

Perancangan Aplikasi Data Crawling Untuk Pencarian Buku Pada Toko Buku Online

1st Bryan Ronald Talisman
Fakultas Rekayasa Industri
Universitas Telkom

Bandung, Indonesia
bryanronaldtalisman@student.telkomuniversity.ac.id

2nd Irfan Darmawan
Fakultas Rekayasa Industri
Universitas Telkom

Bandung, Indonesia
irfandarmawan@telkomuniversity.ac.id

3rd Oktariani Nurul Pratiwi
Fakultas Rekayasa Industri
Universitas Telkom

Bandung, Indonesia
onurult@telkomuniversity.ac.id

Abstrak — Data crawling adalah sebuah engine yang memungkinkan pengumpulan data dari berbagai sumber di halaman website, seperti informasi penjualan buku pada toko buku online, termasuk harga dan ketersediaan stok. Tantangannya adalah variasi ketersediaan stok di setiap toko buku online, yang memerlukan upaya ekstra untuk mengecek satu per satu. Penelitian ini bertujuan untuk membuat sebuah website yang dapat melakukan pencarian buku online dengan informasi yang akurat dan lengkap tanpa biaya tambahan untuk penjual. Data dikumpulkan melalui teknik data crawling karena tidak ada akses langsung ke database toko buku online target. Metode pengembangan menggunakan Rapid Application Development (RAD), dengan perancangan menggunakan UML dan pemrograman Python, HTML, serta PHP, dan menggunakan database MySQL. Hasil penelitian ini adalah sebuah engine data crawling yang mengumpulkan data dari toko buku Grobmart dan Deepublishstore, serta aplikasi search engine untuk mencari harga buku termurah dan ketersediaan dari data yang telah dikumpulkan. Aplikasi ini memudahkan masyarakat dalam mencari buku, memperluas pasar bagi penjual, dan meningkatkan transaksi di Grobmart dan Deepublishstore.

Kata kunci— data crawling, search engine, web crawling, rapid application development (RAD), unified modelling language (UML)

I. PENDAHULUAN

Ilmu pengetahuan adalah hal yang penting bagi semua kalangan dan usia, dan tidak akan pernah habis. Pepatah "buku adalah jendela dunia"[1] mengilustrasikan bahwa membaca buku dapat memberikan beragam ilmu pengetahuan baru, sehingga memperluas wawasan. Oleh karena itu, membaca banyak buku menjadi hal yang sangat penting untuk membuka wawasan dunia, menjadikan buku sebagai kebutuhan esensial.

Di era modern ini, internet menawarkan kemudahan akses yang memungkinkan penggunaannya menghemat waktu dan biaya, sehingga internet menjadi sumber utama untuk mencari informasi dan menambah wawasan. Meski demikian, dalam konteks pembuatan karya ilmiah atau penelitian, artikel dari internet sering tidak disarankan sebagai referensi karena kurangnya kontrol kualitas[1] Buku tetap menjadi pilihan utama yang belum tersisihkan untuk sumber yang jelas dan dapat dipertanggungjawabkan.

Pentingnya buku sebagai sumber ilmu pengetahuan mendorong berkembangnya toko buku yang tersebar di berbagai daerah di Indonesia. Toko buku berperan sebagai penyedia informasi, sumber ilmu pengetahuan, dan sarana penunjang proses belajar mengajar[2]. Dengan perkembangan teknologi, beberapa toko buku ternama menggunakan internet sebagai media penjualan buku, seperti Tokopedia dan Shopee[3]. Namun, kelengkapan dan keragaman koleksi buku di platform e-commerce seringkali tidak terjamin, dan kebijakan yang berbeda antar platform dapat menyebabkan kesenjangan dalam ketersediaan buku[4]. Oleh karena itu, konsumen sering kali harus melakukan pencarian ke berbagai toko buku online untuk mendapatkan buku yang diinginkan.

Berdasarkan latar belakang yang telah dijelaskan diatas, maka pada kesempatan kali ini penulis melakukan penelitian dengan topik data crawling yang berjudul "PERANCANGAN APLIKASI DATA CRAWLING UNTUK Mencari Buku dari Toko Buku Online". Serta membangun sebuah *search engine* dengan tujuan untuk pencarian buku berdasarkan sumber data dari toko buku online yang tersedia, sehingga diharapkan dapat memudahkan pengguna untuk menemukan buku yang tepat.

II. KAJIAN TEORI

Pada bagian kajian teori ini berisikan beberapa proses pengerjaan sistem yang di bangun. Berikut adalah beberapa point penjelasan terkait pengembangan aplikasi *data crawling*.

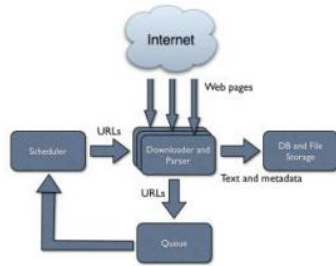
A. Data Crawling

Data *crawling* adalah teknik otomatis untuk mengumpulkan dan mengindeks data dari berbagai sumber online, menggunakan program yang disebut "*crawler*" atau "*spider*" untuk menjelajahi halaman web, mengikuti tautan, dan mengumpulkan data. Data yang diperoleh bisa digunakan untuk analisis data, penelitian, atau pengembangan sistem informasi[5].

Data *crawling* dan web *crawling* mempunyai tujuan yang sama yaitu mengumpulkan data dari sumber online, namun perbedaan dari keduanya:

1. Sumber data : web *crawling* lebih fokus pada data dari internet seperti situs web, sedangkan data *crawling* dapat melibatkan data dari database, file, hingga API.
2. Skala : web *crawling* biasanya dilakukan pada skala besar, sementara data *crawling* bisa skala kecil maupun besar tergantung kebutuhan.

Data *Crawling* dapat diotomatisasi sebagai bot atau robot untuk mengambil data dari halaman web tertentu.



Gambar 1.
Proses umum data *crawling*

B. BeautifulSoup

Beautifulsoup adalah salah satu library dalam Bahasa pemrograman python yang digunakan untuk mengekstrak data dari file HTML. *Beautifulsoup* merupakan library yang populer untuk melakukan ekstraksi informasi dari suatu halaman *website*[6], dikarenakan *Beautifulsoup* menyediakan cara yang mudah untuk menavigasi dan mengekstrak data dari dokumen HTML maupun XML.

Beautifulsoup sangat berguna dalam proses data *crawling* pada suatu *website*, karena dapat membantu proses ekstraksi data secara spesifik dari *website* yang menjadi target *crawling*. *Beautifulsoup* memiliki prinsip kerja yaitu menganalisis dokumen HTML yang memenuhi spesifikasi dan memudahkan pengguna dalam melakukan ekstraksi data yang dibutuhkan.

C. MySql

MySQL merupakan Database Management System (DBMS) yang populer dan berfungsi sebagai relational database management system (RDBMS). Menurut Tuti Susilawati et al. (2020), mysql adalah RDBMS yang cepat, *user-friendly*, dan banyak digunakan untuk berbagai keperluan. MySQL tersedia secara gratis di bawah lisensi *General Public License* (GPL) [7].

MySQL adalah aplikasi open source yang berfungsi sebagai server basis data, memiliki kinerja yang cepat, andal, dan mudah dipahami serta digunakan oleh pengguna. MySQL bekerja dengan arsitektur client-server, memungkinkan pengguna untuk mengelola dan mengakses basis data dengan efisien.

D. Draw.IO

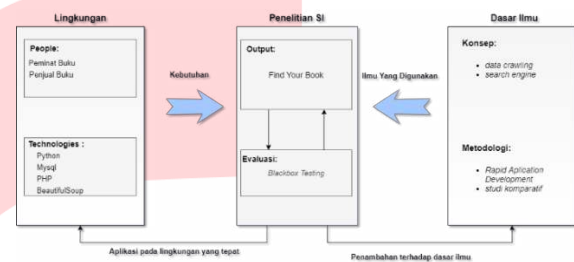
Draw.io adalah platform daring yang dirancang untuk membuat diagram. Platform ini dapat diakses menggunakan browser yang mendukung HTML5 dan koneksi internet. Draw.io terintegrasi dengan Google Drive, memungkinkan penyimpanan file dan ekspor dalam format JPG, PNG, SVG, dan XML[8].

III. METODE

Penelitian ini menggunakan Bahasa pemrograman *python* dengan menggunakan metode *Rapid Application Development* (RAD)

A. Model Konseptual

Model konseptual dalam penelitian ini adalah representasi keterkaitan antara berbagai konsep yang relevan dengan masalah yang diangkat. Model ini menunjukkan hubungan logis antara faktor dan variabel yang telah diidentifikasi. Dengan melibatkan beberapa aspek dan proses aktivitas, model konseptual ini bertujuan untuk mengkaji masalah penelitian secara mendalam dan sistematis, sehingga tujuan penelitian dapat tercapai dengan efektif.

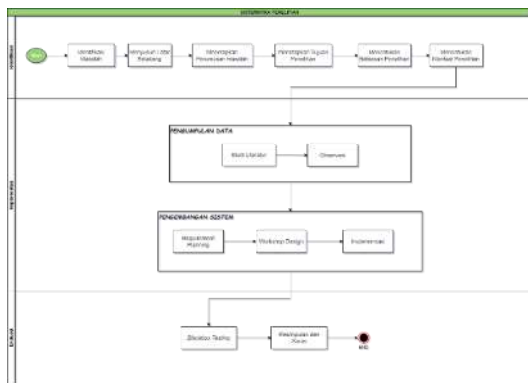


Gambar 2.
Model Konseptual

Model konseptual dalam penelitian ini mengidentifikasi peminat buku dan terbagi menjadi tiga bagian utama: *Environment* (lingkungan penelitian), *IS Research* (penelitian bidang sistem informasi), dan *Knowledge Base* (dasar ilmu). *Environment* terdiri dari *People* (peminat buku dan penjual buku online) dan *Technologies* (Python untuk data *crawling*, MySQL sebagai database, dan BeautifulSoup untuk ekstraksi data). *IS Research* berfokus pada penerapan teknologi dalam penelitian, sementara *Knowledge Base* mencakup teori dan metodologi yang relevan. Penelitian ini menggunakan metode *Rapid Application Development* (RAD) untuk membangun aplikasi *search engine* yang menyediakan informasi buku dari berbagai toko buku online, memudahkan pengguna menemukan buku dengan mudah dan komprehensif. Tujuan akhirnya adalah mengembangkan *website* pencarian buku *online* yang mengintegrasikan berbagai sumber toko buku, sehingga pengguna dapat dengan mudah menemukan buku yang mereka cari.

B. Sistematika Penelitian

Sistematika penelitian menggambarkan tahapan perancangan penelitian dalam menyelesaikan setiap permasalahan.



GAMBAR 3.
Sistematika penelitian

1. Pengumpulan data

Penelitian memerlukan data dan informasi lengkap untuk mendukung kebenaran materi dan pembahasan. Metode pengumpulan data meliputi :

a. Observasi

Peneliti melakukan pengamatan langsung pada website toko buku online yang menjadi target crawling pada penelitian kali ini untuk memahami setiap struktur HTML yang diperlukan untuk proses implementasi crawling.

b. Studi literatur

Peneliti membaca dan mengutip teori-teori yang relevan dari jurnal, dan sumber lain untuk dijadikan informasi pembandingan dalam penelitian.

2. Pengembangan sistem

Pengembangan sistem menggunakan metode *Rapid Application Development*, dimana metode ini memiliki beberapa tahapan pengembangan yaitu:

a. Requirements planning

Pada tahap ini mencakup beberapa tahapan yaitu:

- 1) Lingkup *crawling*
- 2) Identifikasi data yang diperlukan
- 3) Analisis struktur HTML
- 4) Perancangan sistem

b. Workshop design

- 1) Perancangan sistem : pembuatan *flowchart*, *usecase diagrams*, *activity diagrams*, untuk menggambarkan setiap alur dari program *crawling* dan *search engine* yang akan di buat.
- 2) Perancangan interface : desain *interface* yang efisien untuk memudahkan interaksi pengguna dengan *search engine* yang akan di buat.

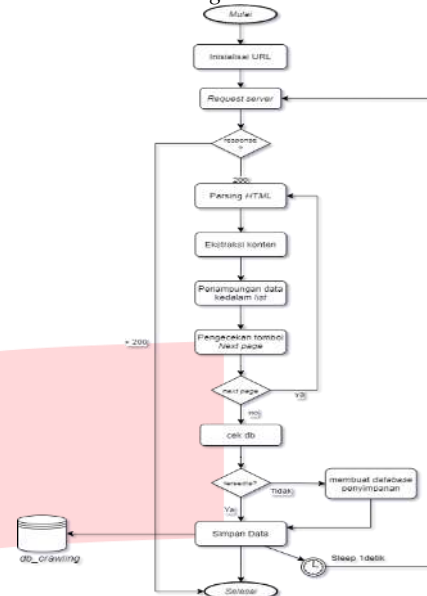
c. Implementation

Implementation merupakan tahap pengembangan sistem yang telah dirancang sebelumnya. Pengembangan kali ini akan melalui 2 tahap yaitu:

- 1) Pemrograman (*coding*): menuliskan setiap kode *Python* yang digunakan untuk mengekstraksi informasi. Dan *PHP* untuk pengembangan aplikasi *search engine*
- 2) Pengujian (*testing*): menguji setiap masing masing sistem menggunakan *Blackbox testing* untuk memastikan setiap

fungsi pada aplikasi *crawling* maupun *search engine* telah berfungsi dengan baik.

C. Flowchart sistem crawling



Gambar 4.
Data Crawling

Pada Gambar 6 terlihat langkah-langkah proses data *crawling*, berikut merupakan penjelasan dari setiap proses *crawling* yang dilakukan:

1. Inisialisasi URL

Langkah pertama dalam proses pengumpulan data atau *crawling* yaitu dengan inisialisasi URL. URL yang digunakan merupakan URL yang sudah ditentukan oleh pengguna untuk melakukan proses *crawling*. Dalam penelitian ini adalah webpage toko buku online yang berisikan daftar buku yang dijual.

2. Request Server

Langkah selanjutnya yaitu sistem akan mengirimkan permintaan HTTP untuk mengakses konten dari *webserver*.

3. Parsing HTML

Setelah melakukan permintaan HTTP pada *webserver*, dan mendapatkan response 200, maka sistem akan menguraikan setiap kode HTML dari halaman web agar dapat diidentifikasi dan mengekstrak setiap informasi yang diperlukan.

4. Ekstraksi Informasi

Langkah selanjutnya yaitu proses pengambilan data atau informasi tertentu dari halaman web yang dilakukan oleh sistem *crawling*. Pengambilan data atau informasi pada penelitian ini juga melibatkan setiap tag HTML yang berisikan informasi seperti judul buku, harga, link buku, gambar buku.

5. Tampung kedalam list

Setelah proses ekstraksi dan pemberishan data telah dilakukan sistem akan menampung data hasil *crawling* tersebut kedalam sebuah list yang didefinisikan dengan variabel "crawl".

6. Pengecekan Halaman

Setelah proses *crawling* pada halaman pertama berhasil dilakukan, maka sistem akan melakukan pengecekan berdasarkan tombol "next page" yang telah di definisikan pada sistem. Dimana jika terdapat tombol "next page" maka

sistem akan kembali melakukan proses crawling pada halaman berikutnya hingga halaman habis atau tidak terdapat tombol “next page”.

7. Pengecekan Database

Langkah berikutnya, adalah proses pengecekan database berdasarkan penamaan database yang di definisikan pada sistem crawling sebelumnya, jika pengecekan berdasarkan penamaan database belum tersedia maka sistem akan membuat database berdasarkan yang telah di definisikan. Jika database telah tersedia maka akan lanjut pada tahap selanjutnya.

8. Simpan data

Langkah terakhir pada sistem crawling ini yaitu sistem akan memasukan setiap data yang telah ditampung sebelumnya pada sebuah list Bernama “crawl” akan dimasukan pada database. Setelah semua proses selesai maka program akan melakukan perulangan untuk setiap proses yang telah dilalui.

D. Identifikasi aktor

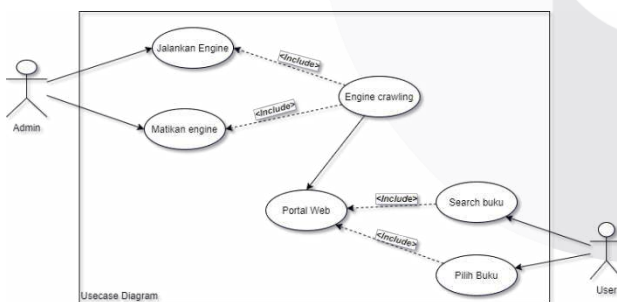
Identifikasi aktor kali ini bertujuan untuk menggambarkan setiap pengguna yang terlibat dalam sistem dan juga dapat memberikan gambaran setiap kerja sistem.

TABLE 1. Identifikasi Aktor

No.	Aktor	Deskripsi
1.	Admin	Admin merupakan orang yang berwajib untuk menjalankan hingga mematikan engine data crawling.
2.	User	User merupakan siapa saja yang menggunakan aplikasi search engine.

E. Usecase diagram

Melalui diagram usecase akan menggambarkan mengenai setiap interaksi yang terjadi antar sistem yang akan disajikan dalam bentuk diagram.



GAMBAR 5. Usecase Diagram

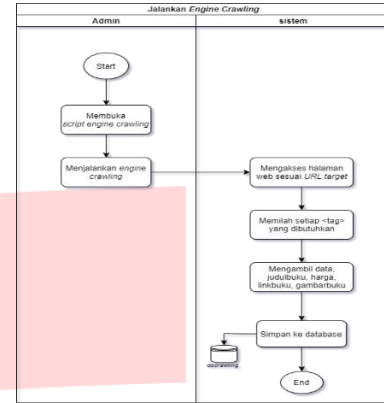
Pada Gambar 7, Usecase diagram menunjukkan setiap interaksi dari kedua aktor yang terlibat dalam sistem dimana setiap aktor yaitu Admin dan User. Admin bertugas untuk menjalankan dan menghentikan program data crawling. User dapat mencari buku yang tersedia pada aplikasi Find Your Book, berdasarkan database hasil crawling yang dilakukan sebelumnya.

F. Activity Diagram

Pada activity diagram akan digambarkan aktivitas yang berjalan pada sistem. Setiap aliran Kerja tersebut digambarkan secara grafis dalam suatu sistem.

1. Jalankan engine crawling

Aktivitas yang terjadi untuk menjalankan engine crawling akan disajikan dalam bentuk grafis agar dapat terlihat setiap alur kerja dari aktivitas tersebut.

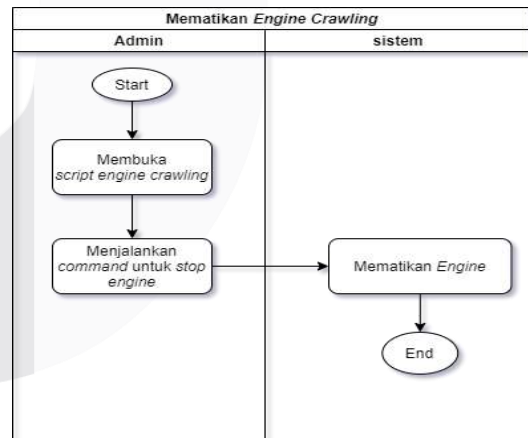


Gambar 6. Usecase Diagram jalankan engine

Pada Gambar 8 aktivitas menjalankan engine crawling dimulai dengan admin membuka dan menjalankan script engine crawling. Engine mengakses URL target, kemudian mengumpulkan data dengan menelusuri setiap tag HTML yang ditentukan dalam script.

2. Mematikan engine crawling

Aktivitas yang terjadi untuk mematikan engine crawling akan disajikan dalam bentuk grafis agar dapat terlihat setiap alur kerja dari aktivitas tersebut

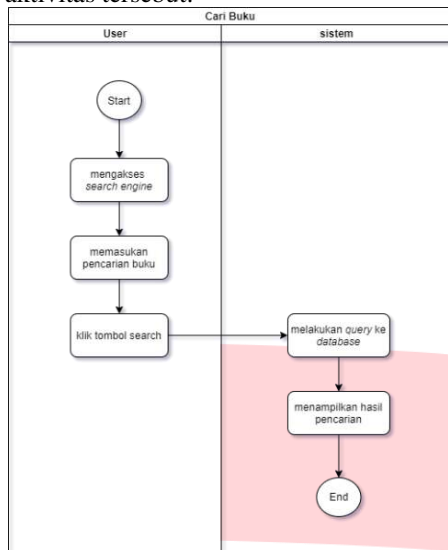


Gambar 7. Usecase Diagram matikan engine

Pada Gambar 9, aktivitas mematikan engine crawling dimulai dengan admin membuka script engine crawling. Admin kemudian menghentikan script dengan menekan Ctrl+C, dan sistem secara otomatis berhenti bekerja, sehingga engine crawling berhasil dimatikan.

3. Mencari Buku

Aktivitas yang terjadi saat *user* akan mencari buku akan disajikan dalam bentuk grafis agar dapat terlihat setiap alur kerja dari aktivitas tersebut.

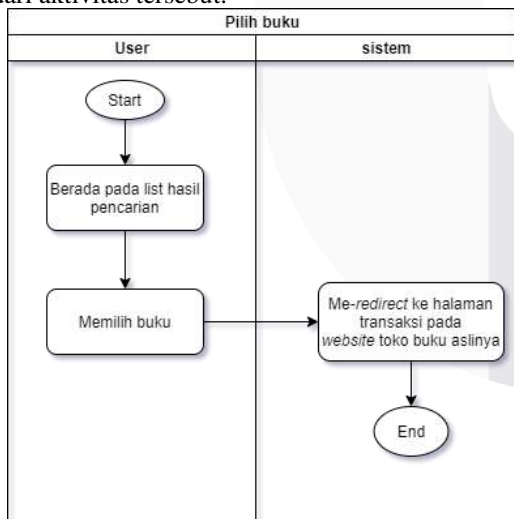


GAMBAR 8. Usecase Diagram search book

Pada Gambar 10, aktivitas search engine dimulai ketika user mengakses search engine dan memasukkan pencarian buku pada search bar. Setelah user mengklik tombol search, sistem melakukan query pada database yang berisi data hasil crawling. Hasil pencarian kemudian ditampilkan berdasarkan keyword yang dimasukkan oleh user.

4. Memilih Buku

Aktivitas yang terjadi saat *user* akan mencari buku akan disajikan dalam bentuk grafis agar dapat terlihat setiap alur kerja dari aktivitas tersebut.



GAMBAR 9. Usecase Diagram memilih buku

Pada Gambar 11, aktivitas dimulai ketika user memilih buku dari hasil pencarian berdasarkan keyword yang digunakan. Setelah user mengklik buku yang diinginkan, sistem akan langsung mengarahkan user ke halaman website asli dari toko buku tersebut.

IV. HASIL DAN PEMBAHASAN

Bab ini menjelaskan implementasi dari perancangan yang telah dilakukan pada bab sebelumnya terkait data crawling. Hasil implementasi merangkum setiap fungsi perangkat lunak yang dibuat dan akan diuraikan secara lengkap di bab ini.

A. Persiapan *crawling*

Untuk mengenali struktur konten web, penting memahami tag-tag HTML yang mengapit setiap elemen pada halaman. Setiap website memiliki struktur HTML khas, sehingga identifikasi tag-tag ini adalah langkah awal dalam pengenalan struktur web target.

TABLE 2. Deepublishstore

Data	Tag	Class
Judul buku	<h2>	“woocommerce-loop-product__title”
Harga		“woocommerce-Price-amount amount”
Linkbuku	<a>	“woocommerce-LoopProduct-link woocommerce-loop-product__link”
Gambarbuku		“attachment-woocommerce_thumbnail sizewocommerce_thumbnail”

TABLE 3. Grobmart

Data	Tag	Class
Judul buku	<div>	“name”
Harga		“Price-new”
Linkbuku	<a>	“name”
Gambarbuku		“Product-img”

Berdasarkan tabel 2 dan 3, informasi yang dibutuhkan diambil sesuai dengan pola tag HTML di setiap website target. Perbedaan penggunaan tag dan class di masing-masing website secara spesifik mendefinisikan setiap informasi yang ditampilkan. Identifikasi pola ini penting untuk proses crawling yang efektif.

B. Proses *Crawling*

Pada tahap ini, proses crawling dilakukan dengan mengambil informasi dari website target melalui tag-tag HTML yang telah ditentukan sebelumnya berdasarkan struktur website. Proses ini dijelaskan lebih lanjut pada gambar di bawah.

```
for item in soup.findAll("li"): # inget text strip lagi dihapus
# nama
nama_books = item.find(
    "h2", class_="woocommerce-loop-product_title")
nama_buku = nama_books.text.strip() if nama_books else None
# harga
prices = item.find("span", "woocommerce-Price-amount amount")
price = prices.text.strip() if prices else None
# link
links = item.find(
    "a", "woocommerce-LoopProduct-link woocommerce-loop-product_link")
link_buku = links["href"] if links and 'href' in links.attrs else None
# gambar
gambar = item.find(
    "img", "attachment-woocommerce_thumbnail size-woocommerce_thumbnail")
gambar_buku = gambar["src"] if gambar and 'src' in gambar.attrs else None
```

GAMBAR 10. Crawling Deepublishstore

```
for item in soup.findAll("div", "product-thumb,"):
# judul
nama_books = item.find("div", "name")
nama_buku = nama_books.text.strip() if nama_books else None

# harga
prices = item.find("span", "price-new")
price = prices.text.strip() if prices else None

# Link
links = item.find("div", class_='name')
link_buku = None
if links:
    a_tag = links.find('a')
    if a_tag and 'href' in a_tag.attrs:
        link_buku = a_tag['href']

# image
img = item.find('a', class_='product-img')
gambar_buku = None

if img:
    b_tag = img.find('img')
    if b_tag and 'data-src' in b_tag.attrs:
        gambar_buku = b_tag['data-src']

# menambahkan kedalam list
if None not in [nama_buku, price, link_buku, gambar_buku]:
    crawl.append([nama_buku, price, link_buku, gambar_buku])
```

GAMBAR 11. Crawling Grobmart

Mengacu pada Gambar 12 dan 13, setelah URL dan tag ditentukan, langkah selanjutnya adalah memasukkan URL tersebut ke dalam program crawling. Selain URL, parameter informasi yang dibutuhkan juga dimasukkan ke dalam aplikasi crawling sesuai yang telah ditentukan sebelumnya, untuk mendapatkan konten yang dimuat dalam setiap tag pada masing-masing website target.

C. Implementasi Crawling

Proses pengambilan data buku dari toko buku online target crawling melibatkan beberapa tahap.

```
PS D:\W_ TUGAS AKHIR\codingan\CRAWLING BUKU ONLINE\yang dipake> C:\Python\python.exe d:/B_ TUGAS AKHIR/BUKU ONLINE/ yang dipake/deep.py
Memulai crawling...
halaman habis
response to server <response [200]>
pengecekan db
Database multibookstore sudah tersedia.
Tabel deepublishstore sudah tersedia.
koneksi DB berhasil
Data lama berhasil dihapus
berhasil memasukan data kedalam database
Memulai crawling...
halaman habis
response to server <response [200]>
```

GAMBAR 12. Proses crawling deepublishstore

```
PS D:\W_ TUGAS AKHIR\codingan\CRAWLING BUKU ONLINE\yang dipake> C:\Python\python.exe d:/B_ TUGAS AKHIR/codingan/CRAWLING BUKU ONLINE/ yang dipake/grobmart.py
Memulai crawling...
response to server <response [200]>
halaman habis
pengecekan db
Database multibookstore sudah tersedia.
Tabel grobmart sudah tersedia.
koneksi DB berhasil
Data lama berhasil dihapus
berhasil memasukan data kedalam database
Memulai crawling...
```

GAMBAR 13. Proses crawling grobmart

Mengacu pada Gambar 14 dan 15, setelah engine crawling dijalankan, program akan mengirimkan HTTP request ke web server dan menjalankan proses pengambilan data sesuai yang telah ditentukan. Setelah seluruh data dari website target diambil, program akan memeriksa database yang telah didefinisikan. Apakah database beserta tabel yang telah di definisikan pada script sudah tersedia apa belum, jika belum maka akan dibuat terlebih dahulu, jika sudah selanjutnya akan di cek lagi Jika terdapat data di dalamnya, data tersebut akan dihapus terlebih dahulu, kemudian data hasil crawling akan dimasukkan ke dalam database yang telah dibuat sebelumnya. Data ini akan digunakan pada aplikasi search engine yang akan dikembangkan.

D. Hasil Implementasi

Hasil proses crawling yang dilakukan pada masing-masing website yang menjadi target akan disimpan kedalam sebuah database yang telah dibuat sebelumnya pada program crawling.

Tabel	Tindakan	Baris	Jenis	Penyortiran	Ukuran	Beban
deepublishstore	Jelajah Struktur Cari Tambahkan Kosongkan Hapus	6,308	InnoDB	utf8mb4_general_ci	2.5 MB	-
grobmart	Jelajah Struktur Cari Tambahkan Kosongkan Hapus	25,485	InnoDB	utf8mb4_general_ci	8.5 MB	-

GAMBAR 14. Database crawling

Mengacu pada Gambar 5.1, proses crawling berhasil mengumpulkan total 31.785 data. Dari jumlah tersebut, Grobmart menyumbang 6.380 data yang disimpan dalam database, sedangkan Deepublishstore menyumbang 25.405 data. Setiap data yang dikumpulkan berisi informasi terkait buku yang dijual di masing-masing toko buku online yang menjadi target crawling.

id	nama_buku	harga	link_buku	gambar_buku
1	Buku Buku Tangkis Dasar	61 000	https://deepublishstore.com/shop/buku-buku-tangkis	https://deepublishstore.com/wp-content/uploads/201...
2	Buku Belajar Microsoft Office (Word, Excel, P...	108 000	https://deepublishstore.com/shop/buku-belajar-micr...	https://deepublishstore.com/wp-content/uploads/201...
3	Buku Eduart a Project Based Learning 2	84 000	https://deepublishstore.com/shop/buku-eduart1	https://deepublishstore.com/wp-content/uploads/202...
4	Buku Eduart a Project Based Learning	65 000	https://deepublishstore.com/shop/buku-eduart-a-pro...	https://deepublishstore.com/wp-content/uploads/202...
5	Buku Kumpulan Soal Matematika Bisnis	61 500	https://deepublishstore.com/shop/buku-soal-matemat...	https://deepublishstore.com/wp-content/uploads/201...
6	Buku Pengetahuan Kependidikan	72 000	https://deepublishstore.com/shop/buku-kependidikan	https://deepublishstore.com/wp-content/uploads/202...
7	Buku Hukum Waris Islam	82 000	https://deepublishstore.com/shop/buku-hukum-waris-	https://deepublishstore.com/wp-content/uploads/202...

GAMBAR 15. Tabel Deepublishstore

id	nama_buku	harga	link_buku	gambar_buku
1	"Ummat, Ummat, Ummat..." (Betapa Nabii Kange...	77 000	https://www.grobmart.com/Buku/ummat-ummat-ummat-	https://www.grobmart.com/image/cache/catalog/00sel...
2	#AboutLife	65 100	https://www.grobmart.com/Buku/aboutlife-978020530...	https://www.grobmart.com/image/cache/catalog/00sel...
3	#BukuAntiKorupsi	89 600	https://www.grobmart.com/Buku/bukuantikorupsi-978923	https://www.grobmart.com/image/cache/catalog/00sel...
4	#BukuMamaBisa BIKIN PLAYDATE SENDIRI	87 120	https://www.grobmart.com/Buku/bukumamabisa-bikin-p...	https://www.grobmart.com/image/cache/catalog/00sel...
5	#CintaPenunBata	52 500	https://www.grobmart.com/Buku/cintapenunbata-978...	https://www.grobmart.com/image/cache/catalog/00prod...
6	#HappyNewNormal	53 200	https://www.grobmart.com/Buku/happynormal-97862	https://www.grobmart.com/image/cache/catalog/00sel...
7	#Korban Phip	18 765	https://www.grobmart.com/Buku/korban-phi-978602770	https://www.grobmart.com/image/cache/catalog/00sel...
8	#Muladirent Mari Investasi Saham Dengan EBsen	69 300	https://www.grobmart.com/Buku/muladirent-mari-in-	https://www.grobmart.com/image/cache/catalog/00sel...
9	#NeshaBuruk	59 500	https://www.grobmart.com/Buku/neshaburuk-9789230	https://www.grobmart.com/image/cache/catalog/00sel...

GAMBAR 16. Tabel Grobmart

Berdasarkan Gambar 17 dan 18, engine crawler yang dibuat dalam penelitian ini berhasil mengumpulkan data sesuai yang diharapkan. Field gambar buku berisi URL yang memuat gambar dari masing-masing website target. Data ini akan mendukung pengembangan aplikasi berikutnya, yaitu search engine.

E. Pengujian Aplikasi

Pada penelitian kali ini menggunakan dua pengujian yaitu pengujian validasi data guna memastikan data yang berhasil


di *crawling* berhasil dan sesuai dengan harapan, pengujian yang berikutnya yaitu menggunakan pengujian *Blackbox* dimana pengujian ini bertujuan untuk memastikan setiap fungsi dari sistem yang telah dibuat dapat berjalan sesuai tujuan.

1. Validasi Data

Dilakukan pengujian keakuratan aplikasi *crawling* untuk mengumpulkan informasi dari Grobmart dan Deepublishstore. Pengujian melibatkan validasi dengan membandingkan data hasil *crawling* dengan informasi pada *website* target.

Deepublishstore	Field	Value	Validasi
	namabuku	Bulu Tangkis Dasar	Sesuai
	harga	61.000	Sesuai
	linkbuku	https://deepublishstore.com/shop/buku-bulu-tangkis-dasar/	Sesuai
	gambarbuku	https://deepublishstore.com/wp-content/uploads/2017/07/Bulutangkis-Dasar-Pl_Dhedhy-Yulianwan-depan-300x400.jpg	Sesuai

GAMBAR 17. Validasi Deepublishstore

Grobmart	Field	Value	Validasi
	namabuku	#AboutLife	Sesuai
	harga	65.100	Sesuai
	linkbuku	https://www.grobmart.com/Buku/aboutlife-9786020630212	Sesuai
	gambarbuku	https://www.grobmart.com/image/cache/catalog/00seller00/201905/9786020630212-250x250h.jpg	Sesuai

GAMBAR 18. Validasi Grobmart

Berdasarkan pengujian dari masing-masing sample data hasil *crawling* diatas dapat dikatakan berhasil karena dari sampel data yang didapat sesuai dengan sumber *website* yang menjadi target *crawling* pada penelitian kali ini.

2. Blackbox Testing

Pengujian yang berikutnya merupakan pengujian *Blackbox* dimana pada pengujian ini akan dilakukan pengujian dari setiap fungsi yang dimiliki pada masing-masing aplikasi yaitu pada aplikasi *crawling* dan aplikasi *search engine*.

TABLE 4. Blackbox Testing

No.	Prosedur Pengujian	Masukan	Hasil yang diharapkan	Hasil
1. Pengujian <i>Crawling</i>				
1.1.	Menjalankan program <i>Crawling</i>	Klik Run pada <i>text editor</i> , masukan <i>command</i> untuk menjalankan program	Program dapat berjalan dan berhasil mengumpulkan data	OK
2. Pengujian Akses				
2.1.	Mengakses <i>interface</i> aplikasi	User menginputkan URL aplikasi	Aplikasi Dapat diakses	OK
3. Pengujian Pencarian				
3.1.	Mencari buku dengan penulisan yang benar	Ketik <i>keyword</i> pada <i>searchbar</i> , lalu tekan	Hasil pencarian muncul	OK

		enter pada keyboard		
3.2.	Mencari buku dengan kesalahan penulisan	Ketik <i>keyword</i> pada <i>searchbar</i> , lalu tekan enter pada keyboard	Hasil Pencarian tidak muncul	OK
4. Pengujian Pemilihan Buku				
4.1.	Klik buku dari Deepublishstore	Klik button <i>Book details</i>	Redirect ke <i>website Deepublishstore</i> , daftar informasi buku yang ditampilkan sesuai	OK
4.2.	Klik buku dari Grobmart	Klik button <i>Book details</i>	Redirect ke <i>website Grobmart</i> , daftar informasi buku yang ditampilkan sesuai	OK

V. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan maka dapat ditarik kesimpulan yaitu :

1. Aplikasi *crawling* yang dikembangkan dalam penelitian ini efektif dalam mengumpulkan dan menyajikan informasi dari daftar buku di toko buku Grobmart dan Deepublishstore, menunjukkan potensi sebagai alat yang berguna dalam proses *crawling* data.
2. Search engine "Find Your Book" memungkinkan pengguna menemukan buku berdasarkan pencarian dan menyediakan beragam pilihan buku. Search engine ini mengakses data dari database hasil *crawling*, memberikan informasi yang lebih komprehensif kepada pengguna.

Beberapa Saran yang dapat dipertimbangkan untuk menunjang keberlangsungan penelitian mengenai topik ini adalah sebagai berikut :

1. Penelitian berikutnya diharapkan dapat mengembangkan sistem data *crawling* yang mencakup lebih banyak *website*, sehingga data yang diperoleh menjadi lebih luas dan beragam.
2. Aplikasi ini dapat diterapkan pada objek lain seperti pencarian tiket pesawat, pemesanan kamar, dan penerapan lainnya dengan konsep serupa.
3. Penelitian selanjutnya perlu mengembangkan metode data *crawling* baru untuk mengatasi teknologi keamanan yang mendeteksi dan memblokir aktivitas crawler pada beberapa *website*.

REFERENSI

[1] N. Nafi'iyah and E. Sulistiono, "PEMANFAATAN ROBOT CRAWLER PADA PEMBUATAN TOKO BUKU ONLINE," 2016.

- [2] Nugroho Wandu, Rully Hendrawan, and Ahmad Mukhlason, "PENGEMBANGAN SISTEM REKOMENDASI PENELUSURAN BUKU DENGAN PENGGALIAN ASSOCIATION RULE MENGGUNAKAN ALGORITMA APRIORI (STUDI KASUS BADAN PERPUSTAKAAN DAN KEARSIPAN PROVINSI JAWA TIMUR)," 2012.
- [3] Siti Humayroh, "Maraknya Penjualan buku Bajakan di E-Commerce," Dec. 2022, Accessed: Dec. 18, 2023. [Online]. Available: <https://kumparan.com/siti-humayroh-1670211402564137241/maraknya-penjualan-buku-bajakan-di-e-commerce-1zOxx9LBqD9>
- [4] Rehni Jayana Purba, "SISTEM INFORMASI BERBASIS WEB PADA TOKO BUKU DI KOTA BATAM," 2021.
- [5] R. (Ryan E.) Mitchell, *Web scraping with Python : collecting more data from the modern web*. 2018.
- [6] A. Purnomo, "Implementasi Web Scraping Pada OJS Dengan Metode CSS Selector," *Media Online*, 2022.
- [7] A. Hidayat, A. Yani, P. Studi Sistem Informasi, and S. Mahakarya, "MEMBANGUN WEBSITE SMA PGRI GUNUNG RAYA RANAU MENGGUNAKAN PHP DAN MYSQL," 2019.
- [8] E. Suharyanto, "PERANCANGAN APLIKASI PENGENALAN BUDAYA NUSANTARA BERBASIS ANDROID DENGAN METODE RAD," *Jurnal Ilmu Komputer JIK*, p. 2022, 2022.