

Web Scraping Situs Berita Menggunakan Bahasa Pemrograman Python

Muhammad Ramadan Fikri¹, Rahmadya Trias Handayanto^{1,*}, Dadan Irwan¹

* Korespondensi: e-mail: rahmadya_handayanto@unismabekasi.ac.id

¹ Teknik Komputer; Universitas Islam 45; Jalan Cut Meutia No.83 Bekasi Timur, Kota Bekasi; e-mail: ramaskets@gmail.com, rahmadya_handayanto@unismabekasi.ac.id, dadanirwan@gmail.com

Submitted : 31 Maret 2022
Revised : 25 April 2022
Accepted : 8 Mei 2022
Published : 30 Mei 2022

DOI:
<https://doi.org/10.31599/jsracs.v3i1.1514>

Abstract

Currently the rapid development of technology provides innovation, one of which is the technique of obtaining information from portal websites, termed web scrapers. This application provides data needs in the form of information where the process of retrieving information from sites will later be taken to observe behavior and perceptions to get the right segmentation. Most data collection is currently still done manually, as a result, this method has several system limitations, namely the length of the data collection process so that it slows down the performance of market segment analysis. The risk is not getting the right market segmentation. To solve this problem, a web scraping news site is needed. In this study, web scraping news sites were created using the python programming language and the flask library to display web scraping. In addition, the Selenium library is used to simplify application creation, facilitate interaction with the web and provide facilities to control a web browser. This program can retrieve data based on keywords, where the results are in the form of the title, posting date, summary, then collect the data that has been taken into a csv file extension automatically.

Keywords: Internet, News, Python, Scraping, Website

Abstrak

Saat ini perkembangan pesat teknologi memberikan inovasi, salah satunya adalah teknik memperoleh informasi dari situs web portal, yaitu *web scraper*. Aplikasi ini menyediakan kebutuhan data berupa informasi dimana proses pengambilan informasi dari situs-situs nantinya diambil untuk diamati perilaku dan persepsi untuk mendapatkan segmentasi yang tepat. Kebanyakan pengambilan data saat ini masih dilakukan secara manual, akibatnya cara ini memiliki beberapa keterbatasan sistem yaitu lamanya proses pengumpulan data sehingga memperlambat kinerja analisa segmen pasar. Resikonya adalah tidak mendapatkannya segmentasi pasar yang tepat. Untuk mengatasi masalah tersebut diperlukan *web scraping* situs berita. Pada penelitian ini, *web scraping* situs berita dibuat dengan menggunakan bahasa pemrograman *python* dan library *flask* untuk tampilan *web scraping*. Selain itu, library *selenium* digunakan untuk mempermudah pembuatan aplikasi, mempermudah interaksi dengan *web* dan menyediakan fasilitas untuk mengontrol suatu peramban *web*. Program ini dapat mengambil data berdasarkan kata kunci, dimana hasilnya berupa judul, tanggal postingnya, rangkuman, lalu mengumpulkan data yang telah di ambil ke *file* berekstensi *csv* secara otomatis.

Kata kunci: Berita, Internet, Python, Scraping, Website

1. Pendahuluan

Sebagian besar data pada internet adalah informasi yang setiap harinya terus bertambah, untuk mengumpulkan informasi sesuai dengan kata kunci yang dimasukkan dan pengindeksan *web* yang dipakai hampir semua *search engine* melalui penyaringan ke semua halaman internet diperlukan program yang dapat memperoleh informasi sesuai kebutuhan dapat dilakukan secara otomatis (Mitra et al., 2017; Ramadhan & Sastramihardja, 2018; Rohman et al., 2019). Dengan teknologi informasi penyebaran informasi begitu cepat seperti melalui situs *web* memberikan manfaat memudahkan mendapatkan berbagai informasi yang diinginkan khususnya berita (Dewantoro, 2016; Yunita, 2020). Salah satu perkembangan teknologi berupa *website* berita *scraper* yang menyediakan kebutuhan data berupa informasi, proses pengambilan data informasi dari situs-situs diambil untuk diamati perilaku dan persepsi agar mendapatkan segmentasi yang tepat.

Web Scraper

Teknik mendapatkan informasi dari *website* secara otomatis tanpa harus menyalinnya secara manual. *Web scraping* berfokus dalam mendapatkan data dan informasi dapat berupa, teks, tautan dengan cara pengambilan dan ekstraksi untuk diambil data tertentu dari halaman tersebut agar bisa digunakan kembali oleh sistem lain maupun di analisis lebih lanjut (Arisandi et al., 2021; Flores et al., 2020; Rahmatulloh & Gunawan, 2020). Umumnya data yang diperoleh berupa laman *web* dokumen *HTML* dengan memilih bagian tertentu, kemudian di transformasi dari bentuk yang tidak terstruktur dalam format *HTML* menjadi format data terstruktur dan disimpan dalam format data tertentu (Priyanto & Ma'arif, 2018; Satriajati et al., 2021). Dalam *web scraping* memiliki langkah: a) *Create scraping template* mempelajari dokumen *HTML* dari *website* yang akan diambil informasinya, b) *Explore site navigation* navigasi pada *website* yang akan diambil informasinya ditirukan pada aplikasi *web scraper*, c) *Automate navigation and extraction* berdasarkan informasi yang telah diperoleh aplikasi *web scraper* dibuat untuk mengotomatisasi pengambilan informasi dari *website* yang ditentukan, d) *Extracted data and package history* informasi yang didapat disimpan dalam tabel *database* (Ayani et al., 2019).

Website Portal Berita

Portal berita sebagai situs halaman *web* yang berisi mengenai berbagai jenis berita, mulai dari berita seputar politik, dalam negeri, luar negeri, olahraga hingga berita ringan. Portal berita terdiri dari dua kata, yaitu portal atau halaman *web* dan berita informasi terbaru mengenai sesuatu yang sedang terjadi yang disajikan melalui bentuk cetak, siaran, internet, atau dari mulut ke mulut kepada orang banyak. Dengan pesatnya arus berita, diperlukan metode yang lebih cepat dan efisien untuk mendapatkan topik salah satunya dengan *scraping* mengambil artikel dari portal berita dan kemudian diolah (Dewantoro, 2016).

Sistem *website* portal berita *scraper* mampu menemukan hasil pencarian di beberapa situs secara bersamaan (Ayani et al., 2019) dengan menginputkan *keyword* mengenai topic artikel yang dicari oleh pengguna dan seperti mesin pencari proses pengumpulan data *website* berita *online* yang *URL*-nya diketikkan oleh pengguna, kemudian sistem akan memberikan

request kepada *server* untuk mencari tautan yang berhubungan dengan *keyword* tersebut, dari *URL* artikel berita kemudian disimpan ke dalam *database*, hasil *keyword* dari proses *scraping* yang telah di temukan sistem secara otomatis mengunduh sebuah file berekstensi *csv* untuk memudahkan pemindahannya dari satu program keprogram lain (Flores et al., 2020; Indra & Trisnawati, 2018).

Bahasa Pemrograman Python

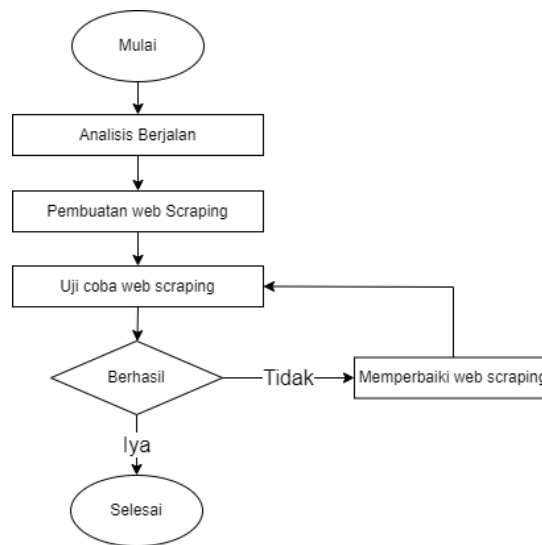
Merupakan bahasa pemrograman interpretatif multiguna bisa melakukan eksekusi sejumlah instruksi multi guna secara langsung (interpretatif) dengan metode *Object Oriented Programming* dan juga menggunakan semantik dinamis yang berfokus pada tingkat keterbacaan *syntax* (Syaefulloh, 2020). *Python* di klaim sebagai bahasa yang menggabungkan kapabilitas, kemampuan, dengan sintaksis kode yang sangat jelas, dan dilengkapi dengan fungsionalitas pustaka standar yang besar serta komprehensif.

Beberapa penelitian terdahulu yang dilakukan tentang *web scraping* antara lain: penelitian oleh Mitra *et al* pada 2017 pada situs *marketplace* meliputi proses penyebaran, pembelian, penjualan, pemasaran barang dan jasa melalui internet yang melibatkan pertukaran data elektronik, sistem manajemen inventori otomatis, transfer dana elektronik, dan sistem pengumpulan data otomatis. Agar pembeli dapat mengambil keputusan yang terbaik ketika perbelanjaan produk penjualan terbanyak atau terlaris perlu melakukan pencarian di beberapa situs secara manual. Dengan teknik *web scraping*, melakukan pencarian di beberapa situs *marketplace* dan menampilkan hasil secara bersamaan, hasil pengujian menunjukkan sistem memberikan hasil terbaik pencarian produk sesuai kata kunci dari pembeli pada gabungan hasil pencarian di tiga situs web *marketplace* metode *HTML DOM* proses *scraping*, tokenisasi, *cleaning*, dan *lowercased*, semua proses tersebut berjalan secara otomatis sehingga sangat menghemat waktu dan biaya (Mitra et al., 2017). Penelitian lain oleh Ayani *et al* pada 2019 pengumpulan secara manual memerlukan waktu lama dan biaya tidak sedikit. Aplikasi *web scraping* dibuat mengkombinasikan dengan metode *HTML DOM* adalah proses *scraping*, tokenisasi, *cleaning*, dan *lowercased*, semua proses tersebut berjalan secara otomatis, Pengujian dengan *blackbox* diperoleh hasil mengambil data *web scraping* mampu melakukan pencarian di beberapa situs *marketplace* dan menampilkan hasil pencarian secara bersamaan sehingga sangat menghemat waktu dan biaya (Ayani et al., 2019). Penelitian oleh Flores *et al* pada 2020 untuk proses pencarian artikel ilmiah seringkali memakan waktu dikarenakan banyaknya data yang ada di Internet. Dengan *Depth First Search (DFS)* sebagai metode pencarian yang dimanfaatkan untuk melakukan *web scraping* di internet, sistem ini bekerja dengan mengolah *keyword* mengenai topik artikel ilmiah yang diberikan oleh pengguna. Kemudian sistem akan mengirimkan *request* kepada *server* untuk mencari tautan yang berhubungan dengan *keyword* tersebut. Setelah ditemukan otomatis akan diunduh artikel berformat *pdf*. Hasil pengujian 134 artikel yang diunduh menunjukkan sistem bekerja sangat baik dengan nilai *precision* 0.87 dan *recall* 0.99 (Flores et al., 2020).

Berdasarkan penelitian terdahulu yang telah dilakukan, pengambilan data secara manual memiliki keterbatasan lamanya waktu mengumpulkan data memperlambat kinerja pengguna serta tidak mendapatkannya segementasi yang tepat sehingga diperlukan sebuah aplikasi *web scraping* yang dapat berjalan otomatis dalam mengumpulkan data dari internet sesuai dengan kebutuhan pengguna dalam waktu bersamaan. Penelitian ini berkontribusi dalam mengatasi masalah pengumpulan data, khususnya berita, dengan sebuah aplikasi *web scraping* dengan bahasa pemrograman *Python*.

2. Metode Penelitian

Penelitian dalam membuat *web scraping* situs berita menggunakan bahasa pemrograman *python* seperti pada Gambar 1.



Sumber: Hasil Penelitian (2022)

Gambar 1. Rancangan Penelitian

Analisis Sistem Berjalan

Melakukan pengumpulan data dengan cara manual risiko yang dapat terjadi yaitu lamanya mengumpulkan data menghambat kinerja dan kurang akuratnya pengumpulan data yang dibutuhkan karena kurang telitinya dalam pencarian data informasi pada situs *web* berita. Adapun solusi dari analisis sistem berjalan untuk mengatasi beberapa keterbatasan ini yaitu membuat *web scraping* situs berita untuk dapat menyediakan kebutuhan data berupa informasi yang terdapat pada *website* portal berita, dari data yang diperoleh akan diambil dan diproses lebih lanjut untuk diamati perilaku dan persepsi dari suatu segmentasi pasar yang telah diamati sebelumnya untuk mendapatkan target pasar yang tepat. Program dibuat dengan bahasa pemrograman *python* serta bantuan *framework selenium* untuk pengambilan data pada *web site* berita, hasil dari proses *scraping* menghasilkan file berekstensi *csv*. Setelah dilakukan analisis sistem berjalan proses selanjutnya yaitu perancangan sistem sebagai tahapan membuat *website scraping*. Pembuatan sistem diawali dengan *flowchart* yang menggambarkan proses pengambilan data menggunakan *web scraping*.

Pembuatan Web scraping

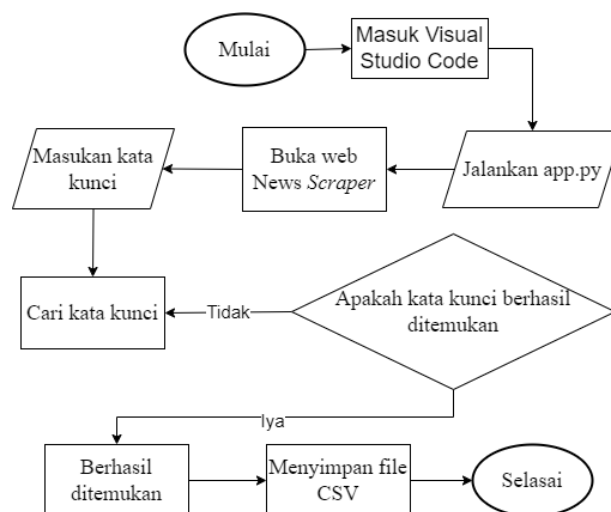
Pada pembuatan *web scraping* beberapa tahapan yang dilakukan antara lain pembuatan *front end*, pencarian data, dan *DOM parsing*. Pembuatan *front end* sebagai tahapan awal tampilan *web scraping* bertujuan untuk mempermudah interaksi dibangun menggunakan bahasa pemrograman *python* memanfaatkan *flask*. Pencarian data memasukan *list* situs *web* berita yang di *scraping* untuk pengambilan informasi data berita, yaitu <http://www.kompas.com>, <http://www.detik.com>, <http://www.kapanlagi.com>, dan <http://www.tribunnews.com>. Data yang telah di *scraping* dari empat situs tersebut file disimpan berbentuk *csv*, data pada *file csv* meliputi pencarian judul, *news* portal, waktu *post* dan rangkuman. *DOM parsing* mengetahui cara kerja internal halaman *website* dan mengekstrak *script* yang berjalan didalamnya, untuk melakukan *webs craping* menggunakan *parsing* DOM (*Document Object Model*) untuk mendapatkan informasi yang dibutuhkan. Dengan bantuan *web browser*, progam dapat mengakses *dynamic content* dari *script client-side* yang sudah dibuat.

Uji Coba Web scraping

Pengujian *web scraping* terdiri dari pengujian mencari kata kunci yang diinginkan oleh pengguna. Pengujian kata kunci bertujuan untuk memastikan apakah kata kunci sudah benar-benar mendapatkan hasil yang sesuai dari kata kunci yang diinginkan. Setelah memperoleh hasil dari pengujian *web scraping* yang telah dibuat, apakah berjalan sesuai atau tidak dengan yang diharapkan.

3. Hasil dan Pembahasan

Langkah awal dilakukan pembuatan *flowchart* seperti pada Gambar 2. *Flowchart* menjelaskan arah alur pada *web scraping*, langkah awal masuk ke *visual studio code*, menjalankan program *app.py* untuk membuka *news scraper*, memasukan kata kunci yang diinginkan, apabila kata kunci berhasil ditemukan maka dapat menyimpan *file csv* sesuai pencarian kata kunci.

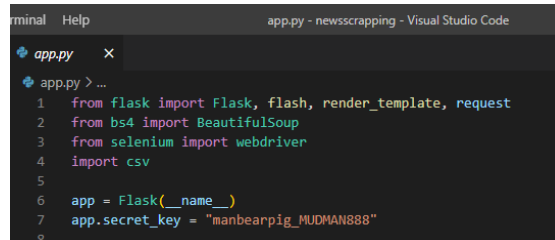


Sumber: Hasil Penelitian (2022)

Gambar 2. Flowchart Web Scraping Portal Berita

Mengedit Kode Program dengan Visual Studio Code

Untuk menjalankan *web scraping* buka *visual studio code* lalu pilih *file* jalankan dengan *Open in Intergrated Terminal*, ketik *script Set-Execution Policy Unrestricted-Scope Process* *enter*, masukan kembali *scriptenv/Scripts/activate* *enter*, masukan kembali *script python -m flask run* *enter*, muncul *ip link web scraping* <http://127.0.0.1:5000> pada *script terminal* tekan *ctrl+klik ip* tujuan maka akan langsung *running* pada *web scraping* pada Gambar 3.



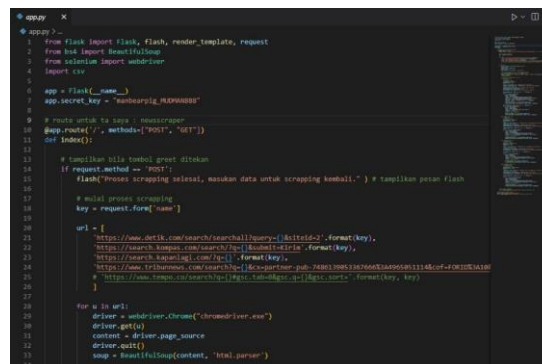
```
terminal Help app.py - newsscraping - Visual Studio Code
app.py x
app.py > ...
1 from flask import Flask, flash, render_template, request
2 from bs4 import BeautifulSoup
3 from selenium import webdriver
4 import csv
5
6 app = Flask(__name__)
7 app.secret_key = "manbearpig_MUDDMAN888"
8
```

Sumber: Hasil Penelitian (2022)

Gambar 3. Tampilan Kode untuk Mengakses Flask

Menjalankan Server Flask dengan Kode *App.py*

Pada gambar 4 membuat *App.py* memakai bahasa pemrograman *python* dengan *flask*, *framework*, *selenium*, *webdriver*, dan bertujuan untuk *scraping* situs berita.



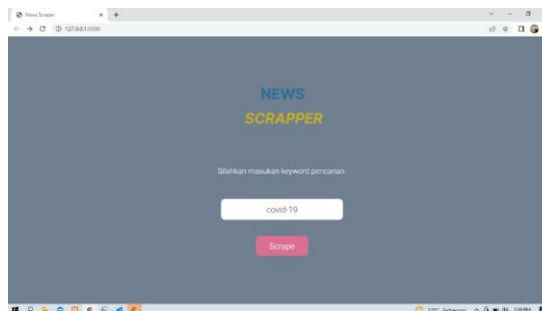
```
app.py x
1 from flask import Flask, flash, render_template, request
2 from bs4 import BeautifulSoup
3 from selenium import webdriver
4 import csv
5
6 app = Flask(__name__)
7 app.secret_key = "manbearpig_MUDDMAN888"
8
9 # route untuk ke page (request)
10 @app.route('/', methods=['POST', 'GET'])
11 def index():
12
13     # tambahkan bila tombol great diklik
14     if request.method == 'POST':
15         flash("Proses scraping selesai, masukan data untuk scraping kembali.") # tampilkan pesan flash
16
17     # mulai proses scraping
18     key = request.form['name']
19
20
21     url = [
22         "http://www.detik.com/search/searchall?query={}&title=2".format(key),
23         "http://search.detik.com/search/2?term={}&title=2".format(key),
24         "http://search.bekasini.com/?q={}".format(key),
25         "http://www.tribunnews.com/search?sr=1&sr-source=pub-1480190613e790651a000a11148-cf-0010101a1a1e",
26         "http://www.republika.com/search?sr=1&sr-source=pub-1480190613e790651a000a11148-cf-0010101a1a1e".format(key, key)
27     ]
28
29     for u in url:
30         driver = webdriver.Chrome("chromedriver.exe")
31         driver.get(u)
32         content = driver.page_source
33         driver.quit()
34         soup = BeautifulSoup(content, 'html.parser')
```

Sumber: Hasil Penelitian (2022)

Gambar 4. Tampilan kode *Web Scraping* Portal Berita

Memasukan Kata Kunci

Saat *web scraping* sudah dijalankan, kata kunci dapat digunakan untuk proses pencarian ke *website* berita. Kata kunci yang sudah dimasukkan maka akan dimulai pencarian sesuai kata kunci yang diinginkan, seperti kata kunci *Covid-19* pada gambar 5.

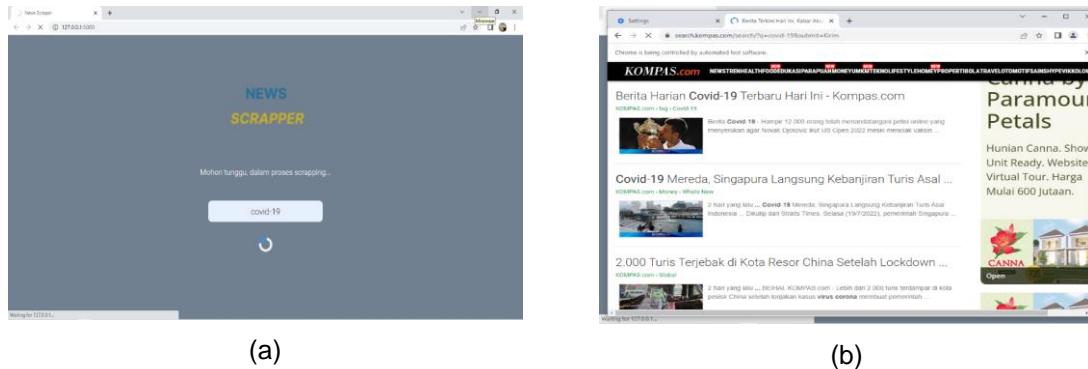


Sumber: Hasil Penelitian (2022)

Gambar 5. Tampilan Aplikasi *Web Scraping* dengan Kata Kunci *Covid-19*

Proses Scraping

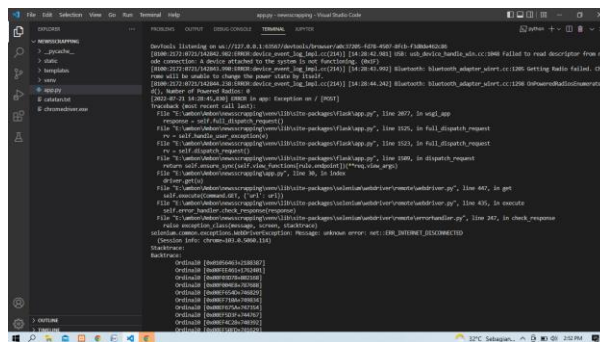
Pada gambar 6(a) dan 6(b) proses *scraping* portal *web* berita pada program berhasil dijalankan tanpa ada kegagalan ataupun *error* koneksi saat memulai *scraping*.



Sumber: Hasil Penelitian (2022)

Gambar 6. (a) Proses Memasukkan Kata Kunci, dan (b) Tampilan Proses *Scraping*

Scraping akan gagal ketika terputusnya koneksi atau tidak adanya koneksi pada perangkat yang digunakan seperti pada Gambar 7.

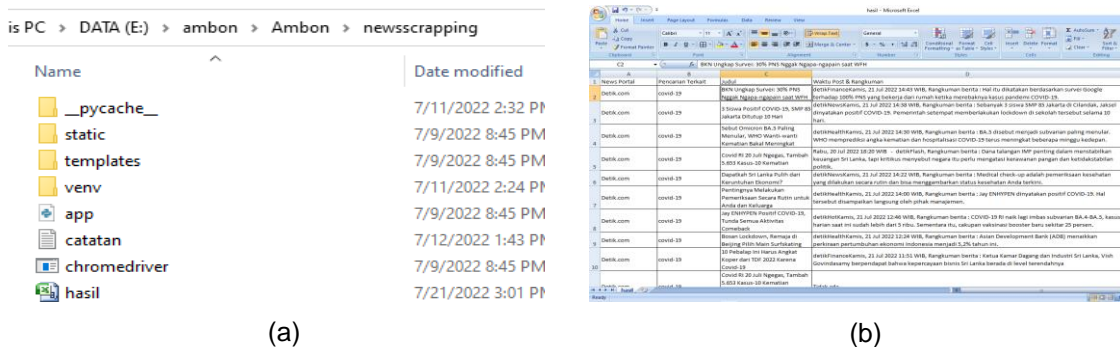


Sumber: Hasil Penelitian (2022)

Gambar 7. *Script Error*

Hasil Scraping

Pada Gambar 8(a) dan 8(b), untuk melihat hasil dari *scraping web* masuk kedalam folder penyimpanan yang sudah ditentukan. Hasil dari *scraping* berbentuk *file excel* sesuai dengan judul, kata kunci, dan waktu.



Sumber: Hasil Penelitian (2022)

Gambar 8. (a) Penyimpanan *File Hasil Scraping*, (b) *File Excell CSV Hasil Scraping*

Program ini dapat mengambil data seperti pada Gambar 8 (a) dan (b), hasil *scrapping* web portal berita berupa *file excel* seauai kata kunci yang diinginkan, judul (jika tertera), tanggal postingnya pada *website* berita dan rangkuman, lalu memasukkan data yang telah di ambil sebelumnya ke *file* berekstensi *csv* yang telah dibuat secara otomatis oleh program. Tabel 1 merupakan hasil *scrapping* kata kunci *Covid-19* berbentuk *file excel* data yang diambil dari *web*.

Tabel 4. 1 Hasil Web Scraping

News Portal	Pencarian Terkait	Judul	Waktu Post & Rangkuman	Relevan	Tidak Relevan
Detik.com	<i>Covid-19</i>	BKN Ungkap Survei: 30% PNS Nggak Ngapa-ngapainsaat WFH	detikFinance Kamis, 21 Jul 2022 14:43 WIB, Rangkuman berita: Hal itu dikatakan berdasarkan survei Google terhadap 100% PNS yang bekerja dirumah ketikamereba knyaka susepandemi COVID-19.	✓	✗
Detik.com	<i>Covid-19</i>	3 Siswa Positif COVID-19, SMP 85 Jakarta Ditutup 10 Hari	detikNews Kamis, 21 Jul 2022 14:38 WIB, Rangkuman berita: Sebanyak 3 siswa SMP 85 Jakarta di Cilandak, Jakseldinyatakan positif COVID-19. Pemerintah setempat memberlakukan 111 lockdown di sekolah tersebut selama 10 hari.	✓	✗
Detik.com	<i>Covid-19</i>	Sebut Omicron BA.5 Paling Menular, WHO Wanti-wanti Kematian Bakal Meningkat	detikHealth Kamis, 21 Jul 2022 14:30 WIB, Rangkuman berita: BA.5 disebut menjadi subvarian paling menular. WHO memprediksi angka kematian dan hospitalisasi COVID-19 terus meningkat beberapa minggu ke depan.	✓	✗
Detik.com	<i>Covid-19</i>	Covid RI 20 Juli Ngegas, Tambah 5.653 Kasus-10 Kematian	Rabu, 20 Jul 2022 18:20 WIB- detikFlash, Rangkuman berita: Dana talangan IMF penting dalam menstabilkan keuangan Sri Lanka, tapi kritikus menyebut negara itu perlu mengatasi kerawanan pangan dan ketidakstabilan politik.	✓	✗
Detik.com	<i>Covid-19</i>	Dapatkah Sri Lanka Pulih dari Keruntuhan Ekonomi?	detikNews Kamis, 21 Jul 2022 14:22 WIB, Rangkuman berita: Medical check-up adalah pemeriksaan kesehatan yang dilakukan secara rutin dan bisa menggambarkan status kesehatan Anda terkini.	✓	✗
Detik.com	<i>Covid-19</i>	Pentingnya Melakukan Pemeriksaan Secara Rutin untuk Anda dan Keluarga	detikHealth Kamis, 21 Jul 2022 14:00 WIB, Rangkuman berita: Jay ENHYPEN dinyatakan positif COVID-19. Hal tersebut disampaikan langsung oleh pihak manajemen.	✓	✗
Detik.com	<i>Covid-19</i>	Jay ENHYPEN Positif COVID-19, Tunda Semu	detikHot Kamis, 21 Jul 2022 12:46 WIB, Rangkuman berita: COVID-19 RI naik lagi imbassubvarian BA.4-BA.5,	✓	✗

News Portal	Pencarian Terkait	Judul	Waktu Post & Rangkuman	Relevan	Tidak Relevan
		aAktivitas Comeback	kasushariansaatinisudahlebihdari 5 ribu. Sementaraitu, cakupanvaksinasi booster barusekitar 25 persen.		
Detik.com	Covid-19	Bosan Lockdown, Remaja di Beijing Pilih Main Surfskating	detikHealthKamis, 21 Jul 2022 12:24 WIB, Rangkumanberita: Asian Development Bank (ADB) menaikkanperkiraanpertumbuhanekonomi Indonesia menjadi 5,2% tahunini.	✓	✗
Detik.com	Covid-19	10 PebalapIni Harus Angkat Koper dari TDF 2022 Karena Covid-19	detikFinanceKamis, 21 Jul 2022 11:51 WIB, Rangkumanberita: Ketua Kamar Dagang dan Industri Sri Lanka, VishGovindasamyberpendapatb ahwakepercayaanbisnis Sri Lanka berada di level terendahnya	✓	✗
Kompas.com	Covid-19	BeritaHarian Covid-19 Terbaru Hari Ini - Kompas.com	Berita Covid-19 - Hampir 12.000 orang telahmenandatanganipetisi online yang menyerukan agar Novak Djokovic ikut US Open 2022 meskimenolakvaksin ...	✓	✗
Kompas.com	Covid-19	BeritaHarian Covid-19 Terbaru Hari Ini - Kompas.com	2 hari yang lalu ... Covid-19 Mereda, Singapura LangsungKebanjiranTurisAsal Indonesia ... Dikutipdari Straits Times, Selasa (19/7/2022), pemerintah Singapura ...	✓	✗
Kompas.com	Covid-19	Covid-19 Mereda, Singapura LangsungKe banjiranTuris Asal ...	2 hari yang lalu ... BEIHAI, KOMPAS.com - Lebihdari 2.000 turisterdampar di kotapesisir China setelahlonjakankasus virus corona membuatpemerintah ...	✓	✗
Kompas.com	Covid-19	Covid-19 Mereda, Singapura LangsungKe banjiranTuris Asal ...	2 hari yang lalu ... JAKARTA, KOMPAS.com - Menteri KoordinatorBidangPolitik, Hukum, dan Keamanan (MenkoPolhukam) Mahfud MD terkonfirmasipositif Covid-19 ...	✓	✗
Kompas.com	Covid-19	2.000 TurisTerjebak di Kota Resor China Setelah Lockdown ...	23 Okt 2020 ... Kasus Covid-19 di Indonesia diumumkan pada 2 Maret 2020, kinilebihdari 377 ribu orang terinfeksi. Pentinguntukmengetahuiklaster pertama ...	✓	✗
Kompas.com	Covid-19	2.000 TurisTerjebak di Kota Resor China Setelah Lockdown ...	9 Jul 2021 ... Penderita Covid-19 perlumewaspadai happy hypoxia karenabisamenyebabkantubuhkuranganoksigentibatanpasesak napas. Halaman all.	✓	✗
Kompas.com	Covid-19	Mahfud MD Positif Covid-19 Sepulang Ibadah Haji	19 Mar 2022 ... Denganpenambahantersebut, total pasien Covid-19 yang meninggal dunia kiniberjumlah 153.599 orang.	✓	✗
Kompas.com	Covid-19	Mahfud MD Positif	28 Mei 2022 ... Pemerintahmelaporkan 279	✓	✗

News Portal	Pencarian Terkait	Judul	Waktu Post & Rangkuman	Relevan	Tidak Relevan
		<i>Covid-19</i> Sepulang Ibadah Haji	kasusbaru <i>Covid-19</i> dalam 24 jam terakhir. Penambahankasusitutersebar di 24 provinsi.		
Kompas.com	<i>Covid-19</i>	MenelusuriKlasterPertamaPenularan <i>Covid-19</i> di Indonesia ...	21 Mar 2022 ... Petugaskesehatanmengambil sampeltes Swab PCR COVID -19 untuk guru dan siswasaatpelacakanklustersekolah di SMA N 1 Bantul, ...	✓	✗
Kompas.com	<i>Covid-19</i>	MenelusuriKlasterPertamaPenularan <i>Covid-19</i> di Indonesia ...	6 Mar 2022 ... Surat yang ditandatangani oleh KetuaSatgasPenanganan <i>Covid-19</i> Letjen TNI Suharyanto pada Rabu (2/3/2022) itubukanmenyatakan <i>Covid-19</i> ...	✓	✗
KapanLagi.com	<i>Covid-19</i>	Termasuk Jimin BTS, 28 Idola K-Pop DiyatakanPositif COVID-19 di ...	31 Jan 2022 ... Berikutinideretan artis K-Pop yang dinyatakansudahpositif COVID-19.	✓	✗
KapanLagi.com	<i>Covid-19</i>	Termasuk Jimin BTS, 28 Idola K-Pop DiyatakanPositif COVID-19 di ...	21 Des 2020 ... Kapanlagi.com - Hinggasaatini virus corona masihmenjadimusuhbanyak negara termasuk Indonesia. Setelahhampir 10 bulan, pandemibelum ...	✓	✗
KapanLagi.com	<i>Covid-19</i>	Pevita Pearce & NirinaZubir Positif <i>Covid-19</i> , Bukti Virus Tak ...	23 Jun 2021 ... MijooNegatif COVID-19. Gambar. Mijoo - JisooLovelyz credit: lovelyz 94's. SetelahJisoodikonfirmasipositif, member Lovelyz lain dan staff ...	✓	✗
KapanLagi.com	<i>Covid-19</i>	Pevita Pearce & NirinaZubir Positif <i>Covid-19</i> , Bukti Virus Tak ...	25 Des 2021 ... 1. RM Sedang MelakukanKarantina. Beginilahpernyataandari Big Hit Music: "Member BTS, RM dan Jindidiagnosispositif COVID-19 pada Sabtumalam ...	✓	✗
KapanLagi.com	<i>Covid-19</i>	JisooLovelyz Positif COVID-19, MijooDipastikanKarantinaMandiri ...	18 Sep 2020 ... Bukan Karena <i>Covid-19</i> . Sebelumnyaপিহকkeluargadiwaili oleh sang adikbernama Dai Tirtamengungkapbila Ade Firman Hakim meninggaldiduga ...	✓	✗
KapanLagi.com	<i>Covid-19</i>	JisooLovelyz Positif COVID-19, MijooDipastikanKarantinaMandiri ...	30 Mar 2020 ... Apasajajenistanaman herbal yang dapatmencegah virus corona covid-19? Cari tahu di artikelinini.	✓	✗
KapanLagi.com	<i>Covid-19</i>	Setelah Suga BTS, RM dan Jin Juga	26 Jul 2021 ... Kapanlagi.com - <i>Covid-19</i> sampaisekarangmasihmenjadi musuhkitasemua. Para	✓	✗

News Portal	Pencarian Terkait	Judul	Waktu Post & Rangkuman	Relevan	Tidak Relevan
KapanLagi.com	Covid-19	DinyatakanP ositif COVID- 19 ... Setelah Suga BTS, RM dan Jin Juga DinyatakanP ositif COVID- 19 ...	selebritis Korea pun juga kinibanyak yang mengumumkankalau ... 13 Des 2021 ... Kareena Kapoor dinyatakanpositif covid- 19 setelahmenjalanites. Pelanggaranapa yang didugadilakukan oleh Kareena?	✓	✗
KapanLagi.com	Covid-19	Hasil Swab Test Mendiang Ade Firman Hakim Telah Keluar ...	5 Jul 2021 ... Kabardukaterusdatangdari dunia hiburanselebritis Tanah Air usaibeberapavarian virus Covid-19 muncul dan mengganas.	✓	✗
KapanLagi.com	Covid-19	Hasil Swab Test Mendiang Ade Firman Hakim Telah Keluar ...	18 Jul 2021 ... Bahkansederetseleb Korea dikonfirmasi juga positif Covid- 19 dalambeberapahariterakhir. Setidaknyaada 12 bintang Korea, mulaidari idol, ...	✓	✗
Tribunnews.com	Covid-19	Jokowi Sebut Stok Vaksin Covid-19 LebihdariCu kup, Segera ...	30 Mei 2022 ... Kondisipandemi Covid-19 di Indonesia kinisudahmelandai. Namun Jokowi mengingatnasyarakat agar segeramelengkapi program vaksinasi.	✓	✗
Tribunnews.com	Covid-19	Jokowi Sebut Stok Vaksin Covid-19 LebihdariCu kup, Segera ...	22 Jun 2022 ... ReaksiSatgas Covid-19 terkaitpernyataanPresiden Joko Widodo menyebutkanjikakesulitanmenc aripesertauntuksuntikvaksin booster.	✓	✗
Tribunnews.com	Covid-19	Presiden Jokowi SebutVaksin asi Booster Sepi Peminat, Begini ...	13 Jun 2022 ... TRIBUNNEWS.COM, JAKARTA - Terjadikenaikankasus Covid-19 di beberapa negara akibat sub varian Omicron yaitu BA.4 dan BA.5. Sub varianini juga ...	✓	✗
Tribunnews.com	Covid-19	Presiden Jokowi SebutVaksin asi Booster Sepi Peminat, Begini ...	17 Mei 2022 ... Presiden Joko Widodo (Jokowi) akhirnyamemutuskankelonggar andalamkebijakanpemakaian masker.	✓	✗
Tribunnews.com	Covid-19	Presiden Jokowi ImbauVaksin Booster DigenjotLagi, Waspada ...	10 Jun 2022 ... Kasus Covid-19 di Indonesia mengalamikenaikansebesar 31 persendalamtiga pekan terakhir, Jokowi sebutmasihamanterkendali.	✓	✗
Tribunnews.com	Covid-19	Presiden Jokowi ImbauVaksin Booster DigenjotLagi, Waspada ...	10 Jun 2022 ... Kasus Covid-19 di Indonesia mengalamikenaikandalambeber apahariterakhirini, Presiden Joko Widodo (Jokowi) imbaumasyrakattetap ...	✓	✗

News Portal	Pencarian Terkait	Judul	Waktu Post & Rangkuman	Relevan	Tidak Relevan
Tribunnews.com	Covid-19	Pandemi Covid-19 Semakin Terkendali, Jokowi Izinkan Masyarakat ...	21 Jun 2022 ... Dalam pertemuan tersebut, Tedros menyampaikan kepada Jokowi bahwa pandemi Covid-19 belum selesai. Menteri Luar Negeri Retno Marsudi awalnya ...	✓	✗
Tribunnews.com	Covid-19	Pandemi Covid-19 Semakin Terkendali, Jokowi Izinkan Masyarakat ...	11 Feb 2022 ... Johnson mengisyaratkan diakhirnya semua pembatasan Covid-19 di Inggris, termasuk persyaratan hukum untuk mengisolasi diri untuk kasus ...	✓	✗
Tribunnews.com	Covid-19	Kasus Covid-19 Naik, Jokowi: Masih Terkendali - Tribunnews.com ...	17 Jun 2022 ... Presiden Joko Widodo (Jokowi) mengatakan perkembangan Penyakit Mulut dan Kuku yang menyerang sapiternaksamasepeti Covid 19.	✓	✗
Tribunnews.com	Covid-19	Kasus Covid-19 Naik, Jokowi: Masih Terkendali - Tribunnews.com ...	6 Apr 2022 ... Ilustrasi virus corona - Sub-varian baru Omicron telah ditemukan di Inggris, disebut dengan Omicron XE. Apagejalanya? TRIBUNNEWS.	✓	✗

Sumber: Hasil Penelitian (2022)

Sebagaimana Tabel 1 hasil *web scraping* situs berita, menggunakan bahasa pemrograman *python* berhasil dijalankan dan mengambil data dari *website* berita <http://www.kompas.com>, <http://www.detik.com>, http://www.kapan_lagi.com, dan <http://www.tribunnews.com> sesuai kata kunci yang dicari *covid-19*.

4. Kesimpulan

Penelitian ini melakukan proses pengambilan informasi dari halaman web yang ditargetkan dengan metode *web scraping*. Pengumpulan data menjadi lebih cepat dari beberapa situs sekaligus menggunakan aplikasi yang telah dibuat dilakukan secara bersamaan secara otomatis, proses pengambilan informasi khususnya yang bersumber dari artikel di internet menggunakan kata kunci agar memperoleh data sesuai yang dibutuhkan. Hasil pengujian *web scraping* yang telah dibuat berjalan sesuai yang diharapkan yaitu data berhasil dikumpulkan secara otomatis, dan disimpan dalam bentuk *file excell csv* yang dibuat secara otomatis oleh program, hasil datanya diperoleh sesuai kata kunci yang diinginkan sesuai judul (jika tertera), tanggal postingnya pada *website* berita serta rangkuman. Untuk implementasi, jumlah situs berita yang dituju dapat ditambahkan sesuai dengan kebutuhan.

Daftar Pustaka

- Arisandi, D., Indra, Z., & Kartini, K. (2021). Mengidentifikasi Hoax Pada Hasil Pencarian Berita Online Dengan Teknik *Web scraping* dan Algoritma C4.5. *Rabit: Jurnal Teknologi dan Sistem Informasi Univrab*, 6(2), 130–137. <https://doi.org/10.36341/rabit.v6i2.690>
- Ayani, D. D., Pratiwi, H. S., & Muhandi, H. (2019). Implementasi *Web Scraping* untuk Pengambilan Data pada Situs Marketplace. *Jurnal Sistem dan Teknologi Informasi (JUSTIN)*, 7(4), 257. <https://doi.org/10.26418/justin.v7i4.30930>
- Dewantoro, P. R. (2016). *Implementasi Teknik Web Scraping pada Proses Topic Modelling Portal Berita* [Universitas Gadjah Mada]. http://etd.repository.ugm.ac.id/index.php?mod=penelitian_detail&sub=PenelitianDetail&act=view&typ=HTML&buku_id=100170&obyek_id=4
- 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>
- Indra, Z., & Trisnawati, L. (2018). Pengembangan Intelligent Data Collector untuk Analisis Big Data Artikel Berita Online. *Rabit: Jurnal Teknologi Dan Sistem Informasi Univrab*, 3(1), 49–59. <https://doi.org/10.36341/rabit.v3i1.388>
- Mitra, V., Sujaini, H., & Negara, A. B. P. (2017). Rancang Bangun Aplikasi *Web Scraping* untuk Korpus Paralel Indonesia - Inggris dengan Metode *HTML DOM*. *Jurnal Sistem Dan Teknologi Informasi (JUSTIN)*, 5(1), 1–6. <https://jurnal.untan.ac.id/index.php/justin/article/view/18130>
- Priyanto, A., & Ma'arif, M. R. (2018). Implementasi *Web Scraping* dan Text Mining untuk Akuisisi dan Kategorisasi Informasi Laman Web Tentang Hidroponik. *Indonesian Journal of Information Systems (IJIS)*, 1(1), 25–33. <https://doi.org/10.24002/ijis.v1i1.1664>
- Rahmatulloh, A., & Gunawan, R. (2020). *Web Scraping* with *HTML DOM* Method for Data Collection of Scientific Articles from Google Scholar. *Indonesian Journal of Information Systems*, 2(2), 95–104. <https://doi.org/10.24002/ijis.v2i2.3029>
- Ramadhan, I., & Sastramihardja, H. (2018). Pemanfaatan *Web Crawler* Dalam Mengumpulkan Informasi Melalui Internet. *Konferensi Nasional Sistem Informasi 2018*, 8–9. http://digilib.mercubuana.ac.id/manager/t!@file_artikel_abstrak/Isi_Artikel_160883504689.pdf
- Rohman, M. S., Santoso, H. A., Saraswati, G. W., Anisa, N., & Winarsih, S. (2019). Pemanfaatan Topic-Focused Crawler untuk Pembangunan Corpus Berita Bencana menggunakan Teknik Scrapy CSS Selector. *Seminar Nasional APTIKOM*, 2019. <http://bpbddjateng.com>
- 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>

Syaefulloh, A. Y. (2020). *Web Scraping Menggunakan Python*. March, 1–2.
https://www.researchgate.net/publication/340265742_Web_Scraping_Menggunakan_Python

Yunita, N. (2020). *Implementasi Teknologi Text Mining dan Web scraping untuk Mengumpulkan Informasi pada Website Pelatihan Keterampilan Khusus Perempuan*. Politeknik Negeri Jakarta.