

IMPLEMENTASI WEB SCRAPING PADA SITUS JURNAL SINTA MENGUNAKAN FRAMEWORK SELENIUM WEBDRIVER PYTHON

Amanny Ulfah Nabiylah Ramadhanty¹⁾, Ina Najiyah²⁾

^{1,2} Sistem Informasi, Fakultas Teknologi Informasi, Universitas Adhirajasa Reswara Sanjaya, Jalan Sekolah Internasional
No. 1-2 Bandung
Co Responden Email: inajiyah@ars.ac.id

Abstract

Article history

Received 14 Oct 2022

Revised 24 Nov 2022

Accepted 01 Jan 2023

Available online 15 Feb 2023

Keywords

Web Scraping,
Scientific Articles,
Python,
CodeIgniter,
Selenium

The search for references to scientific articles is the initial stage of designing a study. Researchers often start by looking for a wide variety of references that have the same topic and purpose as what is being studied. Scientific articles are very easy to find through the internet. It's just that with the large number of scientific articles, the search process will take a long time. This is because some articles have been deleted or need more access to open them. Therefore, to speed up students and researchers to get scientific articles that are in accordance with the research topic, a special website is needed that is able to collect some information about scientific articles. This study aims to create a web page that contains special information about references to scientific articles that can help students of the Bakti Asih College of Analysts and researchers to speed up the search for references to scientific articles. The design of this system uses a web scraping method for data retrieval of scientific articles, which uses the Python programming language using the selenium library, pandas library, BeautifulSoup library and openpyxl library. The result of the scraping data in the form of an excel file created by a web page on the siacad web of the Bakti Asih Bandung College of Analysts using the CodeIgniter Framework with the Php programming language. System development uses the SDLC (Waterfall) method and system testing uses the blackbox testing method

Abstrak

Riwayat

Diterima 14 Okt 2022

Revisi 24 Nov 2022

Disetujui 01 Jan 2023

Terbit online 15 Feb 2023

Kata Kunci

Web Scraping,
Artikel Ilmiah,
Python, CodeIgniter,
Selenium

Pencarian referensi artikel ilmiah merupakan tahap awal dari perancangan sebuah penelitian. Para peneliti seringkali memulai dengan mencari berbagai macam referensi yang memiliki topik dan tujuan yang sama dengan apa yang diteliti. Artikel ilmiah sangat mudah ditemukan melalui internet. Hanya saja dengan banyaknya jumlah artikel ilmiah, proses pencarian akan memakan waktu yang cukup lama. Hal ini dikarenakan beberapa artikel yang sudah terhapus atau membutuhkan akses lebih untuk membukanya. Maka dari itu, untuk mempercepat para mahasiswa dan peneliti mendapatkan artikel ilmiah yang sesuai dengan topik penelitian, diperlukan sebuah laman web khusus yang mampu mengumpulkan beberapa informasi mengenai artikel ilmiah. Penelitian ini bertujuan membuat sebuah laman web yang berisi informasi khusus mengenai referensi artikel ilmiah yang dapat membantu mahasiswa Sekolah Tinggi Analisis Bakti Asih dan para peneliti untuk mempercepat pencarian referensi artikel ilmiah. Perancangan sistem ini menggunakan metode web scraping untuk pengambilan data artikel ilmiah, yang menggunakan bahasa pemrograman Python dengan menggunakan library selenium, library pandas, library BeautifulSoup dan library openpyxl. Hasil data scraping yang berupa file excel yang dibuatkan sebuah laman web pada web siacad Sekolah Tinggi Analisis Bakti Asih Bandung yang menggunakan Framework CodeIgniter dengan bahasa pemrograman Php. Pengembangan sistem menggunakan metode SDLC (Waterfall) dan pengujian sistem menggunakan metode pengujian blackbox.

PENDAHULUAN

Sebelum melakukan sebuah penelitian, mahasiswa atau para peneliti seringkali memulai dengan mencari berbagai referensi yang berkaitan dengan tujuan penelitian tersebut (Flores, Permatasari, & Jasa, 2020). Referensi biasanya berupa artikel ilmiah ataupun berupa teori. Artikel ilmiah sangat mudah ditemukan (Adli & Firgia, 2018). Hanya saja dengan banyaknya jumlah artikel ilmiah, proses pencarian artikel ilmiah memakan waktu yang cukup lama. Hal ini dikarenakan beberapa artikel ilmiah sudah terhapus atau membutuhkan akses lebih untuk membukanya (Flores et al., 2020). Situs yang biasa digunakan untuk pencarian artikel ilmiah adalah Jurnal Sinta (*Science and Technology Index*). Jurnal Sinta merupakan salah satu situs yang digunakan untuk menyimpan hasil-hasil penelitian di Indonesia. Hasil penelitian di Indonesia sangat lah banyak dan diperlukan pengeksekusan yang lebih cepat dan efisien untuk segera mendapatkan jurnal yang sesuai dengan tema penelitian. pengambilan beberapa data yang sesuai bisa dilakukan dengan teknik *web scraping* (Sahria, 2020).

Web Scraping adalah metode untuk mengekstraksi informasi dari situs *web*, yang kemudian menjadi data yang dapat dianalisis, diolah dan dimanfaatkan untuk berbagai tujuan (Djufri, 2020). Teknik *web scraping* dapat digunakan untuk mendapatkan artikel ilmiah secara otomatis dan dapat melakukan pencarian apa-apa yang tidak dapat dilakukan oleh mesin pencari tradisional seperti *google search* (Djufri, 2020). Tujuan dari *web scraping* ini untuk mendapatkan informasi tertentu dari suatu *website* secara otomatis dan tidak harus menyalinnya secara manual (Satriajati, Panuntun, & Pramana, 2021). Data atau informasi yang diambil dapat berupa teks, tautan, video, audio ataupun dokumen (Flores et al., 2020). Menurut, (Sujarwadi & Zailani, 2019) *web scraping* terdiri dari beberapa langkah yaitu membuat *template scraping*, mempelajari teknik navigasi dari *website* yang akan diambil informasinya, mengotomatisasi pengambilan informasi dan memasukan informasi yang sudah diambil kedalam *database*.

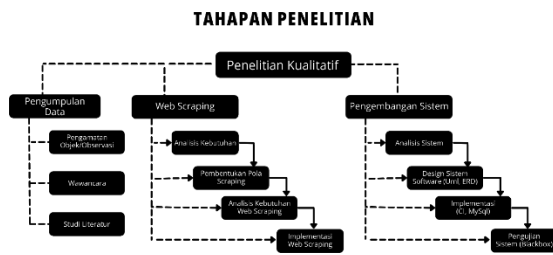
Web Scraping telah diterapkan dalam berbagai bidang sebagai alat bantu untuk pengambilan data seperti analisis topik

penelitian kesehatan indonesia pada jurnal sinta (Sahria & Fudholi, 2020), penerapan *scraping web* pada *marketplace indonesia* (Sembiring et al., 2020), untuk ekstraksi web pada situs berita (M. A. Rohim, 2018), untuk menentukan kualitas jurnal ilmiah (Purnomo & Ayub, 2021), bahkan untuk pengambilan data pada situs *marketplace* yang digunakan untuk membuat *website* perbandingan harga produk (Mufidah, 2018).

Berdasarkan latar belakang yang telah diuraikan penulis penelitian ini dilakukan menggunakan teknik *web scraping* untuk melakukan pengambilan beberapa data pada situs Jurnal Sinta yang berkaitan dengan Jurnal Analisis Kesehatan. Data yang diambil berupa judul jurnal, gambar jurnal, profil jurnal, *link* jurnal, nomor jurnal dan akreditasi jurnal. Penelitian ini dibuat menggunakan bahasa pemrograman *python* dan menggunakan *library selenium, pandas, beautifulsoup*, dan *openpyxl* yang dapat memproses *web scraping*. Pada proses *scraping* dilakukan dengan identifikasi *tag HTML*. *Tag HTML* yang digunakan yaitu *tag* yang menaruh judul jurnal, gambar jurnal, profil jurnal, *link* jurnal, nomor jurnal dan akreditasi jurnal. Yang kemudian dibuatkan *template scraping*. Data yang diperoleh kemudian disimpan dalam bentuk format *.XLSX*. Dokumen dalam format *.XLSX* dan dimasukan dalam *database* yang akan dihubungkan dengan sebuah halaman khusus pada *web* Siakad Sekolah Tinggi Analisis Bakti Asih Bandung Yang menggunakan *framework Code Igniter*. Penelitian ini diharapkan bisa membantu para mahasiswa tingkat akhir Sekolah Tinggi Analisis Bakti Asih Bandung yang mencari referensi penelitian tanpa harus keluar dari *web* siakad kampus.

METODE PENELITIAN

Metode penelitian yang digunakan pada penelitian ini adalah penelitian terapan yang bertujuan untuk mendapatkan solusi dari suatu masalah di lingkungan sekolah tinggi maupun lingkungan pendidikan sebagai kelanjutan dari penelitian dasar. Penelitian ini lebih diarahkan untuk menciptakan inovasi serta pengembangan ilmu pengetahuan dan teknologi. Metode penelitian ini dibagi kedalam tiga teknik yaitu:



Gambar 2. 1 Tahapan Penelitian

1. Teknik Pengumpulan Data

Teknik pengumpulan data terbagi menjadi 3 tahap, dimulai dari tahap pengamatan objek/observasi (melihat dan mengamati sistem yang sudah berjalan), dilanjut dengan wawancara bersama bagian IT STABA dan bagian-bagian lain yang ikut menggunakan sistem, terakhir melakukan studi literatur yang digunakan untuk mencari penelitian terdahulu yang berhubungan dengan objek penelitian.

2. Teknik Web Scraping

Teknik ini terdiri dari beberapa tahap, tahap awal yang dilakukan penulis yaitu analisis kebutuhan untuk mendapatkan analisis struktur situs jurnal sinta dan pembentukan pola. Tahap selanjutnya yaitu desain aplikasi *web scraping* untuk mengimplementasikan *web scraping* pada situs jurnal sinta, terakhir melakukan pengujian untuk melihat apakah pola *web scraping* yang dipakai sudah berjalan dengan baik atau tidak.

3. Teknik Pengembangan Sistem

Teknik yang digunakan untuk pengembangan sistem yaitu SDLC (*System Development Life-Cycle*) dengan model *waterfall*. karena metode ini mengerjakan pengembangan sistem yang dilakukan secara berurutan dan juga merupakan metode yang sering dipakai oleh peneliti dalam pengembangan sistem pada umumnya.

HASIL DAN PEMBAHASAN

Pengembangan implementasi *web scraping* menggunakan bahasa pemrograman python yang memakai *framework Selenium*, *library pandas*, *Library BeautifulSoup* dan *library Openpyxl*. Berikut tahapan pengimplementasian yang berdasarkan urutan tahapan penelitian.

3.1. Analisis Struktur HTML Website Jurnal Sinta dan Pembentukan Pola Web Scraping

Tahapan ini merupakan tahapan awal yang dilakukan penulis. Tahapan ini melakukan analisis terhadap struktur html situs jurnal sinta. Analisis ini dilakukan untuk mencari struktur penulisan sembilan konten yang ingin di ekstrak seperti, judul jurnal, gambar jurnal, profil jurnal, link jurnal, nomor jurnal, akreditasi jurnal, volume jurnal, tahun terbit jurnal, dan nomor DOI jurnal.

A. Penulisan Judul Jurnal

- Analisa Penempatan

Stuktur penempatan nama jurnal yang diletakkan di antara tag `<div> ... </div>`. nama jurnal ditulis secara umum adalah `<div class="nama class"> Judul Jurnal <i class="nama class"> </div>`.

- Pola Web Scraping

Berdasarkan pola html tersebut, pola ekstraksi nama jurnal yaitu menggunakan tag `<div> .. </div>` dan atribut `class` sudah cukup untuk membentuk sebuah pola *web scraping*. Pola yang digunakan adalah

Judul = area.find('div', class_="affil-name mb-3").get_text()

B. Penulisan Gambar Jurnal

- Analisa Penempatan

Stuktur penempatan gambar jurnal yang diletakkan di antara tag `<div> ... </div>`. Gambar jurnal ditulis secara umum adalah `<div class="nama class"> </div>`.

- Pola Web Scraping

Berdasarkan pola html tersebut, pola ekstraksi gambar jurnal yaitu menggunakan tag `` sudah cukup untuk membentuk sebuah pola *web scraping*. Pola yang digunakan adalah

Gambar = area.find('img')['src']

C. Penulisan profil Jurnal

- Analisa Penempatan

Stuktur penempatan profil jurnal yang diletakkan di antara tag `<div> ... </div>`. Profil jurnal ditulis secara umum

adalah `<div class="nama class"> Judul Jurnal <i class="nama class"> </div>`.

- Pola Web Scraping

Berdasarkan pola html tersebut, pola ekstraksi nomor jurnal yaitu menggunakan tag `<div> .. </div>` dan atribut class sudah cukup untuk membentuk sebuah pola *web scraping*. Pola yang digunakan adalah

Profil = area.find('div', class_="affil-loc mt-2").get_text()

D. Penulisan Link Jurnal

- Analisa Penempatan

Struktur penempatan link jurnal yang diletakkan di antara tag `<a> `. Link jurnal ditulis secara umum adalah ` nama Jurnal `.

- Pola Web Scraping

Berdasarkan pola html tersebut, pola ekstraksi nomor jurnal yaitu menggunakan tag `<a> .. ` dan atribut class sudah cukup untuk membentuk sebuah pola *web scraping*. Pola yang digunakan adalah

link = area.find('a')['href']

E. Penulisan Nomor Jurnal

- Analisa Penempatan

Struktur penempatan nomor jurnal yang diletakkan di antara tag `<div> </div>`. Nomor jurnal ditulis secara umum adalah `<div class="nama class"><i class="nama class"></i> Nomor Jurnal </div>`.

- Pola Web Scraping

Berdasarkan pola html tersebut, pola ekstraksi nomor jurnal yaitu menggunakan tag `<div> .. </div>` dan atribut class sudah cukup untuk membentuk sebuah pola *web scraping*. Pola yang digunakan adalah

Nomor = area.find('div', class_="profile-id").get_text()

F. Penulisan Akreditasi Jurnal

- Analisa Penempatan

Struktur penempatan akreditasi jurnal yang diletakkan di antara tag ` `. akreditasi jurnal ditulis secara umum adalah ` S4`

`<span class=Accredited`.

- Pola Web Scraping

Berdasarkan pola html tersebut, pola ekstraksi akreditasi jurnal yaitu menggunakan tag ` .. ` dan atribut class sudah cukup untuk membentuk sebuah pola *web scraping*. Pola yang digunakan adalah

akreditasi = area.find('span', class_="num-stat accredited").get_text()

G. Penulisan Volume Jurnal

- Analisa Penempatan

Struktur penempatan volume jurnal yang diletakkan di antara tag `<i> </i>`. Volume jurnal ditulis secara umum adalah `<i class="nama class"></i> Volume Jurnal`.

- Pola Web Scraping

Berdasarkan pola html tersebut, pola ekstraksi volume jurnal yaitu menggunakan tag `<div> .. </div>` dan atribut class sudah cukup untuk membentuk sebuah pola *web scraping*. Pola yang digunakan adalah

volume = area.find('a', class_="ar-pub").get_text()

H. Penulisan Tahun Terbit Jurnal

- Analisa Penempatan

Struktur penempatan tahun terbit jurnal yang diletakkan di antara tag `<a> `. Tahun terbit jurnal ditulis secara umum adalah ` Tahun Terbit Jurnal `.

- Pola Web Scraping

Berdasarkan pola html tersebut, pola ekstraksi tahun terbit jurnal yaitu menggunakan tag `<a> .. ` dan atribut class sudah cukup untuk membentuk sebuah pola *web scraping*. Pola yang digunakan adalah

tahun = area.find('a', class_="ar-year").get_text()

I. Penulisan Nomor DOI Jurnal

- Analisa Penempatan

Struktur penempatan profil jurnal yang diletakkan di antara tag `<a> `. Nomor DOI jurnal ditulis secara umum adalah `</i> nomor DOI `.

- Pola *Web Scraping*

Berdasarkan pola html tersebut, pola ekstraksi nomor DOI yaitu menggunakan tag `<div> .. </div>` dan atribut class sudah cukup untuk membentuk sebuah pola *web scraping*. Pola yang digunakan adalah

```
doi = area.find('a', class_="ar-cited").get_text()
```

3.2. Implementasi Web Scraping

Seperti yang sudah dijelaskan pada tahap penelitian, sebelum mengeksekusi *web scraping*, sebelumnya harus menganalisa struktur halaman web untuk mengetahui letak tag html yang menyimpan informasi yang akan diambil. Berikut struktur html yang ada pada *website* jurnal sinta. Setelah mengetahui tag html yang akan diambil informasinya, teknik *scraping* ditulis dalam bahasa pemrograman *Python* dengan *framework Selenium* untuk mempermudah proses *scraping*. Berikut cuplikan *coding scraping website* jurnal sinta.

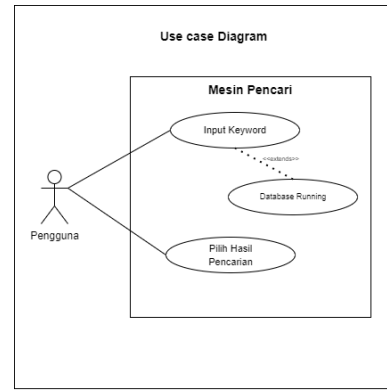
3.3. Perancangan dan Desain Sistem

A. Perancangan Sistem

Berdasarkan analisis sebelumnya, maka akan dirancang model desain dan pengembangan sistem sesuai dengan kebutuhan. Model pengembangan yang digunakan adalah UML. UML digunakan dalam proses desain sistem untuk menggambarkan hubungan antara pengguna dan aplikasi (Sanjaya & Susanti, 2021). Berikut beberapa diagram yang digunakan dalam perancangan aplikasi web *scraping* yaitu :

i. Usecase Diagram

Use case merupakan penggambaran fungsi pada sistem yang berbentuk simbol sebagai dokumentasi dan juga merupakan gambaran sekumpulan aktor pada sistem yang ada (A. A. Rohim & Rachman, 2022).



Gambar 1. Use Case diagram Aplikasi jurnal

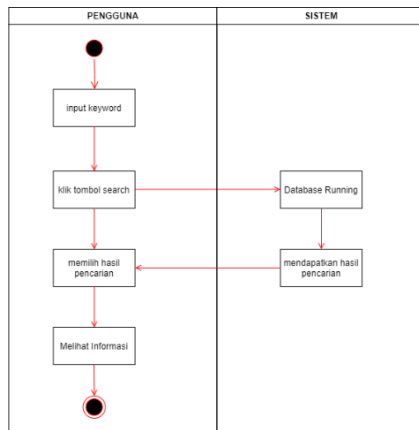
Usecase diagram aplikasi pencarian jurnal hanya terdapat 1 actor yaitu pengguna yang bisa menginputkan keyword pencarian dan memilih hasil pencarian.

Tabel 1. Deskripsi *Use Case Diagram input keyword*

<i>Use Case Name</i>	Melakukan Input Keyword
Requirements	B1
Goal	Pengguna masuk ke halaman utama
Pre-condition	Pengguna memasukan keyword pencarian
Post-Condition	Pengguna dapat memilih hasil pencarian
Failed condition and	Pengguna dapat membatalkan input keyword
Primary Actor	Pengguna
Main Flow/Basic Path	Pengguna meng-inputkan Keyword
Invariant	-

ii. Activity Diagram

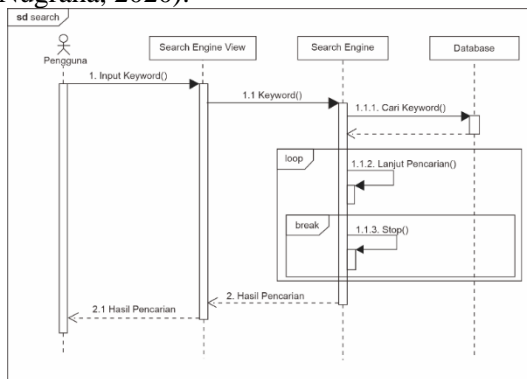
Activity diagram memodelkan seluruh alur aktifitas yang sedang terjadi pada sistem. Berikut *activity diagram* aplikasi pencarian jurnal.



Gambar 2. Activity diagram Aplikasi pencarian jurnal

iii. Sequence Diagram

Sequence diagram berfungsi untuk menjelaskan interaksi yang terjadi pada objek berdasarkan urutan waktu serta dapat menggambarkan tahapan yang harus dilakukan dalam pembuatan sistem (Syarif & Nugraha, 2020).

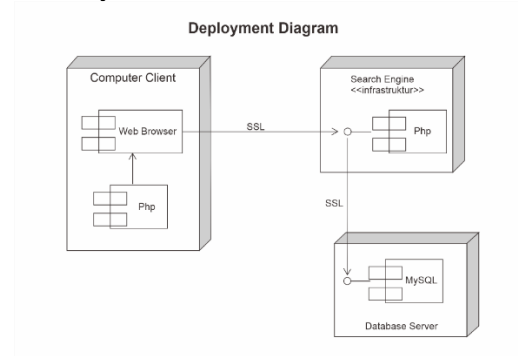


Gambar 3. Sequence diagram Aplikasi pencarian jurnal

iv. Deployment Diagram

Deployment diagram adalah diagram yang menggambarkan tataletak sistem secara fisik yang terdapat pada bagian perangkat lunak (software) dan (hardware) yang menggambarkan relasi atau hubungan antara

keduanya.



Gambar 4. Deployment diagram Aplikasi pencarian jurnal

3.4. Perancangan Database

Perancangan database (basis data) pada aplikasi pencarian jurnal dibuat berdasarkan kebutuhan masukan (input) yang diperlukan. Rancangan database atau basis data pada aplikasi pencarian jurnal terdiri dari beberapa tabel sebagai berikut:

v. Tabel Jurnal

Tabel 2. Tabel Jurnal

No.	Nama Field	Type Data	Keterangan
1.	Id_jurnal	Int (11)	Primary key
2.	Judul_jurnal	Varchar(100)	
3.	Cover_jurnal	Varchar(100)	
4.	Sumber_jurnal	Varchar(100)	
5.	Link_jurnal	Varchar(100)	
6.	Nomor_jurnal	Varchar(100)	
7.	Akreditasi_jurnal	Varchar(100)	

vi. Tabel Artikel

Tabel 3. Tabel Artikel

No.	Nama Field	Type Data	Keterangan
1.	Id_artikel	Int(11)	Primary key
2.	Id_jurnal	Int(11)	Foreign key
3.	Judul_artikel	Text	
4.	Link_artikel	Varchar(100)	
5.	Volume_artikel	Text	
6.	Tahun_terbit	Year(4)	
7.	Doi_artikel	Varchar(100)	

3.5. Implementasi Sistem/Code

Implementasi merupakan tahapan selanjutnya dari perancangan sistem yang sudah dibahas sebelumnya. Mengenai hasil dan pembahasan akan dijelaskan mengenai cara kerja dan hasil dari sistem yang sudah dibuat. Dalam pembuatan aplikasi ini peneliti menggunakan bahasa pemrograman PHP dengan framework CodeIgniter 3 untuk

merancang aplikasi berbasis web (Zharfan & Najiyah, 2022). Penyimpanan atau *database* menggunakan MySQL, *Visual Studio Code* digunakan sebagai teks editornya.

3.6. Perancangan Antarmuka Sistem

Antarmuka atau *interface* yang dibutuhkan sesuai dengan hasil analisa diatas hanya ada 2 (dua) tampilan antarmuka, yaitu sebagai berikut:

1. Halaman *search-engine*

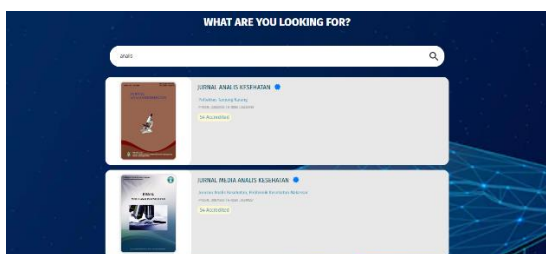
Pada halaman ini pengguna yang akan melakukan pencarian jurnal, dengan mengetikkan kata kunci sesuai dengan topik yang akan dicari.



Gambar 5. Halaman *search-engine*

2. Halaman *search*

Halaman ini merupakan halaman hasil dari pencarian dari kata kunci yang diketikkan dihalaman sebelumnya. Pengguna bisa memilih hasil pencarian yang sesuai dengan apa yang dicari. Halaman ini berisi artikel-artikel ilmiah yang sudah dikelompokkan sesuai dengan jurnal masing-masing.



Gambar 6. Halaman *search*

3.7. Pengujian Sistem

Pengujian sistem yang dipakai yaitu metode *blackbox testing*. Pengujian ini berfokus pada persyaratan fungsionalitas perangkat lunak. Berikut ini merupakan pengujian yang dilakukan dengan *blackbox* (kotak hitam).

Tabel 4. Pengujian Sistem

No	Skenario pengujian	Case	Hasil diharapkan	Hasil Pengujian	Keterangan
1.	Button Search	Mengetikkan <i>Keyword</i> lalu tekan tombol <i>search</i>	Sistem akan menampilkan hasil pencarian sesuai dengan <i>keyword</i>	Sesuai harapan	valid

KESIMPULAN

Berdasarkan hasil pembahasan dalam penelitian ini, sehingga dapat diambil kesimpulan yaitu:

- Berdasarkan hasil penelitian terhadap struktur HTML situs jurnal sinta, ditemukan beberapa pola *web scraping* yang dapat digunakan untuk melakukan proses ekstraksi data dari *keyword* Analisis Kesehatan. Terdapat perbedaan pola pada masing-masing informasi yang diambil. Perbedaan tersebut terletak pada *tag* apa yang digunakan dan atribut apa saja yang dipakai.
- Telah dihasilkan sebuah halaman *web* yang berupa *web search engine* yang berisi informasi khusus mengenai referensi artikel ilmiah.

REFERENSI

- Adli, M. A., & Firgia, L. (2018). Rancang Bangun Web Scraping Pada Media Online Berita Nasional. *Jurnal ENTER : Jurnal Online Mahasiswa Program Studi Teknik Informatika STMIK PONTIANAK*, 1, 118–128.
- Djufri, M. (2020). Penerapan Teknik Web Scraping Untuk Penggalan Potensi Pajak (Studi Kasus Pada Online Market Place Tokopedia, Shopee Dan Bukalapak). *Jurnal BPPK : Badan Pendidikan Dan Pelatihan Keuangan*, 13(2), 65–75. <https://doi.org/10.48108/jurnalbppk.v13i2.636>
- Flores, V. A., Permatasari, P. A., & Jasa, L. (2020). Penerapan Web Scraping

- Sebagai Media Pencarian dan Menyimpan Artikel Ilmiah Secara Otomatis Berdasarkan Keyword. *Majalah Ilmiah Teknologi Elektro*, 19(2), 157. <https://doi.org/10.24843/mite.2020.v19i02.p06>
- Mufidah, U. (2018). Perancangan Aplikasi Perbandingan Harga Produk (Historical Data) Menggunakan Teknik Web Scraping. *Skripsi*, 1(1), 1–14.
- Purnomo, L. M., & Ayub, M. (2021). Analisis data hasil web scraping untuk menentukan kualitas jurnal ilmiah. *Jurnal STRATEGI-Jurnal Maranatha*, 3(1), 122–132. Retrieved from <http://strategi.it.maranatha.edu/index.php/strategi/article/view/237>
- Rohim, A. A., & Rachman, R. (2022). Sistem Informasi Pemesanan Makanan Minuman berbasis Android Hybrid Sun and Grass Coffee, 3(1), 94–104.
- Rohim, M. A. (2018). Digital Digital Repository IMPLEMENTASI EKSTRAKSI WEB (WEB SCRAPING) PADA SITUS BERITA MENGGUNAKAN METODE EKSPRESI REGULER. *Skripsi*, 68–74.
- Sahria, Y. (2020). Implementasi Teknik Web Scraping pada Jurnal SINTA Untuk Analisis Topik Penelitian Kesehatan Indonesia. *URECOL (Unversity Research Colloqium)*, 11(2020), 297–306. Retrieved from <http://repository.urecol.org/index.php/proceeding/article/view/1079>
- Sahria, Y., & Fudholi, D. H. (2020). Analisis Topik Penelitian Kesehatan di Indonesia Menggunakan Metode Topic Modeling LDA. *Jurnal Rekayasa Sistem Dan Teknologi Informasi*, 4(2), 336–344.
- Sanjaya, L., & Susanti, S. (2021). Perancangan Sistem Informasi Penerimaan Siswa Baru Berbasis Web Di Smp Taruna Mandiri Cimahi. *E-PROSIDING SISTEM INFORMASI* Vol. 2, No. 2, Desember 2021, 2(2), 84–92.
- Satriajati, S., Panuntun, S. B., & Pramana, S. (2021). Implementasi Web Scraping Dalam Pengumpulan Berita Kriminal Pada Masa Pandemi Covid-19. *Seminar Nasional Official Statistics*, 2020(1), 300–308. <https://doi.org/10.34123/semnasoffstat.v2020i1.578>
- Sembiring, F., Yudistyril, D., Sari, D. P., Sistem Informasi, P., Pendidikan, P., Dan, S., & Informasi, T. (2020). Penerapan Teknik Scraping Python Pada Website Marketplace Indonesia. *INTEGRATED (Information Tecknology and Vocational Education)*, 2(1), 15–21.
- Sujarwadi, F., & Zailani, A. U. (2019). Prosiding Seminar Nasional Informatika PERANCANGAN SISTEM INFORMASI WEB SCRAPING RESEP MASAKAN BERBASIS PHP DESIGN WEB SCRAPING INFORMATION SYSTEM OF PHP-BASED FOOD RECIPES. *Prosiding Seminar Nasional Informatika Dan Sistem Informasi*, 4(1), 34–45.
- Syarif, M., & Nugraha, W. (2020). Pemodelan Diagram UML Sistem Pembayaran Tunai Pada Transaksi E-Commerce. *Jurnal Teknik Informatika Kaputama (JTIK)*, 4(1), 70 halaman. Retrieved from <http://jurnal.kaputama.ac.id/index.php/JTIK/article/view/240>
- Zharfan, R. N., & Najiyah, I. (2022). Rancang Bangun Aplikasi Cyber Rongsok Berbasis Website Menggunakan Framework Codeigniter. *E-PROSIDING SISTEM INFORMASI*, 3(1).