

OPTIMASI METODE *SCRAPING* DATA PRODUK DARI *PLATFORM* TOKOPEDIA.COM

Muhammad Daffa Al Farizi¹, Rahmad Hidayat^{2*}, Musta'inul Abdi

^{1,2,3} Teknik Informatika, Teknologi Informasi dan Komputer, Politeknik Negeri Lhokseumawe, Kota Lhokseumawe, Indonesia

Email: ¹daffaalfarizi1945@gmail.com, ^{2*}rahmad_hidayat@pnl.ac.id, ³mustainul.abdi@pnl.ac.id

Abstrak- Tingkat keefisienan dalam pengumpulan data produk dari *platform e-commerce*, seperti Tokopedia.com, telah menjadi tantangan karena memiliki tampilan dan data yang dinamis. Dalam upaya untuk mengatasi kendala ini, penelitian ini bertujuan untuk mengoptimalkan metode *scraping* data produk pada Tokopedia.com, dengan fokus pada pencarian yang cepat dan efisien. Dalam konteks *platform e-commerce* yang dinamis seperti Tokopedia.com, dimana jumlah produk terus meningkat, metode pengumpulan data manual telah terbukti tidak efektif dan memakan waktu. Oleh karena itu, penelitian ini mengusulkan solusi melalui penggunaan metode *scraping* yang dioptimalkan dengan cara memanfaatkan dari kecepatan request http dan memanfaatkan pergerakan yang dinamis menggunakan ilmu dari *Robotic Process Automation* (RPA). Dari hasil penelitian metode *Asynchronous Playwright HTTPX* lebih efektif untuk digunakan berdasarkan jumlah data, durasi waktu, dan penggunaan sumber daya, Metode HTTPX berhasil mengambil jumlah data (750 data), durasi waktunya (20 menit 34 detik), Sedangkan Metode *Asynchronous Playwright HTTPX* mencapai hasil dalam jumlah data yang sangat besar (8228 data) dengan durasi waktu (21 menit 7 detik). Berdasarkan penggunaan sumber daya Metode HTTPX memerlukan penggunaan sumber daya yang rendah, termasuk tingkat penggunaan CPU (2,1%), RAM (83,0%), dan Paket Internet (108.34 MB) yang rendah, Sedangkan Metode *Asynchronous Playwright HTTPX* tetap efisien dalam penggunaan sumber daya, dengan penggunaan CPU (29,2%), RAM (82,8%) dan Paket Internet (1,19 GB).

Kata Kunci: Pengumpulan Data, *Scraping*, Data Produk, Optimasi, *Robotic Process Automation* (RPA).

OPTIMIZATION OF DATA *SCRAPING* METHODS FOR PRODUCTS FROM TOKOPEDIA.COM PLATFORM

Abstract- Efficiency in collecting product data from *e-commerce* platforms, such as Tokopedia.com, has emerged as a challenge due to their dynamic interface and data structure. In an attempt to address this challenge, this study aims to optimize the method of *scraping* product data from Tokopedia.com, focusing on achieving rapid and efficient data retrieval. Within the context of the dynamic *e-commerce* platform like Tokopedia.com, where the product count continues to grow, manual data collection methods have proven to be inefficient and time-consuming. Consequently, this research proposes a solution through the implementation of an optimized *scraping* method that leverages the speed of HTTP requests and exploits dynamic behaviors using the principles of *Robotic Process Automation* (RPA). Based on the research findings, the *Asynchronous Playwright HTTPX* method demonstrates superior effectiveness in terms of data volume, time duration, and resource utilization. The HTTPX method successfully retrieves a data volume of 750 entries within a time frame of 20 minutes and 34 seconds. On the other hand, the *Asynchronous Playwright HTTPX* method achieves substantial results in terms of data volume (8228 entries) within a slightly longer time frame of 21 minutes and 7 seconds. In the aspect of resource consumption, the HTTPX method requires minimal resources, including low CPU utilization (2.1%), significant RAM consumption (83.0%), and minimal Internet data usage (108.34 MB). In contrast, the *Asynchronous Playwright HTTPX* method maintains efficiency in resource utilization, with CPU usage (29.2%), RAM consumption (82.8%), and Internet data usage (1.19 GB).

Keywords: Data Collection, *Scraping*, Product Data, Optimization, *Robotic Process Automation* (RPA).

1. PENDAHULUAN

Dalam era digital yang serba cepat dan kompetitif ini, akses terhadap data yang akurat dan komprehensif menjadi krusial bagi pelaku bisnis dan analis untuk mengambil keputusan strategis yang tepat. Salah satu sumber data yang sangat berharga adalah *platform e-commerce*, khususnya Tokopedia.com, Tokopedia merupakan *platform* perdagangan daring pertama di Indonesia yang membuka peluang bisnis secara online [1]. Tokopedia menyediakan berbagai produk dari ribuan penjual. Namun, pengumpulan data dari Tokopedia.com melalui metode *scraping* masih menghadapi berbagai kendala yang mempengaruhi efisiensi dan ketepatan waktu dalam pengambilan data. Oleh karena itu, diperlukan sebuah penelitian yang fokus pada optimasi metode *scraping* untuk memungkinkan pengambilan data produk sebanyak mungkin dalam durasi waktu sesingkat mungkin. *scraping* merupakan sebuah teknik untuk secara otomatis mendapatkan informasi dari situs web tanpa perlu melakukan

penyalinan manual [2]. Proses *parsing HTML* mencakup ekstraksi dan pengolahan informasi yang relevan, seperti judul halaman, aset halaman, dan bagian utama, selanjutnya, hasil tersebut dapat disimpan dalam file yang telah diproses [3].

Jumlah produk yang tersedia di Tokopedia.com terus bertambah dan berubah secara dinamis seiring dengan adanya penambahan penjual baru, perubahan harga, serta produk-produk terbaru yang diluncurkan. Dalam konteks ini, waktu menjadi faktor kritis dalam pengumpulan data produk. Semakin lama waktu yang dibutuhkan untuk mengumpulkan data, semakin besar kemungkinan sejumlah produk telah berubah status atau tidak lagi tersedia, sehingga mengakibatkan informasi yang didapatkan tidak lagi akurat atau relevan. Oleh karena itu, optimasi metode *scraping* menjadi penting untuk memastikan bahwa data yang diambil selalu mutakhir dan mencakup sebanyak mungkin produk dalam interval waktu yang singkat. *Asynchronous programming* adalah cara pemrograman *paralel*. Ini beroperasi dengan eksekusi acak dari beberapa proses. CPU dapat mengalokasikan pekerjaan baru meskipun pekerjaan pertama belum selesai, dan seterusnya[4]. Pemrosesan paralel, program akan berjalan dengan cepat karena dapat memanfaatkan lebih banyak sumber daya (dalam hal ini, komputer)[5].

Selain itu, pentingnya pengumpulan data yang cepat dan akurat juga terkait dengan persaingan bisnis yang ketat di *platform e-commerce*. Pelaku bisnis yang mampu mengakses dan menganalisis data produk dengan cepat akan memiliki keunggulan kompetitif dalam mengidentifikasi tren pasar, memahami preferensi pelanggan, dan menyesuaikan strategi pemasaran. Dengan metode *scraping* yang dioptimasi, para pelaku bisnis dapat dengan mudah mengakses informasi produk yang relevan dan *up-to-date* untuk mendukung pengambilan keputusan yang tepat waktu. *Playwright* adalah kemampuannya untuk berinteraksi dengan berbagai browser seperti *Chromium*, *Firefox*, dan *WebKit*. Ini berarti pengguna dapat menguji aplikasi web di berbagai browser yang berbeda dengan menggunakan satu set kode[6]. *RPA* merupakan teknologi otomatisasi proses bisnis yang bertujuan untuk melakukan otomatisasi interaksi manusia dengan desktop atau computer[7]. *Python* merupakan salah satu bahasa pemrograman yang sering digunakan oleh perusahaan besar maupun para pengembang untuk menciptakan berbagai jenis aplikasi berbasis desktop, web, dan mobile[8].

Pentingnya optimasi metode *scraping* juga berkaitan dengan aspek efisiensi dan efektivitas pengumpulan data. Metode *scraping* yang kurang dioptimasi dapat mengakibatkan beban server yang berlebihan, menyebabkan gangguan dalam operasional situs web, dan bahkan dapat menyebabkan blokir oleh pihak *platform*. Dengan demikian, penelitian mengenai optimasi metode *scraping* ini tidak hanya berfokus pada kuantitas data yang diambil, tetapi juga pada cara yang efisien dan etis dalam melakukan pengumpulan data tanpa mengganggu operasional situs *e-commerce* yang bersangkutan. *Web scraping* menggunakan *library BeautifulSoup* dengan *html5lib* sebagai *library parser*, *BeautifulSoup* dapat diimplementasikan dengan beberapa bahasa pemrograman, salah satunya adalah *Python*[9].

Penelitian ini membedakan dari penelitian-penelitian sebelumnya adalah Penelitian ini berusaha meningkatkan efisiensi dan performa dalam proses *scraping* data produk dari Tokopedia.com. Hal ini bertujuan untuk mengoptimalkan pengambilan data dalam waktu yang lebih singkat dan dengan sumber daya yang lebih efisien. Penelitian ini menitikberatkan pada optimasi metode *scraping* data produk dari Tokopedia.com. Optimasi ini dilakukan berdasarkan keyword produk yang diinputkan oleh pengguna serta jumlah halaman yang diinputkan, sehingga pengguna memiliki kendali lebih besar terhadap data yang ingin diambil. Penelitian ini memfokuskan pada perbandingan berbagai metode *scraping* yang berbeda untuk mengumpulkan data produk dari Tokopedia.com. Ini membantu mengidentifikasi metode mana yang paling efektif dan cocok untuk skenario tertentu. Penelitian ini merinci hasil perbandingan dan evaluasi dari metode-metode *scraping* yang digunakan. Aspek efisiensi, akurasi, dan keterbatasan dari masing-masing metode dianalisis secara mendalam. Penelitian ini mencakup pengambilan data dari berbagai jenis halaman, seperti halaman hasil pencarian, halaman detail produk, dan halaman toko di Tokopedia.com. Hal ini memastikan data yang diambil lebih representatif. Penelitian ini menggabungkan teknik *Robotic Process Automation (RPA)* seperti *Selenium* dan *Playwright* untuk mengotomatisasi interaksi dengan situs, serta menggunakan *HTTPX* untuk efisiensi dalam melakukan permintaan *HTTP* pada halaman-halaman yang dituju.

Diharapkan bahwa hasil dari penelitian ini akan memberikan kontribusi signifikan dalam pengembangan teknik *scraping* yang dioptimasi untuk *platform Tokopedia.com*, serta dapat diadaptasi dan diterapkan pada *platform e-commerce* lainnya. Keberhasilan dalam optimasi metode *scraping* ini akan membuka peluang baru bagi para pelaku bisnis, peneliti, dan analis untuk mengakses data produk secara cepat dan menyeluruh, sehingga mendukung perkembangan industri *e-commerce* di Indonesia. Selain itu, penelitian ini juga dapat menjadi acuan penting bagi studi-studi mendatang yang ingin mengeksplorasi dan mengatasi tantangan dalam pengumpulan data dari *platform-platform e-commerce* pada skala yang lebih luas. Perlu diingat bahwa keputusan ini tidak memberikan izin bagi *scraping* untuk menggunakan data yang diperoleh untuk tujuan komersial [10].

2. METODE PENELITIAN

2.1 Metode Perancangan Sistem

Berisi penjelasan tentang tahapan penelitian yang menggambarkan urutan logis untuk mendapatkan hasil penelitian sesuai dengan harapan dan gambaran sistem. Jika ada gambar dan tabel, itu harus disajikan dengan nama tabel dan gambar yang disertai dengan nomor urut. Berikut adalah gambaran umum perancangan sistem penelitian dapat dilihat pada Gambar 1 berikut



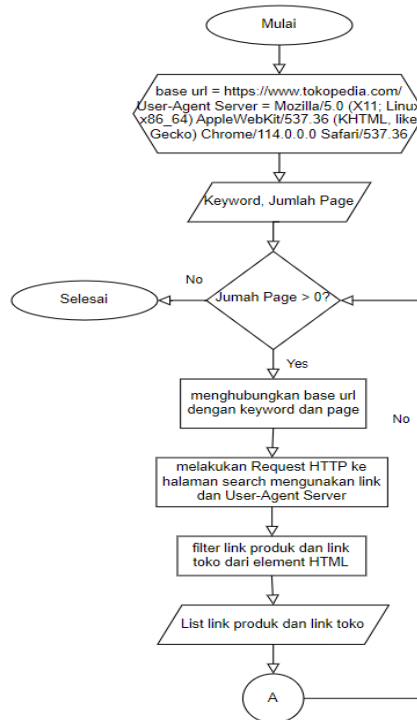
Gambar 1. Bagan Tahapan Penelitian

1. Analisis Kebutuhan:
 - a. Identifikasi kebutuhan data: Data produk yang akan diambil dari Tokopedia.com adalah nama produk, harga, deskripsi, kategori, dan lain-lain.
 - b. Tentukan skala *scraping*: halaman produk yang perlu diambil adalah halaman search produk, halaman detail produk, dan halaman toko.
2. Pemilihan Teknologi:
 - a. Bahasa pemrograman: Bahasa pemrograman yang digunakan untuk mengimplementasikan metode *scraping* pada penelitian ini yaitu python.*Library* dan Framework: Identifikasi *library* atau framework yang mendukung proses *scraping* data dari situs web, yaitu Fastapi, BeautifulSoup, HTTPX, Playwright dan Selenium.
3. Pengembangan Metode *Scraping*:
 - a. Lakukan studi literatur: Teknik *scraping* yang telah ada yaitu melakukan request http ke server kemudian menyeleksi *response* berupa elemen html dari server dan teknologi terkini dalam domain *scraping* data yaitu dengan mengimplementasikan cabang ilmu Robotic Process Automation (RPA).
 - b. Identifikasi kendala dan hambatan: Kendala yang mungkin dihadapi selama *scraping* dari Tokopedia.com yaitu halaman Tokopedia.com yang dinamis, *response* dari server yang terbatas, dan kebijakan anti-*scraping*.
4. Implementasi:
 - a. Buat skrip *scraping*: Implementasikan metode *scraping* yang telah dioptimasi menggunakan bahasa pemrograman python dan *library* yang dipilih adalah Playwright dan Selenium.
 - b. Uji coba: Lakukan uji coba untuk memastikan skrip berjalan dengan benar dan mampu mengambil data sesuai kebutuhan.
5. Validasi dan Evaluasi:
 - a. Validasi data: Verifikasi data yang diambil dari Tokopedia.com.
 - b. Evaluasi performa: Ukur performa metode *scraping* dengan membandingkan performa dari cpu, ram, dan paket internet yang digunakan saat proses *scraping* dilakukan.
 - c. Evaluasi hasil *scraping*: Bandingkan durasi *scraping* dan jumlah data berdasarkan metode *scraping*, *keyword*, dan jumlah *page* yang digunakan saat *scraping*.
6. Penyusunan Dokumentasi:
 - a. Dokumentasikan metode *scraping* yang telah dioptimasi dengan rinci.
 - b. Catat hasil evaluasi dan uji coba secara terperinci.
7. Simpulan:
 - a. Analisis hasil: Berdasarkan validasi data, evaluasi performa, dan evaluasi hasil *scraping*.

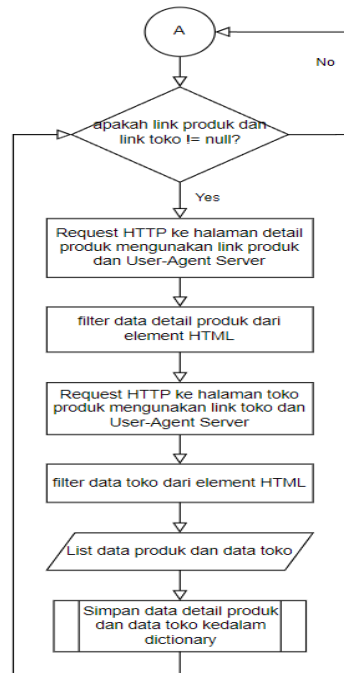
2.2 Desain Sistem

A. Metode HTTPX

Gambar 2 dan Gambar 3 menampilkan flowchart yang menggambarkan metode *scraping* HTTPX.



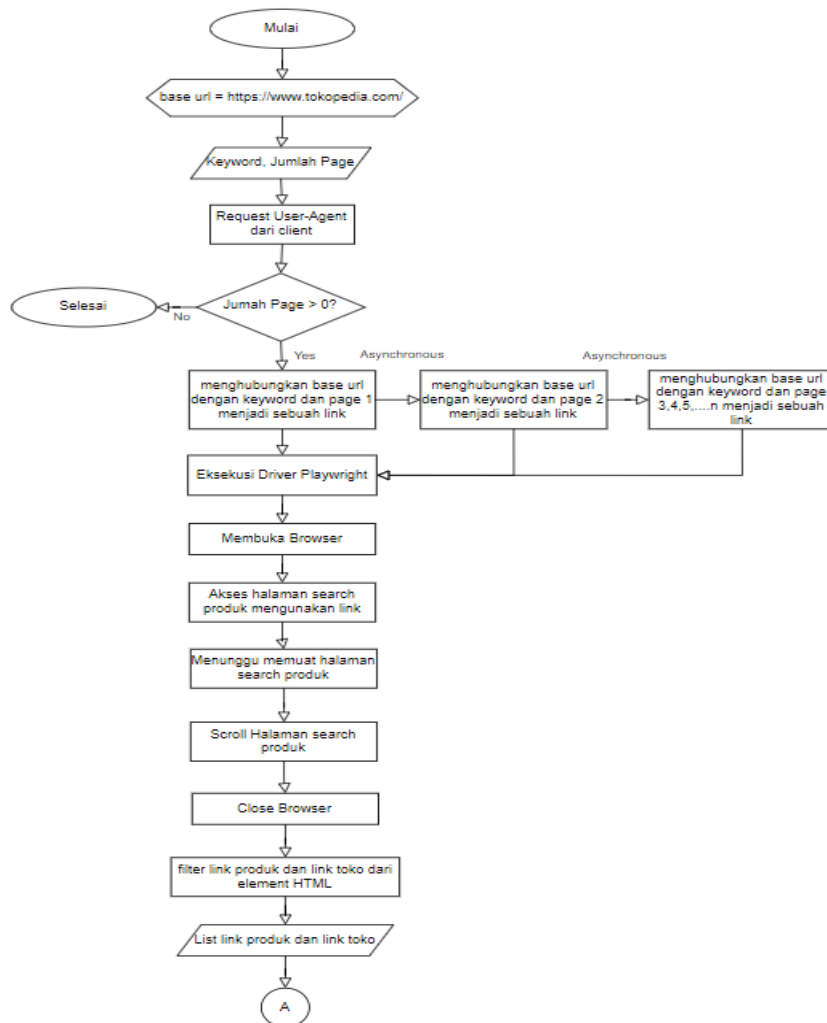
Gambar 2. Flowchart HTTPX Bagian 1



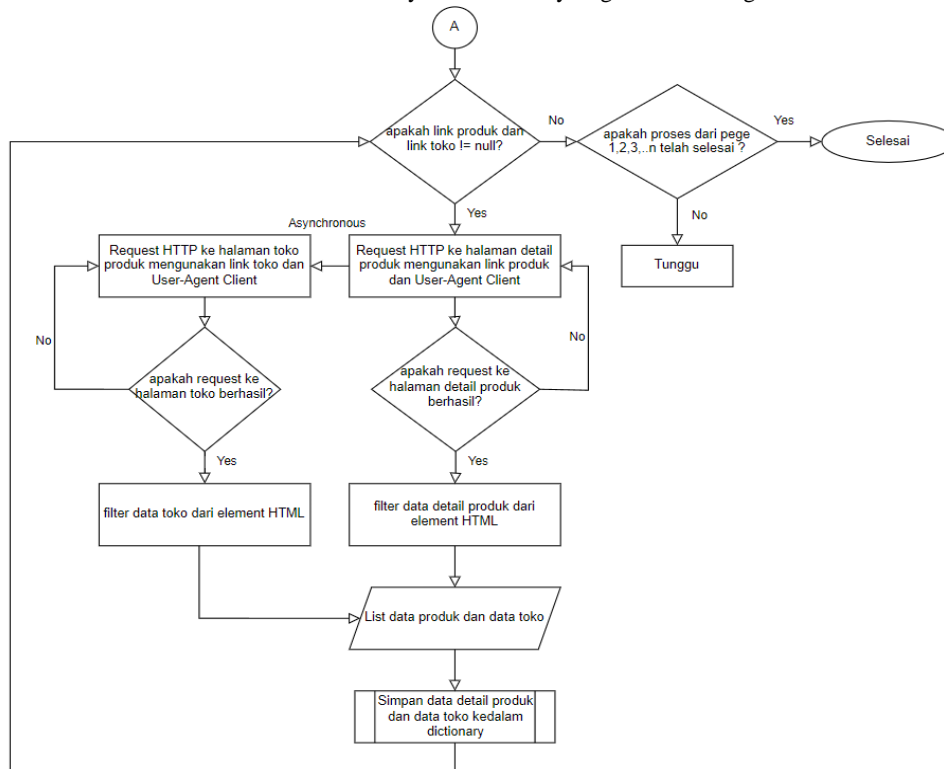
Gambar 3. Flochart HTTPX Bagian 2

B. Metode *Asynchronous Playwright* HTTPX

Metode ini merupakan kombinasi antara kekuatan *Asynchronous* dari Playwright dan HTTPX. Dalam pendekatan ini, Playwright digunakan untuk *scraping* halaman search produk dari platform Tokopedia.com, sementara HTTPX digunakan untuk melakukan permintaan HTTP ke halaman detail produk dan halaman toko yang bersifat *Asynchronous*. Gambar 4 dan Gambar 5 menampilkan flowchart yang menggambarkan metode *scraping Asynchronous Playwright* HTTPX.



Gambar 4. Flowchart Asynchronous Playwright HTTPX Bagian 1



Gambar 5. Flowchart Asynchronous Playwright HTTPX Bagian 2

3. HASIL DAN PEMBAHASAN

Berikut adalah penjelasan lebih rinci tentang perbedaan, kelebihan, dan kekurangan masing-masing metode:

Metode HTTPX:

Metode HTTPX melakukan *scraping* secara sinkron, yang berarti permintaan ke server dan penerimaan respons dilakukan secara berurutan. Setelah satu permintaan selesai, barulah permintaan berikutnya dieksekusi.

- a. Kelebihan
 1. Lebih sederhana dan mudah diimplementasikan.
 2. Cocok untuk pengambilan data dalam skala kecil hingga menengah.
 3. Penggunaan sumber daya (CPU dan RAM) lebih terkontrol dan rendah.
 4. Cocok untuk kasus-kasus di mana waktu eksekusi tidak menjadi faktor kritis.
- b. Kekurangan:
 1. Lebih lambat dalam pengambilan data dalam jumlah besar atau dalam situasi yang membutuhkan banyak permintaan.
 2. Tidak efisien dalam pengambilan banyak data dalam waktu singkat.

Metode *Asynchronous Playwright HTTPX*:

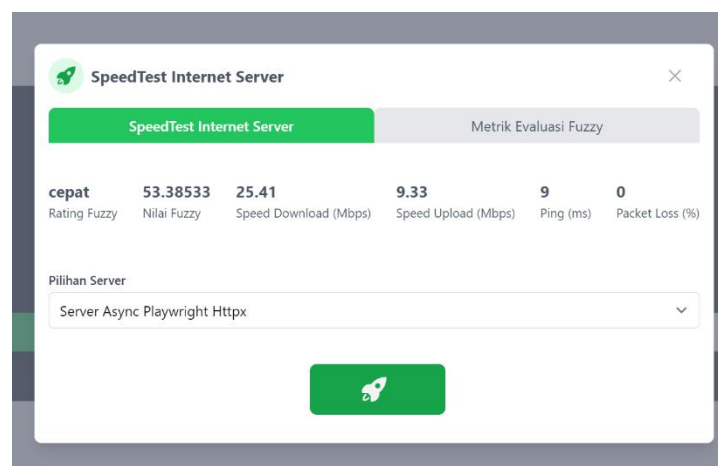
Metode ini menggunakan pendekatan asinkron, yang memungkinkan permintaan-permintaan yang berbeda dieksekusi secara bersamaan tanpa harus menunggu respons sebelumnya selesai.

- a. Kelebihan:
 1. Cepat dalam pengambilan banyak data sekaligus dalam waktu yang relatif singkat.
 2. Cocok untuk kasus-kasus di mana waktu eksekusi dan jumlah data sangat penting.
 3. Dapat memanfaatkan sumber daya yang lebih optimal dalam pengambilan data dalam skala besar.
 4. Cocok untuk kasus-kasus yang membutuhkan *scraping* di halaman web yang berbeda secara bersamaan.
- c. Kekurangan:
 1. Penggunaan sumber daya (CPU dan RAM) lebih tinggi.
 2. Lebih kompleks untuk diimplementasikan dan memerlukan pemahaman tentang konsep asinkron.

3.1 Implementasi *User Interface*

Test Speed Internet Server

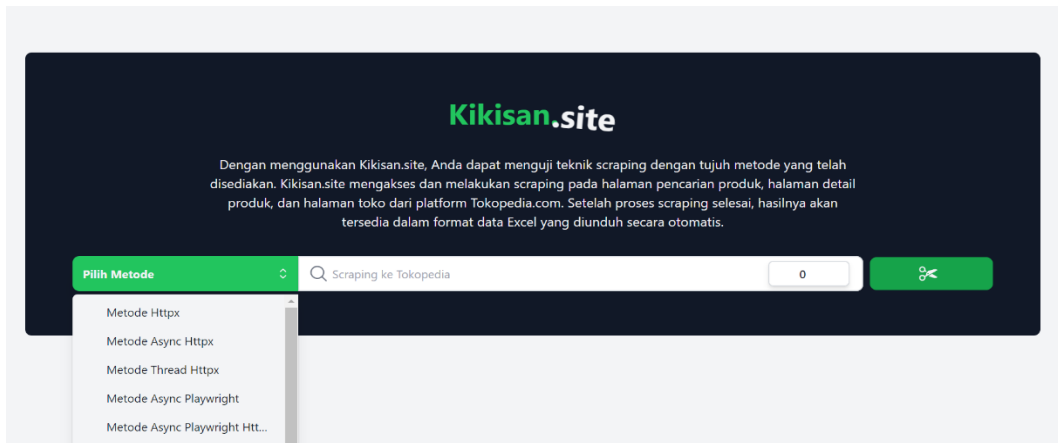
Gambar 6 adalah tampilan antarmuka pengguna (UI) dari Test Speed Internet Server yang bertujuan untuk memberikan pengguna kemampuan untuk menguji kecepatan koneksi internet mereka ke server tertentu. Melalui antarmuka pengguna yang sederhana, pengguna dapat memulai tes dengan memilih server berdasarkan nama metode *scraping* dan mengklik tombol "Icon Roket". Setelah proses pengujian selesai, hasilnya akan ditampilkan dalam bentuk angka yang menggambarkan kecepatan unduh (*download*) dan unggah (*upload*) koneksi internet server *scraping* yang dipilih ke server Tokopedia.com.



Gambar 6. UI *Test Speed Internet Server*

Form Scraping

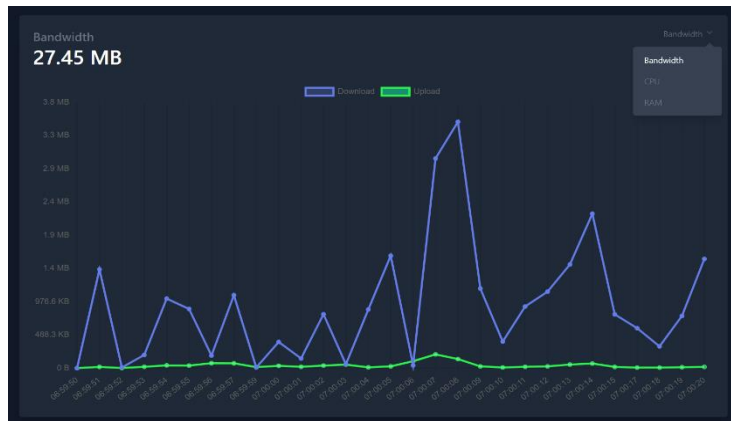
Gambar 7 adalah UI dari form *scraping*, form ini memudahkan pengguna untuk mengatur parameter dan menginisiasi proses *scraping* data produk. Dengan elemen-elemen seperti *select options* metode *scraping*, input *keyword*, input jumlah *page*, dan button *scraping*, pengguna memiliki kendali penuh dalam proses ini, memastikan bahwa data yang diambil sesuai dengan kebutuhan dan preferensi.



Gambar 7. Form Scraping

Grafik Monitoring

Pada gambar 8 menyajikan grafik yang memvisualisasikan data paket internet yang direkam saat proses *scraping* dilakukan, kemudian ditotalkan seluruh nilai paket internet upload dan *download* dari data grafik tersebut. Informasi ini membantu pengguna dalam melacak penggunaan paket internet dengan lebih efektif.



Gambar 8. Monitoring Paket Internet

Pada gambar 9 adalah grafik yang memberikan pengguna pemahaman tentang penggunaan CPU saat proses *scraping* dieksekusi dan menampilkan nilai rata-rata dari persentase CPU pada data grafik tersebut. Pengguna dapat dengan mudah memantau tingkat beban CPU sistem.



Gambar 9. Monitoring CPU

Pada gambar 10 adalah grafik persentase RAM memberikan visualisasi data RAM yang direkam saat proses *scraping* dilakukan dan menampilkan nilai rata-rata dari persentase RAM pada data grafik tersebut. Informasi ini membantu pengguna memantau penggunaan RAM dan memudahkan dalam menganalisis metode *scraping*.



Gambar 10. Monitoring RAM

Komponen Informasi

Pada gambar 11 memaparkan tentang informasi yang telah dirancang untuk memberikan pengguna akses terhadap detail penting terkait informasi hasil *scraping* dan informasi server. Komponen ini menyajikan berbagai informasi yang relevan untuk memberikan pemahaman lebih dalam.



Gambar 11. Komponen Informasi

3.2 Pengujian Sistem

Berikut Tabel 1 adalah pengujian yang dilakukan pada server 1 dengan sumber daya sebagai berikut :

Tabel 1. Pengujian Server 1

Metode	Keyword	Jumlah Page	Total Paket internet	Rata-Rata Persentase CPU	Rata-Rata Persentase RAM	Durasi Waktu	Jumlah Data
HTTPX	sepatu	1	1.43 MB	5.6%	92.0%	00:00:11	10
Asynchronous Playwright HTTPX	sepatu	1	19.19 MB	9.8%	74.9%	00:00:24	113
HTTPX	Smartphone Samsung terbaru	5	7.23 MB	6.1%	91.9%	00:00:53	50
Asynchronous Playwright HTTPX	sepatu	5	66.03 MB	19.0%	81.5%	00:00:57	396
HTTPX	tv	10	13.65 MB	6.7%	62.2%	00:01:33	100
Asynchronous Playwright HTTPX	tv	10	165.12 MB	30.0%	70.6%	00:02:42	1087
HTTPX	Laptop	100	108.34 MB	2.1%	83.0%	00:20:34	750
Asynchronous Playwright HTTPX	Laptop	100	1.19 GB	29.2%	82.8%	00:21:07	8228

Metode HTTPX:

- a. Pengujian dilakukan pada kata kunci "sepatu" dengan 1 halaman.
- b. Total paket internet yang digunakan adalah 1.43 MB.
- c. Rata-rata persentase CPU adalah 5.6%, dan rata-rata persentase RAM adalah 92.0%.
- d. Durasi waktu eksekusi adalah 00:00:11 dengan jumlah data sebesar 10.

Metode Asynchronous Playwright HTTPX:

- a. Pengujian dilakukan pada kata kunci "sepatu" dengan 1 halaman.
- b. Total paket internet yang digunakan adalah 19.19 MB.
- c. Rata-rata persentase CPU adalah 9.8%, dan rata-rata persentase RAM adalah 74.9%.
- d. Durasi waktu eksekusi adalah 00:00:24 dengan jumlah data sebesar 113.

Metode HTTPX:

- a. Pengujian dilakukan pada kata kunci "Smartphone Samsung terbaru" dengan 5 halaman.
- b. Total paket internet yang digunakan adalah 7.23 MB.
- c. Rata-rata persentase CPU adalah 6.1%, dan rata-rata persentase RAM adalah 91.9%.
- d. Durasi waktu eksekusi adalah 00:00:53 dengan jumlah data sebesar 50.

Metode Asynchronous Playwright HTTPX:

- a. Pengujian dilakukan pada kata kunci "sepatu" dengan 5 halaman.
- b. Total paket internet yang digunakan adalah 66.03 MB.
- c. Rata-rata persentase CPU adalah 19.0%, dan rata-rata persentase RAM adalah 81.5%.
- d. Durasi waktu eksekusi adalah 00:00:57 dengan jumlah data sebesar 396.

Metode HTTPX:

- a. Pengujian dilakukan pada kata kunci "tv" dengan 10 halaman.
- b. Total paket internet yang digunakan adalah 13.65 MB.
- c. Rata-rata persentase CPU adalah 6.7%, dan rata-rata persentase RAM adalah 62.2%.
- d. Durasi waktu eksekusi adalah 00:01:33 dengan jumlah data sebesar 100.

Metode Asynchronous Playwright HTTPX:

- a. Pengujian dilakukan pada kata kunci "tv" dengan 10 halaman.
- b. Total paket internet yang digunakan adalah 165.12 MB.
- c. Rata-rata persentase CPU adalah 30.0%, dan rata-rata persentase RAM adalah 70.6%.
- d. Durasi waktu eksekusi adalah 00:02:42 dengan jumlah data sebesar 1087.

Metode HTTPX:

- a. Pengujian dilakukan pada kata kunci "Laptop" dengan 100 halaman.
- b. Total paket internet yang digunakan adalah 108.34 MB.
- c. Rata-rata persentase CPU adalah 2.1%, dan rata-rata persentase RAM adalah 83.0%.
- d. Durasi waktu eksekusi adalah 00:20:34 dengan jumlah data sebesar 750.

Metode Asynchronous Playwright HTTPX:

- a. Pengujian dilakukan pada kata kunci "Laptop" dengan 100 halaman.
- b. Total paket internet yang digunakan adalah 1.19 GB.
- c. Rata-rata persentase CPU adalah 29.2%, dan rata-rata persentase RAM adalah 82.8%.
- d. Durasi waktu eksekusi adalah 00:21:07 dengan jumlah data sebesar 8228.

4. KESIMPULAN

Dari hasil penelitian Metode *Asynchronous Playwright HTTPX* lebih efektif untuk digunakan berdasarkan jumlah data, durasi waktu, dan penggunaan sumber daya, Metode HTTPX berhasil mengambil jumlah data (750 data), durasi waktunya (20 menit 34 detik), Sedangkan Metode *Asynchronous Playwright HTTPX* mencapai hasil dalam jumlah data yang sangat besar (8228 data) dengan durasi waktu (21 menit 7 detik). Berdasarkan penggunaan sumber daya Metode HTTPX memerlukan penggunaan sumber daya yang rendah, termasuk tingkat penggunaan CPU (2,1%), RAM (83,0%), dan Paket Internet (108.34 MB) yang rendah, Sedangkan Metode *Asynchronous Playwright HTTPX* tetap efisien dalam penggunaan sumber daya, dengan penggunaan CPU (29,2%), RAM (82,8%) dan Paket Internet (1,19 GB).

Metode *Asynchronous Playwright HTTPX* cenderung menggunakan lebih banyak sumber daya (CPU dan RAM) dibandingkan dengan metode HTTPX dalam hampir semua kasus. Meskipun demikian, *Asynchronous Playwright HTTPX* menghasilkan total paket internet yang lebih besar, yang mungkin disebabkan oleh kinerja yang lebih efisien dalam mengambil data dari situs web Tokopedia.com. Pengujian dengan kata kunci yang berbeda menghasilkan variasi performa. Beberapa kata kunci menghasilkan lebih banyak data yang di-*scrape*, mengakibatkan penggunaan sumber daya yang lebih tinggi dan durasi eksekusi yang lebih lama. Kecepatan internet dan spesifikasi server secara signifikan mempengaruhi performa web scraping. Metode dengan persentase CPU dan RAM tinggi cenderung menghabiskan waktu eksekusi yang lebih lama, namun juga menghasilkan total paket internet yang lebih besar.

DAFTAR PUSTAKA

- [1] F. R. Lupi and N. Nurdin, "Analisis Strategi Pemasaran Dan Penjualan E-commerce Pada Tokopedia.com," *Jurnal Elektronik Sistem Informasi dan Komputer*, vol. 2, no. 1, pp. 20–29, 2016.
- [2] H. S. P. and H. M. D. D. Ayani, "Implementasi Web Scraping untuk Pengambilan Data pada Situs Marketplace," *JUSTIN (Jurnal Sistem dan Teknologi Informasi)*, vol. VII, no. 4, pp. 257–262, 2019.
- [3] A. N. H. and D. A. A. M. Komarudin, "PENERAPAN TEKNIK WEB SCRAPING PADA SITUS IMDb DENGAN NODE JS," *Jurnal Siliwangi Seri Sains dan Teknologi*, vol. VIII, no. 2, pp. 55–59, 2022.
- [4] M. M. R. and M. R. Z. Ohidujjaman, "Clustering Algorithm with Asynchronous Programming," *American Journal of Engineering Research (AJER)*, vol. VI, no. 8, pp. 286–294, 2017.
- [5] D. T. and S. B. A. S. Yondra, "Implementasi Web Scraping Untuk Mengumpulkan Informasi Produk Dari Situs E-Commerce Dan Marketplace Dengan Teknik Pemrosesan Paralel," *Coding : Jurnal Komputer dan Aplikasi*, vol. X, no. 1, pp. 93–102, 2022.
- [6] M. A. D. M. M. B. T. S. and R. U. M. BANSAL, "Data Ingestion and Processing using Playwright," *Section A -Research paper*, vol. XII, no. 4, pp. 7241–7246, 2023.
- [7] D. A. Puryono and I. Sa'roni, "Penerapan Robotic Process Automation (RPA) Untuk Otomatisasi Proses Penilaian Pada Aplikasi Raport Digital Raudhatul Athfal," *Jurnal Informatika dan Rekayasa Perangkat Lunak*, vol. IV, no. 2, pp. 106–112, 2022.
- [8] M. Romzi and B. Kurniawan, "Pembelajaran Pemrograman Python Dengan Pendekatan Logika Algoritma," *JTIM: Jurnal Teknik Informatika Mahakarya*, vol. III, no. 2, pp. 37–44, 2020.
- [9] D. B. Pratama, A. Sofwan, and Y. A. A. Soetrisno, "Implementasi Teknik Web Scraping Dan Fitur Data Eksternal Pada Sistem Informasi Dosen Penelitian Dan Pengabdian Dosen Fakultas Teknik Universitas Diponegoro," *TRANSIEN*, vol. X, no. 2, pp. 292–299, 2021.
- [10] P. A. P. and L. J. V. A. Flores, "Penerapan Web Scraping Sebagai Media Pencarian dan Menyimpan Artikel Ilmiah Secara Otomatis Berdasarkan Keyword," *Majalah Ilmiah Teknologi Elektro*, vol. XIX, no. 2, pp. 157–162, 2020.