



## PENERAPAN TEKNOLOGI WEB SCRAPING SEBAGAI PENGUMPULAN DATA COVID-19 DI PROVINSI LAMPUNG

Nurman Fadhlullah<sup>1</sup>, Setiawansyah<sup>2</sup>, Ade Surahman<sup>3</sup>

<sup>1,2,3</sup>Fakultas Teknik dan Ilmu Komputer Universitas Teknokrat Indonesia

<sup>1</sup>nurman\_fadhlullah@teknokrat.ac.id, <sup>2</sup>setiawansyah@teknokrat.ac.id, <sup>3</sup>adesurahman@teknokrat.ac.id

Received: (1 januari 2022) Accepted: (3 Februari 2022) Published: (30 Maret 2022 )

### Abstract

On January 30, 2020, the World Health Organization declared the novel coronavirus (COVID-19) outbreak a world public health emergency. Covid-19 itself is a group of viruses found in animals and humans. Responding to COVID-19 requires critical preparation and response, such as equipping health workers with complete equipment and managing health facilities to provide information, safe and effective procedures. The amount of data displayed from each website makes data seekers confused which website they should choose to find the data they want. From this problem, the authors use the method web scraping as an effort to find data Covid-19 and compare which websites are the most active in making news about Covid-19 in Lampung Province. Data will be taken from 3 (three) news websites in Lampung province, namely Lampung Post, Radar Lampung, and Tribun Lampung. The data that will be taken from the three websites are in the form of titles and links. Web scraping will be very appropriate in this study, because of structured data retrieval from these techniques. In this study, the author used python library to facilitate the application of web scraping. The results of this study, techniques web The scraping that the author applies can retrieve data from three websites, so that the data can be displayed as an effort to compare which websites are more active in publishing news Covid-19 in Lampung province.

**Keywords:** Social Media, covid-19, Web Scraping, Python

### Abstrak

Pada 30 Januari 2020, Organisasi Kesehatan Dunia menyatakan wabah virus corona baru (COVID-19) sebagai darurat kesehatan masyarakat dunia. Covid-19 sendiri merupakan kelompok virus yang terdapat pada hewan dan manusia. Untuk merespon COVID-19 diperlukan persiapan dan respon kritis, seperti membekali petugas kesehatan dengan peralatan yang lengkap dan mengelola fasilitas kesehatan untuk memberikan informasi, prosedur yang aman dan efektif. Banyaknya data yang ditampilkan dari masing-masing website membuat pencari data bingung website mana yang harusnya dia pilih untuk mencari data yang dia inginkan. Dari permasalahan tersebut, maka penulis menggunakan metode web scraping sebagai upaya mencari data covid-19 dan membandingkan website mana yang paling aktif dalam membuat berita tentang covid-19 di provinsi lampung. Data akan diambil dari 3 (tiga) website berita yang ada di provinsi lampung yaitu lampung post, radar lampung, dan tribun lampung. Data yang akan diambil dari tiga website tersebut yaitu berupa judul dan link. Web scraping akan sangat tepat dalam penelitian ini, karena pengambilan data yang terstruktur dari teknik tersebut. Dalam penelitian ini digunakan library python untuk memudahkan dalam penerapan web scraping. Hasil penelitian ini, teknik web scraping yang penulis terapkan dapat mengambil data dari tiga website, sehingga data tersebut dapat ditampilkan sebagai upaya membandingkan website mana yang lebih aktif memuat berita covid-19 di provinsi lampung..

**Kata Kunci:** Media Sosial, covid-19, Web Scraping, Python

### To cite this article:

Fadhlullah et.al. (2022). Penerapan teknologi web scraping sebagai pengumpulan data covid-19 di provinsi lampung. *Jurnal Informatika dan Rekayasa Perangkat Lunak*, Vol.3 No.1, 25-30.

## PENDAHULUAN

Pada 30 Januari 2020, Organisasi Kesehatan Dunia menyatakan wabah virus corona baru (COVID-19) sebagai darurat kesehatan masyarakat dunia. Covid-19 sendiri merupakan kelompok virus yang terdapat pada hewan dan manusia. Untuk merespon COVID-19 diperlukan persiapan dan respon kritis, seperti membekali petugas kesehatan dengan peralatan yang lengkap dan mengelola fasilitas kesehatan untuk memberikan informasi, prosedur yang aman dan efektif.

Di satu sisi, banyaknya sumber informasi akan membawa keuntungan bagi pencari data, karena informasi yang diberikan dapat ditambah antara satu halaman dengan halaman lainnya. Namun di sisi lain, jumlah informasi yang sangat banyak dan beragam, dapat menyulitkan pencari data untuk menemukan data yang akan diambilnya.

Berdasarkan permasalahan yang telah dijelaskan pada paragraf sebelumnya, dapat disimpulkan bahwa diperlukan portal terintegrasi yang secara otomatis dapat mengambil informasi dari halaman sebanyak mungkin dan kemudian memberikannya kepada pencari data lainnya.

Penelitian ini bertujuan untuk mengeksplorasi halaman-halaman yang memberikan informasi mengenai data atau berita COVID-19, kemudian menggunakan teknologi web scraping untuk mengumpulkan semua informasi pada halaman tersebut secara otomatis. Dengan cara ini, dapat menjadi lebih ringkas, tetapi dapat memberikan informasi yang lengkap dan akurat kepada pencari data. Oleh karena itu, peneliti menamakan judul "Penerapan teknologi web scraping sebagai pengumpulan data covid-19 di provinsi Lampung" sebagai media pengumpulan data untuk membantu para pencari data COVID-19 agar dapat digunakan untuk pemanfaatan data yang lebih lanjut. Tujuan dalam penelitian ini adalah menghasilkan informasi atau data yang akurat dengan menerapkan metode Web Scraping sebagai media pencari data COVID-19 di provinsi Lampung.

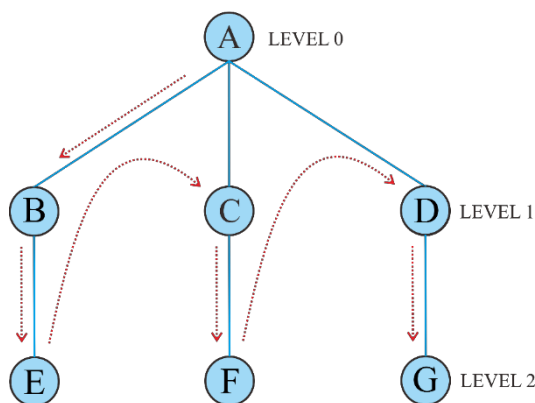
## METODE

### *Web Scraping*

Web Scraping adalah sebuah program atau code otomatis yang dapat mengunjungi dan mencari informasi yang sesuai dengan kata kunci dalam sebuah website. Web scraping ini berkelana di web dan berfungsi untuk mengumpulkan segala informasi tentang suatu halaman web dan mengindeksnya ke dalam suatu database. Informasi tentang halaman web tersebut didapat dari kata-kata yang terdapat didalam halaman web tersebut. Kata-kata tersebut kemudian di indeks menjadi sebuah daftar kata-kata yang biasanya dipakai sebagai kata kunci untuk menemukan halaman web. Proses pengumpulan informasi tentang halaman web dari situs web ini disebut Web scraping.

### *Depth First Search (DFS)*

Pencarian dengan metode ini dilakukan dari node awal secara mendalam hingga yang paling akhir (dead-end) atau sampai ditemukan. Dengan kata lain, simpul cabang atau anak yang terlebih dahulu dikunjungi.



**Gambar 1.** Alur Pencarian Algoritma Depth First Search

Pada gambar diatas, tiap lingkaran berabjad disebut node. Tiap node bisa memiliki anak berupa node yang lainnya. Pada pencarian dengan algoritma Depth First Search, pencarian dimulai dari level paling pertama (level 0) kemudian dilanjutkan ke anak paling kiri pada level berikutnya (level 1) demikian seterusnya sampai tidak terdapat

anak lagi atau level yang lebih dalam lagi. Jika pencarian sudah mencapai node atau anak paling dalam maka akan dilakukan penelusuran mundur atau backtracking untuk melakukan pencarian ke node anak berikutnya. Pada gambar diatas, pencarian dilakukan mulai dari node A-B-E, pada node ini akan melakukan penelusuran mundur menuju C-F kemudian melanjutkan ke node anak berikutnya yaitu D-G. Demikian seterusnya sampai tujuan ditemukan.

### Python

Python adalah bahasa pemrograman tingkat tinggi (higt-level programing language), berjalan dengan sistem interpreted, dan bisa dipakai untuk berbagai jenis tujuan (general-purpose). Sebutan bahasa pemrograman tingkat tinggi merujuk level kedekatan sebuah bahasa pemrograman ke kode-kode listrik yang dipakai komputer. Sebuah bahasa programan disebut sebagai bahasa pemrograman tingkat tinggi (higt- level programinfg language) karena perintah atau kode program yang dipakai sudah mirip dengan bahasa manusia (bahasa inggris). Hampir semua bahasa pemrograman moderen masuk ke dalam bahasa tingkat tinggi, diantara Pascal, C++, Java, PHP, Java Script dan juga python. Bahasa C masih sering disebut bahasa tingkat menengah karena sebagian perintah bisa langsung mengakses hardware komputer, namun karena hal ini juga bahasa C sedikit lebih rumit untuk dipelajari Python menggunakan metode pemrosesan interpreted, dimana kode program akan diproses baris per baris langsung dari kode program (tidak butuh proses compile). Ini mirip seperti bahasa script seperti PHP dan Java Script(Anwar, 2019).

### CSS

CSS (Cascading Style Sheet) adalah sebuah mekanisme sederhana untuk memberikan style (seperti font, warna, jarak, spasi, dll) kepada dokumen web yang ditulis dalam HTML atau XML (termasuk beberapa variasi bahasa XML seperti XHTML dan SVG) [17].

### Javascrip

Javascript adalah suatu bahasa pemrograman web yang bersifat Client Side Programming Language. Client Side Programming Language merupakan tipe pemrograman yang prosesnya dilakukan oleh client atau user. Proses yang dilakukan oleh user yang dimaksud adalah proses yang dilakukan melalui aplikasi browser seperti Google Crome, Mozilla Firefox ataupun Internet Explorer [17].

### eXtreme Programming (XP)

eXtreme Programming (XP) adalah metode pengembangan perangkat lunak yang sederhana dan mencakup salah satu metode tangkas yang dipelopori oleh Kent Beck, Ron Jeffries, dan Ward Cunningham. XP adalah salah satu metode tangkas yang paling banyak digunakan dan menjadi pendekatan yang sangat terkenal. Tujuan XP adalah tim yang terbentuk antara kursus berukuran kecil hingga menengah, tidak perlu menggunakan tim besar. Hal ini dimaksudkan untuk mengatasi persyaratan yang tidak jelas dan perubahan persyaratan dengan sangat cepat [19].

## HASIL DAN PEMBAHASAN

### Implementasi Web Scraping

Pada tahapan ini dilakukan proses implementasi atau penerapan dari metode web robot, berdasarkan dengan apa yang telah direncanakan. Sebelum mengimplementasikan web robot, terlebih dahulu harus diketahui struktur HTML dari halaman web untuk menentukan dalam tag element yang mana informasi yang ingin didapatkan direpresentasikan.

Berikut ini adalah daftar tag element yang digunakan untuk mengambil data yang kemudian akan ditampilkan pada program yang akan dibuat.

NO	URL	Tag HTML
1	<a href="https://radarlampung.co.id/?s=covid">https://radarlampung.co.id/?s=covid</a>	h3[class=box-content]
2	<a href="https://m.lampost.co/search?q=covid">https://m.lampost.co/search?q=covid</a>	h2[class=media-heading]
3	<a href="https://lampung.tribunnews.com/search?q=covid">https://lampung.tribunnews.com/search?q=covid</a>	h3[class=gs-title]

Tabel 1. Tag element halaman website

### Pseudocode Proses Web Scraping

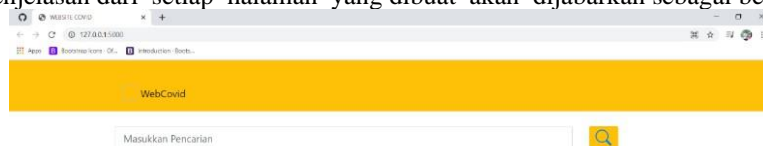
Pada bagian ini akan dijelaskan bagaimana proses dari web robot dalam bentuk Pseudocode. Tujuan dari penggunaan pseudocode adalah untuk mempermudah pembaca dalam pemahaman, dibandingkan menggunakan bahasa pemrograman yang umum digunakan. Untuk proses pseudocode dari web scraping dapat dilihat pada gambar:

```
Function scraping (url)
  Check_is_valid_url(url)
  Urlpage = this -> url
  Data = parse(urlpage)
  Foreach parseData in data
    : getSpecialWrapperClass()
    constructParameters()
    worth =
    worthtoSave() if
    worth = true
    saveData()
  end
```

**Gambar 2.** Pseudocode web scraping

### Implementasi Antar Muka

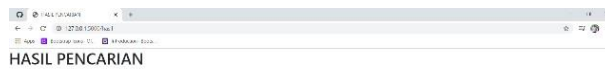
Dalam pembuatan program, program disusun menjadi dua halaman yaitu, halaman utama dan halaman hasil pencarian. Tampilan dan penjelasan dari setiap halaman yang dibuat akan dijabarkan sebagai berikut :



**Gambar 3.** Halama utama

Halaman utama dapat di lihat pada gambar 3 dimana terdapat kolom pencarian bagi user untuk memasukkan kata kunci pencarian ,sebagai contoh penulis memasukan kata kunci “jumlah positif covid di provinsi lampung”

pada kolom pencarian, setelah itu tekan tombol enter atau klik tombol cari maka hasil pencarian akan ditampilkan pada halaman hasil pencarian seperti pada gambar 4.



Gambar 4. Halaman hasil

Setelah mengklik tombol cari, pengguna akan masuk ke halaman hasil pencarian, yang menampilkan hasil pencarian. Jika pengguna ingin membaca detail isi dari berita, hanya dengan mengklik link yang berada dibawah judul maka pengguna akan langsung diarahkan ke website resmi penyedia berita tersebut, seperti pada gambar 5.



Gambar 5. Penyedia berita

## PENUTUP

Berdasarkan hasil penerapan dan pengujian yang penulis lakukan tentang penerapan web scraping sebagai perbandingan website mana yang lebih sering memuat berita tentang covid-19 di provinsi lampung dengan menggunakan media website berita, dapat diambil kesimpulan yaitu: Teknik web scraping yang penulis terapkan pada lima website berita yaitu lampung post, radar lampung, tribun lampung, berhasil mengambil data yang ditampilkan dari hasil pencarian website tersebut. Dengan menggunakan library python dapat memudahkan proses web scraping dengan hanya memanggil fungsi yang ada di library tersebut. Program yang dibuat dapat mencari berita berdasarkan judul berita yang dicari, jika kata kunci yang dimasukan tidak terkait. Tidak semua website berita mempunyai url pencarian, sehingga tidak dapat menerapkan teknik web scraping pada website yang tidak memiliki url.

## REFERENSI

- [1]. Alkhasawneh, S., & Alqahtani, M. A. M. (2019). Fostering students' self-regulated learning through using a learningmanagement system to enhance academic outcomes at the University of Bisha. *TEM Journal*, 8(2), 662–669. <https://doi.org/10.18421/> Aditya, B. R., 2015. Penggunaan Web Crawler Untuk Menghimpun Tweets dengan Metode Pre-Processing Text Mining.
- [2]. Anwar, A., 2019. Pengertian dan Sejarah Bahasa Pemrograman Python.
- [3]. Babar, M. & Iqbal, H., 2016. An Approach for Analyzing ISO / IEC 25010 Product Quality Requirements based on Fuzzy Logic and Likert Scale for Decision Support Systems.
- [4]. Batoebara, M. U., Suyani, E. & Nurafiah, C. A., 2020. Literasi Media dalam Menanggulangi Berita Hoaks (Studi pada Siswa SMKN 5 Medan).

- [5]. Effendy, I., 2019. Penerapan Metode Cross Selling pada Aplikasi Web Crawler E- Commerce.
- [6]. Gunawan, H. & Triantoro, A., 2017. Sistem Informasi Pengolahan Rapor Kurikulum 2013.
- [7]. Homaidi, A. & Ibad, S., 2019. Analisis Pemodelan Sistem Pengaduan Kasus Menggunakan Object Oriented Method (Unified Modelling Language).
- [8]. Lubis, B. O., 2016. PENERAPAN GLOBAL EXTREME PROGRAMMING PADA SISTEM INFORMASI WORKSHOP SEMINAR DAN PELATIHAN DI LEMBAGA EDUKASI. INFORMATIKA, Volume 3.
- [9]. 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).TEM82-47
- [10]. *International Journal of Educational Technology in Higher Education*, 14(1). <https://doi.org/10.1186/s41239-017-0076-8>
- [11].Pasaribu, A., Sihombing, M. & Buaton, R., 2020. Perancangan Sistem Pendeteksi Berita Hoax Menggunakan Algoritma Levenshtein Distance Berbasis PHP.
- [12].Putu Kussa Laksana Utama, 2018. Identifikasi Hoax pada Media Sosial dengan Pendekatan Machine Learning.
- [13].Rinaldi, R., 2019. Penerapan Unified Modelling Language (UML) dalam Analisis dan Perancangan Aplikasi E-Learning.
- [14].Shita, R. T. & S., 2016. Implementasi Algoritma BFS(Breadth-First Search) pada Aplikasi Web Crawler.
- [15].Singh , A. V., V. & Mishra, A., 2014. A Review of Web Crawler Algorithms.
- [16].Supriyatna & Adi, 2018. METODE EXTREME PROGRAMMING PADA PEMBANGUNAN WEB. JURNAL TEKNIK INFORMATIKA,Volume 11.
- [17].Surahman, A., Octaviansyah, A. F. & Darwis, D.,
- [18].2020. EKSTRAKSI DATA PRODUK E- MARKETPLACE SEBAGAI STRATEGI PENGOLAHAN SEGMENTASI PASAR MENGGUNAKAN WEB CRAWLER.
- [19].Syarifudin, Y. W., Syafiandini, A. F. & Prisdana, H. R., 2018. Aplikasi Pencarian Penjualan Laptop Menggunakan Tenologi Web Scraping.
- [20].Wati, R., 2020. Penerapan Algoritma Naive Bayes dan Particle Swarm Optimization untuk Klasifikasi Berita Hoax pada Media Sosial.
- [21].Zuliarso, E. & Mustofa, K., 2009. Crawling Web berdasarkan Ontology