

IMPLEMENTASI DAN EVALUASI WEB SCRAPING PADA DATA JURNAL SINTA

Indra Nasution¹, Muhammad Iqbal²

^{1,2}Pascasarjana, Teknologi Informasi, Universitas Pembangunan Panca Budi

Email: ¹sutanindrabungsu@gmail.com, ²muhammadiqbalpb@gmail.com

Abstract: *The development of digital technology has increased the need for access to scientific information, especially related to academic journal publications. The Science and Technology Index (SINTA) website is one of the main platforms that provides information on accredited scientific journals in Indonesia. However, the process of searching and collecting journal data on the website is still done manually, so it is less efficient when the amount of data needed is large enough. This research aims to implement and evaluate web scraping techniques in collecting journal data on the SINTA website so that the data acquisition process can be carried out automatically, quickly, and in a structured manner. The research method uses an implementive approach through the stages of literature study, analysis of the structure of website pages, development of a system using the Python programming language with the BeautifulSoup library, data collection, database storage, and evaluation of extraction results. The results of the study show that the implementation of the web scraping technique succeeded in obtaining as many as 15,456 journal data from the SINTA website in 2026 which includes information on journal names, publishing institutions, accreditation categories, ISSN, fields of science, and journal website links. The developed system is also capable of storing data in a structured manner into a database and supports regular information updates. The results of the evaluation show that the web scraping technique is able to improve the efficiency of data collection, minimize manual processes, and produce more systematic journal data. Thus, the implementation of web scraping techniques can be an effective solution in supporting the needs of academics and researchers to obtain journal information quickly and accurately.*

Keywords: *Web Scraping, SINTA, Scientific Journals, Python, Beautifulsoup, Data Extraction.*

Abstrak: Perkembangan teknologi digital telah meningkatkan kebutuhan akses terhadap informasi ilmiah, terutama yang berkaitan dengan publikasi jurnal akademik. Situs web Indeks Sains dan Teknologi (SINTA) merupakan salah satu platform utama yang menyediakan informasi tentang jurnal ilmiah terakreditasi di Indonesia. Namun, proses pencarian dan pengumpulan data jurnal di situs web tersebut masih dilakukan secara manual, sehingga kurang efisien ketika jumlah data yang dibutuhkan cukup besar. Penelitian ini bertujuan untuk mengimplementasikan dan mengevaluasi teknik web scraping dalam pengumpulan data jurnal di situs web SINTA sehingga proses akuisisi data dapat dilakukan secara otomatis, cepat, dan terstruktur. Metode penelitian menggunakan pendekatan implementasi melalui tahapan studi pustaka, analisis struktur halaman web, pengembangan sistem menggunakan bahasa pemrograman Python dengan pustaka BeautifulSoup, pengumpulan data, penyimpanan basis data, dan evaluasi hasil ekstraksi. Hasil penelitian menunjukkan bahwa implementasi teknik web scraping berhasil memperoleh sebanyak 15.456 data jurnal dari situs web SINTA pada tahun 2026 yang mencakup informasi tentang nama jurnal, lembaga penerbitan, kategori akreditasi, ISSN, bidang ilmu, dan tautan situs web jurnal. Sistem yang dikembangkan juga mampu menyimpan data secara terstruktur ke dalam basis data dan mendukung pembaruan informasi secara berkala. Hasil evaluasi menunjukkan bahwa teknik web scraping mampu meningkatkan efisiensi pengumpulan data, meminimalkan proses manual, dan

menghasilkan data jurnal yang lebih sistematis. Dengan demikian, implementasi teknik web scraping dapat menjadi solusi efektif dalam mendukung kebutuhan akademisi dan peneliti untuk memperoleh informasi jurnal dengan cepat dan akurat.

Kata kunci: Web Scraping, SINTA, Jurnal Ilmiah, Python, BeautifulSoup, Ekstraksi Data.

PENDAHULUAN

Kemajuan teknologi digital telah membawa perubahan signifikan terhadap pola akses dan kebutuhan informasi di berbagai bidang, termasuk dalam dunia akademik. Informasi ilmiah saat ini menjadi salah satu kebutuhan utama bagi mahasiswa, dosen, maupun peneliti, terutama yang berkaitan dengan publikasi artikel pada jurnal ilmiah. Jurnal ilmiah berfungsi sebagai media penyebaran hasil penelitian yang memuat gagasan, inovasi, serta temuan ilmiah yang disusun secara sistematis berdasarkan kaidah akademik[1], [2]. Keberadaan jurnal tidak hanya menjadi sarana publikasi, tetapi juga menjadi indikator perkembangan ilmu pengetahuan dan kualitas penelitian suatu institusi pendidikan tinggi. Dalam lingkungan perguruan tinggi, publikasi karya ilmiah telah menjadi bagian penting dari aktivitas akademik[3]. Hal ini diperkuat dengan kebijakan pemerintah yang mewajibkan mahasiswa pada jenjang pendidikan tertentu untuk mempublikasikan hasil penelitiannya sebagai salah satu persyaratan akademik. Kebijakan tersebut bertujuan untuk mendorong peningkatan kualitas penelitian, memperluas kontribusi akademik, serta memperkuat daya saing institusi pendidikan Indonesia dalam menghasilkan publikasi ilmiah yang berkualitas.

Sebagai upaya mendukung pengelolaan informasi publikasi ilmiah, pemerintah Indonesia mengembangkan Sistem Informasi Sains dan Teknologi (Science and Technology Index atau SINTA) [4], [5], [6]. Platform berbasis web ini dirancang untuk mengintegrasikan data peneliti, institusi, dan jurnal ilmiah di Indonesia. Selain berfungsi sebagai sistem pengindeksan, SINTA juga digunakan untuk menilai kinerja publikasi melalui

pengelompokan jurnal berdasarkan tingkat akreditasi yang terbagi menjadi enam kategori, mulai dari SINTA 1 hingga SINTA 6. Dengan adanya sistem ini, akademisi memperoleh kemudahan dalam menemukan jurnal yang sesuai dengan bidang keilmuan dan kebutuhan publikasi. Meskipun demikian, proses pencarian dan pengumpulan informasi jurnal melalui website SINTA masih menghadapi sejumlah kendala. Pengguna sering kali harus melakukan pencarian secara berulang untuk memperoleh data tertentu, seperti nama jurnal, tautan website, bidang kajian, status akreditasi, maupun informasi pendukung lainnya. Ketika jumlah data yang dibutuhkan semakin besar, proses tersebut menjadi kurang efektif karena membutuhkan waktu dan tenaga yang cukup banyak. Kondisi ini menunjukkan perlunya metode yang mampu melakukan pengumpulan data secara otomatis agar informasi yang diperoleh lebih cepat, sistematis, dan mudah dianalisis.

Salah satu metode yang dapat dimanfaatkan untuk mengatasi permasalahan tersebut adalah web scraping[7], [8], [9]. Teknik ini memungkinkan proses ekstraksi informasi dari halaman website dilakukan secara otomatis dengan memanfaatkan struktur halaman berbasis HTML. Melalui pendekatan ini, data yang tersedia pada suatu website dapat dikumpulkan secara terstruktur tanpa memerlukan proses penginputan manual. Sejumlah penelitian sebelumnya menunjukkan bahwa web scraping efektif digunakan dalam proses pengambilan data dari berbagai platform digital. Namun, implementasi teknik ini pada pengumpulan data jurnal di website SINTA masih belum banyak dieksplorasi, khususnya yang disertai evaluasi terhadap efektivitas, keakuratan, dan efisiensi proses ekstraksi data.

Berdasarkan kondisi tersebut, penelitian ini dilakukan untuk mengimplementasikan sekaligus mengevaluasi penerapan web scraping pada pengumpulan data jurnal di website SINTA. Penelitian ini diharapkan mampu menghasilkan mekanisme pengambilan data yang lebih efisien dan terstruktur, sehingga dapat membantu akademisi maupun peneliti dalam memperoleh informasi jurnal secara lebih cepat dan akurat. Selain itu, hasil penelitian diharapkan menjadi landasan awal bagi pengembangan sistem pengelolaan dan analitik data jurnal berbasis otomatisasi pada platform akademik

Tinjauan Literatur Dan Pernyataan Masalah

Beberapa penelitian terdahulu telah mengimplementasikan teknik web scraping untuk memperoleh data dari berbagai platform berbasis web guna mendukung kebutuhan analisis informasi. Penelitian yang dilakukan oleh Purnomo, dkk membahas pemanfaatan web scraping untuk membantu peneliti dalam menentukan kualitas jurnal ilmiah. Dalam penelitian tersebut, data diperoleh melalui proses ekstraksi informasi dari portal SINTA dan SCImagoJR. Hasil penelitian menunjukkan bahwa sistem yang dikembangkan mampu menyajikan informasi jurnal berdasarkan bidang ilmu tertentu beserta indikator kualitas jurnal yang tersedia pada kedua platform tersebut. Temuan ini memperlihatkan bahwa teknik web scraping dapat digunakan sebagai sarana pendukung dalam pemilihan jurnal ilmiah yang sesuai bagi peneliti [10]. Penelitian lain dilakukan oleh Arum dan Sukmasetya yang memanfaatkan web scraping untuk analisis berita pendidikan dari beberapa portal berita daring, yaitu Detik.com, Liputan6, dan CNN Indonesia. Penelitian tersebut menerapkan algoritma Depth-First Search (DFS) untuk membantu proses penelusuran URL dan validasi data berdasarkan tanggal publikasi berita. Data yang berhasil diperoleh kemudian disimpan ke dalam basis data NoSQL guna

mendukung proses pengelolaan informasi secara lebih fleksibel. Penelitian ini menunjukkan bahwa kombinasi web scraping dengan algoritma pencarian dapat meningkatkan relevansi data yang diperoleh dari berbagai sumber digital [11]. Selanjutnya, Rahmatulloh dan Gunawan mengembangkan sistem pengumpulan data artikel ilmiah pada Google Scholar menggunakan metode web scraping berbasis HTML DOM. Penelitian tersebut menghasilkan aplikasi yang mampu merekap informasi profil peneliti, afiliasi, sitasi, dan daftar artikel ilmiah secara otomatis. Selain itu, sistem juga menyediakan fitur ekspor data ke berbagai format dokumen seperti PDF dan Excel sehingga mempermudah pengelolaan data publikasi ilmiah [12].

Penerapan web scraping juga dilakukan oleh Mufidah dan Siahaan pada bidang perdagangan elektronik melalui pengembangan aplikasi pemantauan perubahan harga produk di platform e-commerce. Penelitian ini bertujuan untuk mengumpulkan data historis harga produk secara otomatis berdasarkan alamat URL tertentu. Data hasil ekstraksi kemudian divisualisasikan dalam bentuk grafik untuk mempermudah pengguna dalam memonitor perubahan harga dari waktu ke waktu. Hasil penelitian menunjukkan bahwa teknik web scraping efektif digunakan dalam proses otomatisasi pengambilan data berbasis web [13]. Penelitian yang lebih dekat dengan topik ini dilakukan oleh Sahria yang mengimplementasikan teknik web scraping pada jurnal SINTA untuk menganalisis tren penelitian kesehatan di Indonesia. Data yang diekstraksi meliputi judul jurnal, topik penelitian, nama penulis, dan afiliasi institusi. Data tersebut kemudian dianalisis untuk mengidentifikasi kecenderungan topik penelitian kesehatan di Indonesia. Penelitian ini menegaskan bahwa hasil ekstraksi data dari SINTA dapat dimanfaatkan lebih lanjut untuk kebutuhan analisis tren penelitian secara lebih sistematis [14]. Selain itu, Priyanto dan Ma'arif menggabungkan web scraping

dengan text mining dalam proses akuisisi informasi pada laman web yang membahas hidroponik. Penelitian tersebut bertujuan mengelompokkan informasi berdasarkan kategori topik secara otomatis sehingga pengguna dapat memperoleh informasi yang lebih spesifik tanpa harus menelusuri banyak halaman web. Hasil penelitian menunjukkan bahwa integrasi web scraping dan text mining mampu meningkatkan efisiensi pengelompokan informasi berbasis topik[15].

Berdasarkan beberapa penelitian terdahulu, dapat diketahui bahwa teknik web scraping telah banyak dimanfaatkan dalam pengambilan data dari berbagai website untuk kebutuhan analisis, klasifikasi, maupun visualisasi informasi. Namun demikian, penelitian terkait implementasi web scraping pada data jurnal SINTA masih terbatas, khususnya yang mengkaji aspek evaluasi efektivitas pengumpulan data secara otomatis. Sebagian besar penelitian sebelumnya lebih menitikberatkan pada hasil ekstraksi atau analisis data tertentu, sedangkan evaluasi terhadap tingkat keberhasilan, efisiensi, dan struktur data hasil ekstraksi masih belum banyak dibahas. Oleh karena itu, penelitian ini memiliki kebaruan pada implementasi sekaligus evaluasi teknik web scraping dalam pengumpulan data jurnal SINTA guna menghasilkan proses akuisisi data yang lebih cepat, akurat, dan terstruktur

METODE

Penelitian ini menggunakan pendekatan implementatif dengan menerapkan teknik web scraping untuk memperoleh data jurnal pada website SINTA (Science and Technology Index). Tahapan penelitian dilakukan secara sistematis mulai dari studi literatur, pengumpulan data, implementasi sistem, hingga evaluasi hasil ekstraksi data[16][17]. Proses penelitian dirancang agar dapat menghasilkan data jurnal yang terstruktur serta mendukung proses pembaruan data secara otomatis.

1. Studi Literatur

Tahap awal penelitian dilakukan melalui kajian pustaka dengan mengumpulkan berbagai sumber ilmiah yang relevan, baik dari jurnal, prosiding, maupun buku yang berkaitan dengan publikasi ilmiah, website SINTA, teknik web scraping, bahasa pemrograman Python, serta mekanisme penjadwalan sistem menggunakan cron job[18]. Studi literatur dilakukan untuk memperoleh pemahaman konseptual sekaligus menjadi dasar dalam perancangan sistem pengambilan data jurnal secara otomatis.

2. Tahap Implementasi Sistem

Tahap implementasi dimulai dengan melakukan eksplorasi terhadap website SINTA guna mengidentifikasi struktur data yang akan diekstraksi. Pada tahap ini dilakukan analisis elemen halaman web, seperti struktur HTML, tag, serta pola navigasi halaman yang digunakan untuk menampilkan informasi jurnal. Data yang menjadi fokus pengambilan meliputi nama jurnal, tautan website jurnal, kategori SINTA, bidang keilmuan, serta metadata lain yang tersedia pada platform.

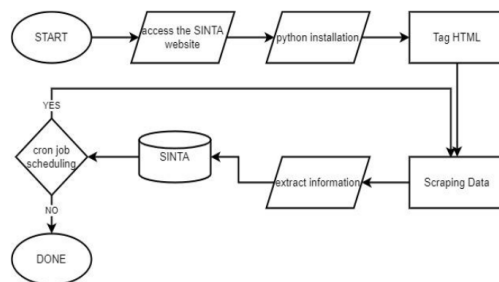
Dalam penelitian ini digunakan bahasa pemrograman Python karena memiliki fleksibilitas tinggi dalam pengolahan data serta mendukung berbagai pustaka (library) untuk ekstraksi informasi berbasis web. Proses web scraping dikembangkan menggunakan pustaka BeautifulSoup, yaitu library Python yang berfungsi untuk membaca, menelusuri, dan mengekstraksi informasi dari dokumen HTML maupun XML. Pemanfaatan BeautifulSoup memungkinkan proses identifikasi elemen halaman dilakukan secara lebih efektif berdasarkan struktur DOM (Document Object Model)[19].

Tahapan implementasi web scraping dalam penelitian ini meliputi beberapa langkah, yaitu:

- a. Perancangan Template Scraping, dilakukan dengan mempelajari struktur HTML pada website SINTA untuk mengidentifikasi tag yang memuat informasi jurnal yang dibutuhkan.

- b. Eksplorasi Navigasi Website, dilakukan untuk memahami pola perpindahan halaman, pengelompokan data jurnal, serta mekanisme akses data pada website SINTA
- c. Automasi Pengambilan Data, dilakukan dengan membangun skrip Python yang mampu mengakses halaman website secara otomatis dan mengekstraksi data jurnal berdasarkan parameter yang telah ditentukan.
- d. Penyimpanan Hasil Ekstraksi, data hasil scraping disimpan dalam format basis data agar lebih mudah diolah, diperbarui, dan dianalisis lebih lanjut.

Tahapan prosedur web scraping pada penelitian ini dapat dilihat pada Gambar 1.



Gambar 1 Tahapan Proses Web Scraping pada Website SINTA

3. Pengolahan dan Evaluasi Data

Data hasil ekstraksi dari website SINTA selanjutnya disimpan ke dalam basis data (database) untuk memudahkan pengelolaan informasi secara terstruktur. Penyimpanan data dilakukan agar informasi jurnal yang telah diperoleh dapat digunakan kembali untuk proses analisis maupun pembaruan data berikutnya.

Untuk menjaga keterbaruan data, penelitian ini menerapkan mekanisme pembaruan otomatis menggunakan cron job scheduler[20]. Sistem ini berfungsi untuk menjalankan skrip pengambilan data secara berkala berdasarkan interval waktu tertentu yang telah ditentukan sebelumnya. Dengan adanya penjadwalan otomatis, sistem mampu memperbarui informasi jurnal secara periodik tanpa perlu menjalankan proses secara manual.

Tahap evaluasi dilakukan dengan mengukur keberhasilan sistem dalam mengekstraksi data jurnal dari website SINTA. Evaluasi meliputi tingkat keberhasilan pengambilan data, kesesuaian informasi yang diperoleh, serta efisiensi proses ekstraksi berdasarkan jumlah data yang berhasil dikumpulkan. Hasil evaluasi digunakan untuk mengetahui efektivitas implementasi web scraping dalam mendukung pengumpulan data jurnal secara otomatis dan terstruktur.

HASIL DAN PEMBAHASAN

1. Implementasi Web Scraping pada Website SINTA

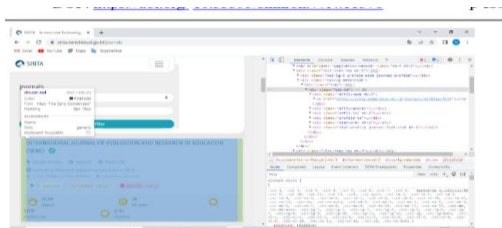
Implementasi teknik web scraping pada website SINTA berhasil menghasilkan dataset jurnal sebanyak 15.456 data jurnal yang diperoleh dari website SINTA pada tahun 2026. Proses pengambilan data dilakukan dengan mengakses halaman jurnal pada website SINTA sebagai sumber utama informasi publikasi ilmiah terakreditasi di Indonesia. Data yang berhasil diperoleh mencakup informasi nama jurnal, institusi penerbit, kategori akreditasi, ISSN, bidang ilmu, serta alamat website jurnal yang tersedia pada platform SINTA.

Tahap awal implementasi dimulai dengan mengakses halaman utama website SINTA untuk memahami struktur navigasi dan pola penyajian data jurnal. Halaman ini menjadi titik awal pengambilan data karena memuat daftar jurnal berdasarkan klasifikasi akreditasi nasional. Tampilan website SINTA yang digunakan sebagai objek penelitian dapat dilihat pada Gambar 2.



Gambar 2. Tampilan Website SINTA

Setelah halaman website berhasil diakses, tahapan berikutnya adalah melakukan analisis struktur halaman untuk mengidentifikasi elemen HTML yang menjadi sumber informasi jurnal. Pemeriksaan dilakukan menggunakan fitur inspection element pada browser untuk menemukan tag HTML yang memuat data target. Informasi yang diidentifikasi meliputi nama jurnal, institusi penerbit, kategori akreditasi, nomor ISSN, bidang keilmuan, dan tautan jurnal. Tahapan ini bertujuan untuk memperoleh pola struktur Document Object Model (DOM) yang digunakan sebagai dasar penyusunan skrip web scraping. Hasil identifikasi struktur HTML website SINTA ditunjukkan pada Gambar 3.



Gambar 3. Pemeriksaan Struktur HTML pada Website SINTA

Berdasarkan struktur HTML yang telah diperoleh, selanjutnya dilakukan pengembangan skrip web scraping menggunakan bahasa pemrograman

Python. Implementasi dilakukan dengan memanfaatkan pustaka (library) BeautifulSoup untuk membaca struktur HTML serta mengekstraksi data yang diperlukan. Selain itu, beberapa pustaka pendukung lain digunakan untuk mengoptimalkan proses pengolahan data dan konektivitas dengan sistem basis data. Konfigurasi pustaka yang digunakan pada implementasi sistem ditunjukkan pada Gambar 4.

```
import requests
from bs4 import BeautifulSoup
import mysql.connector
from mysql.connector import errorcode
```

Gambar 4. Library pada Implementasi Web Scraping

Hasil implementasi menunjukkan bahwa sistem mampu melakukan ekstraksi data jurnal secara otomatis dalam jumlah besar. Dataset awal yang diperoleh berjumlah 15.456 data jurnal yang tersusun secara terstruktur berdasarkan atribut informasi jurnal. Data yang berhasil diekstraksi meliputi identitas jurnal, institusi pengelola, informasi ISSN, kategori akreditasi SINTA, dan alamat website jurnal. Contoh hasil ekstraksi data jurnal website SINTA ditunjukkan pada Tabel 1.

Tabel 1. Hasil Ekstraksi Data Jurnal Website

No	Nama Jurnal	Institusi	Informasi Jurnal	Akreditasi	Link Jurnal
1	International Journal of Electrical and Computer Engineering (IJECE)	Institute of Advanced Engineering and Science (IAES) Yogyakarta	E-ISSN : 25024760	S1	https://ijece.iaescore.com/index.php/IJECE
2	Cogito Smart Journal	Universitas Klabat	E-ISSN : 24778079	S2	https://cogito.unklab.ac.id/index.php/cogito
3	IT Journal Research and Development,	Universitas Islam Riau	E-ISSN:25284061	S3	https://journal.uir.ac.id/index.php/ITJRD

Data hasil ekstraksi kemudian disimpan ke dalam sistem basis data untuk memudahkan proses pengelolaan, klasifikasi, dan pembaruan informasi jurnal. Penyimpanan data dilakukan secara otomatis menggunakan konektivitas antara Python dan database sehingga data yang diperoleh dapat

diakses kembali untuk kebutuhan analisis lebih lanjut.

2. Evaluasi Implementasi Web Scraping

Evaluasi sistem dilakukan untuk mengetahui tingkat keberhasilan implementasi web scraping dalam

memperoleh data jurnal pada website SINTA. Evaluasi difokuskan pada aspek keberhasilan ekstraksi data, kelengkapan informasi yang diperoleh, serta kemampuan sistem dalam menyusun data ke dalam format terstruktur. Hasil evaluasi menunjukkan bahwa sistem berhasil mengekstraksi 15.456 data jurnal dari website SINTA tanpa proses input manual. Informasi yang diperoleh mencakup atribut utama jurnal seperti nama jurnal, institusi penerbit, kategori akreditasi, ISSN, dan tautan website. Data hasil ekstraksi juga berhasil tersimpan dalam basis data sehingga dapat digunakan kembali untuk kebutuhan klasifikasi maupun analisis data.

Selain itu, penggunaan pustaka BeautifulSoup pada bahasa pemrograman Python menunjukkan kemampuan yang baik dalam membaca struktur HTML website dan melakukan ekstraksi elemen data secara otomatis. Sistem juga mampu menyesuaikan proses pengambilan data berdasarkan pola struktur halaman website SINTA sehingga informasi dapat diperoleh secara lebih cepat dibandingkan proses pencarian manual.

Berdasarkan hasil implementasi dan evaluasi yang telah dilakukan, teknik web scraping terbukti mampu mendukung proses pengumpulan data jurnal pada website SINTA secara otomatis, terstruktur, dan efisien. Hasil penelitian ini dapat dimanfaatkan sebagai dasar pengembangan sistem analitik data jurnal maupun pemetaan publikasi ilmiah berbasis otomatisasi

3. Diskusi

Hasil penelitian menunjukkan bahwa implementasi teknik web scraping pada website SINTA mampu menghasilkan dataset jurnal dalam jumlah besar secara otomatis, yaitu sebanyak 15.456 data jurnal. Temuan ini menunjukkan bahwa pendekatan berbasis web scraping efektif digunakan dalam proses akuisisi data dari website yang memiliki struktur halaman terorganisasi. Kemampuan sistem dalam memperoleh data jurnal secara otomatis memperlihatkan bahwa proses

pengumpulan informasi dapat dilakukan dengan lebih efisien dibandingkan metode manual yang memerlukan pencarian satu per satu melalui halaman website. Kondisi ini sejalan dengan pendapat Mitchell yang menyatakan bahwa web scraping merupakan teknik pengumpulan data berbasis website yang memungkinkan ekstraksi informasi dilakukan tanpa interaksi manual melalui antarmuka web secara langsung [19].

Keberhasilan ekstraksi data pada penelitian ini dipengaruhi oleh proses identifikasi struktur HTML yang dilakukan sebelum pengembangan skrip scraping. Analisis terhadap elemen Document Object Model (DOM) memungkinkan sistem mengenali pola tag HTML yang menyimpan informasi jurnal seperti nama jurnal, institusi penerbit, ISSN, kategori akreditasi, serta tautan website jurnal. Pemanfaatan struktur DOM tersebut menjadi faktor penting karena data pada website umumnya tersimpan dalam elemen HTML tertentu yang harus dikenali terlebih dahulu sebelum dilakukan ekstraksi. Hasil penelitian ini mendukung temuan Sahria yang menjelaskan bahwa implementasi web scraping pada jurnal SINTA memerlukan identifikasi struktur halaman website agar informasi penelitian dapat diekstraksi secara terstruktur dan siap dianalisis lebih lanjut [14].

Penggunaan bahasa pemrograman Python dengan pustaka BeautifulSoup juga memberikan kontribusi signifikan terhadap keberhasilan sistem. BeautifulSoup memungkinkan proses parsing HTML dilakukan secara sistematis sehingga informasi yang tidak terstruktur dapat diubah menjadi data yang lebih mudah diproses. Dalam penelitian ini, pustaka tersebut mampu membaca struktur halaman website SINTA dan mengekstraksi data jurnal secara konsisten. Hasil ini sejalan dengan penelitian Rahmatulloh dan Gunawan yang menemukan bahwa penggunaan metode HTML DOM berbasis Python efektif untuk memperoleh informasi publikasi ilmiah dari platform digital secara otomatis [12].

Selain menghasilkan data dalam jumlah besar, penelitian ini juga menunjukkan bahwa hasil ekstraksi dapat disimpan ke dalam basis data untuk mendukung pengelolaan informasi yang lebih sistematis. Penyimpanan data memungkinkan proses klasifikasi, penyaringan, dan pembaruan informasi dilakukan secara lebih mudah dibandingkan penyimpanan konvensional. Integrasi antara Python dan sistem basis data memberikan fleksibilitas dalam pengelolaan data hasil scraping sehingga informasi jurnal dapat digunakan kembali untuk berbagai kebutuhan, termasuk analisis tren publikasi dan pemetaan jurnal ilmiah. Temuan ini memperkuat penelitian Purnomo,dkk yang menyatakan bahwa data hasil web scraping dari platform SINTA dapat dimanfaatkan sebagai dasar dalam pengembangan sistem analisis kualitas jurnal ilmiah [10].

Pada aspek evaluasi, sistem yang dikembangkan menunjukkan kemampuan dalam memperoleh data jurnal secara terstruktur tanpa memerlukan proses input manual. Hal ini menunjukkan bahwa implementasi web scraping dapat meningkatkan efisiensi waktu dan mengurangi potensi kesalahan manusia (human error) pada saat pengumpulan data. Dari sisi efektivitas, sistem mampu menghasilkan informasi yang mencakup atribut utama jurnal, seperti nama jurnal, institusi penerbit, ISSN, kategori SINTA, dan alamat website jurnal. Dengan demikian, implementasi teknik ini dapat menjadi alternatif solusi dalam mendukung kebutuhan akademisi dan peneliti untuk memperoleh informasi jurnal secara cepat dan sistematis.

Meskipun demikian, penelitian ini memiliki keterbatasan pada ketergantungan terhadap struktur halaman website SINTA. Apabila terjadi perubahan struktur HTML atau pembaruan tampilan halaman, maka skrip web scraping perlu disesuaikan kembali agar proses ekstraksi tetap berjalan dengan baik. Oleh karena itu, pengembangan penelitian selanjutnya dapat diarahkan pada penerapan metode scraping yang lebih adaptif atau integrasi dengan teknologi machine learning untuk

meningkatkan fleksibilitas sistem terhadap perubahan struktur website.

SIMPULAN

Penelitian ini berhasil mengimplementasikan teknik web scraping pada website SINTA untuk memperoleh data jurnal secara otomatis dan terstruktur. Hasil implementasi menunjukkan bahwa sistem mampu mengekstraksi sebanyak 15.456 data jurnal dari website SINTA tahun 2026 yang mencakup informasi nama jurnal, institusi penerbit, kategori akreditasi, ISSN, bidang ilmu, serta tautan website jurnal. Proses pengambilan data dilakukan melalui identifikasi struktur HTML website dan penerapan bahasa pemrograman Python dengan pustaka BeautifulSoup, sehingga data dapat diperoleh secara lebih cepat dibandingkan pencarian manual. Hasil evaluasi menunjukkan bahwa sistem yang dikembangkan mampu mendukung proses pengumpulan data jurnal secara efektif dengan menghasilkan data dalam format yang lebih terorganisasi dan mudah dikelola melalui basis data. Implementasi sistem juga memberikan kemudahan dalam proses pembaruan data secara berkala sehingga informasi jurnal yang diperoleh dapat tetap mutakhir. Dengan demikian, teknik web scraping dapat dijadikan sebagai alternatif solusi dalam mendukung kebutuhan akademisi, peneliti, maupun institusi pendidikan dalam memperoleh informasi jurnal secara efisien untuk keperluan publikasi, pemetaan jurnal, maupun analisis data ilmiah. Penelitian ini memiliki peluang pengembangan lebih lanjut, terutama pada peningkatan kemampuan sistem agar lebih adaptif terhadap perubahan struktur website SINTA serta integrasi fitur analitik untuk memetakan tren publikasi ilmiah berdasarkan bidang keilmuan, akreditasi jurnal, maupun pola penerbitan. Selain itu, penelitian berikutnya dapat mengembangkan sistem berbasis dashboard visualisasi data agar hasil ekstraksi informasi lebih mudah dipahami dan dimanfaatkan oleh pengguna.

DAFTAR PUSTAKA

- N. E. Widiyastuti *et al.*, *Inovasi \& pengembangan karya tulis ilmiah: Panduan lengkap untuk penelitian dan mahasiswa*. PT. Sonpedia Publishing Indonesia, 2023.
- A. Julia, S. Febriani, and S. Nurjanah, “PUBLIKASI ILMIAH DAN INOVASI PENDIDIKAN,” *J. Ilmu Pendidik. Islam*, vol. 23, no. 4, pp. 225–240, 2025.
- N. Agustin and A. Fithriyah, “Pendampingan penulisan karya ilmiah bagi mahasiswa sebagai upaya peningkatan budaya akademik di perguruan tinggi,” *J. Pengabd. Masy.*, vol. 2, no. 2, pp. 235–246, 2025.
- K. T. Ilmiah, “Peningkatan Pengetahuan Dan Keterampilan Dalam Penyusunan Karya Tulis Ilmiah Terakreditasi Sinta,” *Community Dev. J.*, vol. 4, no. 2, 2023.
- A. Saputra, “Pemanfaatan science and technology index (sinta) untuk publikasi karya ilmiah dan pencarian jurnal nasional terakreditasi,” *Media Pustak.*, vol. 27, no. 1, pp. 56–68, 2020.
- D. Darman, T. S. Maulana, and R. Mansur, “Pendampingan optimalisasi profil SINTA bagi dosen melalui perbaikan data dan penguatan dokumentasi publikasi,” *Eastasouth J. Eff. Community Serv.*, vol. 4, no. 02, pp. 309–317, 2025.
- D. Chrisinta and J. E. Simarmata, “Eksplorasi teknik web scraping pada data mining: Pendekatan pencarian data berbasis Python,” *Fakt. Exacta*, vol. 17, no. 1, 2024.
- M. R. Fikri, R. T. Handayanto, and D. Irwan, “Web Scraping Situs Berita Menggunakan Bahasa Pemrograman Python,” *J. Students' Research Comput. Sci.*, vol. 3, no. 1, pp. 123–136, 2022.
- N. ADILA, “IMPLEMENTASI WEB SCRAPING UNTUK PENGUMPULAN DATA JURNAL PADA WEBSITE SINTA,” Nusa Putra, 2022.
- L. M. Purnomo and M. Ayub, “Analisis data hasil web scraping untuk menentukan kualitas jurnal ilmiah,” *J. Strateg. Maranatha*, vol. 3, no. 1, pp. 122–132, 2021.
- E. R. Arumi and P. Sukmasetya, “Exploiting web scraping for education news analysis using depth-first search algorithm,” *J. Online Inform.*, vol. 5, no. 1, pp. 19–26, 2020.
- A. Rahmatulloh, R. Gunawan, and others, “Web scraping with HTML DOM method for data collection of scientific articles from Google Scholar,” *Indones. J. Inf. Syst.*, vol. 2, no. 2, pp. 95–104, 2020.
- U. Mufidah and M. Siahaan, “Perancangan aplikasi perbandingan harga produk (historical data) menggunakan teknik web scraping,” *Pusdansi. org*, vol. 1, no. 1, pp. 1–14, 2021.
- Y. Sahria, “Implementasi Teknik Web Scraping pada jurnal SINTA untuk analisis topik penelitian kesehatan Indonesia,” *Proceeding of The URECOL*, pp. 297–306, 2020.
- A. Priyanto, M. R. Ma'arif, and others, “Implementasi web scrapping dan text mining untuk akuisisi dan kategorisasi informasi dari internet (studi kasus: Tutorial hidroponik),” *Indones. J. Inf. Syst.*, vol. 1, no. 1, pp. 25–33, 2018.
- M. F. Adham, “Analisis implementasi sistem informasi: studi literatur,” *J. Teknol. Sist. Inf.*, vol. 5, no. 1, pp. 264–275, 2024.
- P. Deepublish, “Studi Literatur: Pengertian, Ciri, Teknik Pengumpulan Datanya,” *Diakses pada tanggal*, vol. 7, 2023.
- F. Sembiring and A. Erfina, *Bahasa Ular untuk Pemrograman Python*. Insan Cendekia Mandiri, 2020.
- R. Mitchell, *Web scraping with python*. “O'Reilly Media, Inc.,” 2024.
- C. Chusna, S. Ilham, A. C. Fauzan, and others, “Implementasi Penjadwalan Round Robin pada Task Scheduler untuk Pembaruan Aplikasi Otomatis,” *ILKOMNIKA*, vol. 1, no. 1, pp. 11–14, 2019.