 187, 2019.
- [4] Maulidiansyah, Z. A: Implementasi Analisa Semantik Pada Webdriver Untuk Otomatisasi Scraping. *explore it*, 81, 2020.
- [5] A. Priyanto and M. R. Ma'arif.: S Implementasi Web Scrapping dan Text Mining untuk Akuisisi danKategorisasi Informasi dari Internet (Studi Kasus: Tutorial Hidroponik), Indones. J. Inf. Syst., vol. 1, no. 1, pp. 25–33, 2018.
- [6] Sahria, Y.: Implementasi Teknik Web Scraping pada Jurnal SINTA Untuk Analisis Topik Penelitian Kesehatan Indonesia. *URECOL*, 296-298.
- [7] S. Mahajan and N. Kumar, “A Web Scraping Approach in Node.js,” vol. 4, no. April, pp. 909–912, 2015.
- [8] N. Chomsky, “Three models for the description of language,” *IEEE Trans. Inf. Theory*, vol. IT-2, no. 3, pp. 113–124, Sep. 1956.
- [9] M. Sipser, “Introduction to the theory of computation,” *ACM Sigact News*, vol. 27, no. 1, pp. 27–29, 1996
- [10] M. Joko U. H.B, Hardan G, “SCRAPING WEB MARKETPLACE MENGGUNAKAN METODE DOM PARSING UNTUK PENGUMPULAN DATA PRODUK,” . 2020
- [11]E. Uzun, T. Yerlikaya, and M. Kurt,

“A lightweight parser for extracting useful contents from Web pages,” in Proc. 2nd Int. Symp. Comput. Sci. Eng. (ISCSE), Kuşadasi, Turkey, 2011, pp. 67–73.

- [12]P. W. Geoff Boeing, "New Insights into Rental Housing Markets across the United States: Web Scraping and Analyzing Craigslist Rental Listings," Journal of Planning Education and Research , 2016.