

PENERAPAN TEKNIK *WEB SCRAPING* UNTUK MEMBANGUN *WEBSITE* PEMANTAUAN BERITA TERPADU KLHK

Fadhilah Nur Hidayat Mulyono^{1*}, Kartika Maulida Hindrayani²

*Email : 21083010082@student.upnjatim.ac.id

Universitas Pembangunan Nasional “Veteran” Jawa Timur¹²

Abstrak Banyaknya berita yang bermunculan setiap hari menjadi tantangan tersendiri bagi Kementerian Lingkungan Hidup dan Kehutanan khususnya bagi biro Humas yang memantau dan menganalisis setiap informasi yang ada. Penelitian ini bertujuan meningkatkan efisiensi dan efektivitas pemantauan berita untuk Kementerian Lingkungan Hidup dan Kehutanan (KLHK) melalui teknik *web scraping*. Data dikumpulkan dari portal berita daring seperti CNBC Indonesia, Mongabay Indonesia, Detik, dan Tempo menggunakan *Python* dan *BeautifulSoup*. Data yang diperoleh disimpan dalam *database MySQL* dan diakses melalui antarmuka pengguna berbasis HTML, CSS, dan *JavaScript*. Hasilnya menunjukkan bahwa *web scraping* signifikan meningkatkan kemampuan biro Humas KLHK dalam memantau dan menganalisis berita secara *real-time*, menggantikan proses manual yang tidak efisien. Sistem ini mendukung pengambilan keputusan yang lebih tepat dan menunjukkan komitmen KLHK dalam memanfaatkan teknologi digital untuk tugas dan fungsinya. Penelitian ini membuktikan bahwa *web scraping* adalah solusi efektif untuk pemantauan berita dan dapat diterapkan dalam konteks serupa lainnya..

Kata kunci: KLHK, *Web scraping*, *Python*, *Real-time Monitoring*.

Abstract The amount of news that appears every day is a challenge for the Ministry of Environment and Forestry, especially for the Public Relations bureau that monitors and analyzes every existing information. This research aims to improve the efficiency and effectiveness of news monitoring for the Ministry of Environment and Forestry (KLHK) through web scraping techniques. Data is collected from online news portals such as CNBC Indonesia, Mongabay Indonesia, Detik, and Tempo using Python and BeautifulSoup. The data obtained is stored in a MySQL database and accessed through an HTML, CSS, and JavaScript-based user interface. The results show that web scraping significantly improves the ability of MoEF's PR bureau to monitor and analyze news in real-time, replacing inefficient manual processes. The system supports more informed decision-making and demonstrates KLHK's commitment to utilizing digital technology for its tasks and functions. This research proves that web scraping is an effective solution for news monitoring and can be applied in other similar contexts.

Keywords: KLHK, *Web scraping*, *Python*, *Real-time Monitoring*.

Pendahuluan

Kementerian Lingkungan Hidup dan Kehutanan Republik Indonesia disingkat Kementerian LHK atau KLHK adalah salah satu Kementerian penting dalam pemerintahan Indonesia yang membidangi urusan lingkungan hidup dan kehutanan. Kementerian LHK berada di bawah dan bertanggung jawab langsung kepada presiden dalam menyelenggarakan pemerintahan negara. Kementerian LHK berkomitmen mewujudkan Pembangunan lingkungan hidup dan kehutanan secara berkelanjutan dalam meningkatkan kesejahteraan rakyat menuju Indonesia Maju. Oleh karena itu, Kementerian LHK menyadari bahwasanya informasi yang semakin pesat dan penting seiring dengan semakin canggihnya era digital. Situs berita daring adalah salah satu sumber informasi yang dapat memberikan manfaat tersendiri bagi Masyarakat terutama dari sisi akses yang mudah. Berbagai macam berita tentang subjek termasuk politik, ekonomi, lingkungan, teknologi dan lain sebagainya tersedia di situs berita daring.

Dikutip dari wikipedia, surat kabar daring atau situs berita daring (*Online News*) adalah berita yang tersaji secara daring di internet dengan berbagai format. Berita daring adalah produk jurnalistik *online*. Berita daring juga disebut *enews* atau *e-news*. Menurut *Cambridge Dictionary*, *e-news* adalah berita yang dipublikasikan melalui internet (*news that is published over the internet*).

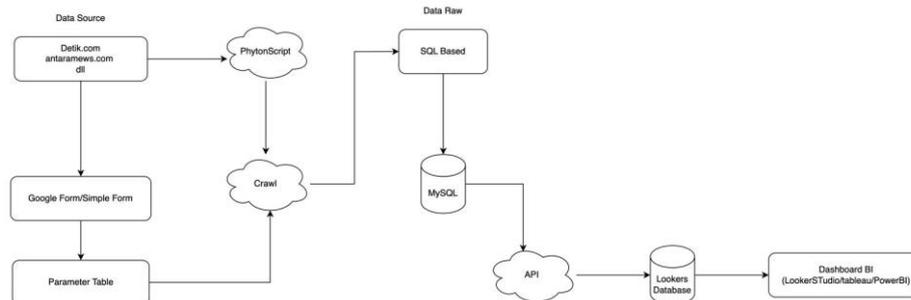
Banyaknya berita yang bermunculan setiap hari menjadi tantangan tersendiri bagi Kementerian Lingkungan Hidup dan Kehutanan khususnya bagi biro Humas yang memantau dan menganalisis setiap informasi yang ada. Metode yang selama ini digunakan sangatlah tradisional seperti mengumpulkan dan memilah artike berita daring (*Online News*) secara manual dengan mencari pada setiap laman berita daring yang ada dan menyimpan di *spreadsheet Excel* terbukti memakan waktu dan tidak efisien. Artikel berita merupakan sumber informasi yang berharga untuk pemantauan dan analisis, namun proses manual ini menghambat pemanfaatannya secara optimal. Maka dari itu kita memerlukan teknik yang efektif untuk mengumpulkan artikel berita secara otomatis.

Untuk mengatasi tantangan tersebut, biro Humas Kementerian Lingkungan Hidup dan kehutanan berinisiatif menggunakan teknik *web scraping* dalam membangun *website* terpadu untuk pemantauan berita. *Web scraping* merupakan suatu teknik ekstraksi atau pengambilan data, baik berupa teks, dokumen, foto, dan lain sebagainya dari sebuah laman di internet. *Web scraping* memanfaatkan pola HTML yang terdapat pada sebuah lama. Berbagai macam tools yang dapat digunakan untuk melakukan *scraping*. Salah satunya adalah Bahasa pemograman *Python*. Bahasa pemograman *Python* menyediakan *library* yang bisa di buat untuk *scraping*, seperti *BeautifulSoup*.

Penerapan teknik *web scraping* untuk *website* terpadu pemantauan berita Kementerian Lingkungan Hidup dan Kehutanan merupakan langkah maju yang signifikan dalam meningkatkan kemampuan Kementerian untuk memantau, menganalisis, dan memanfaatkan informasi berita secara efektif. Dengan memanfaatkan teknologi *web scraping*, biro Humas Kementerian Lingkungan Hidup dan Kehutanan dapat menyederhanakan proses pemantauan berita, memperoleh wawasan *real-time*, dan mengambil Keputusan tepat yang mendukung misinya.

Metode Pelaksanaan

Dalam mengerjakan proyek pembuatan *website* pemantauan berita terpadu KLHK ini, terdapat beberapa tahapan yang perlu dilakukan dalam menyelesaikan proyek tersebut.



Sumber: Hasil Rancangan (2024)
Gambar 1. Metodologi Pengerjaan

1. Pengumpulan Data

Proses pembuatan *website* pemantauan berita terpadu KLHK dimulai dengan pengumpulan data dari berbagai sumber berita *online*. Data ini diperoleh melalui penerapan teknik *web scraping*, dimana penulis menggunakan bahasa pemrograman *Python* dengan Pustaka *BeautifulSoup*. Pada tahap ini, analisis struktur HTML dari setiap *website* berita *online* dilakukan untuk mengekstrak informasi penting seperti headline, tanggal publikasi, URL, dan isi berita. Proses ini memastikan bahwa data yang telah diambil relevan dan dapat diolah lebih lanjut.

2. Penyimpanan Data

Setelah data berhasil diambil, langkah selanjutnya adalah menyimpan dalam sebuah *database* yang handal seperti, *MySQL*, *PostgreSQL*, atau *MongoDB*. Konektivitas antara proses *scraping* dan *database* dicapai melalui penggunaan API, yang memungkinkan data yang telah berhasil *discrease* disimpan langsung ke dalam *database*. Pada tahap ini, desain skema *database* dibuat untuk memastikan bahwa data disimpan dengan struktur yang terorganisir dan mudah diakses. Penggunaan Pustaka seperti *pymysql* membantu dalam menghubungkan aplikasi *Python* dengan *database*.

3. Pengembangan *Simple Form*

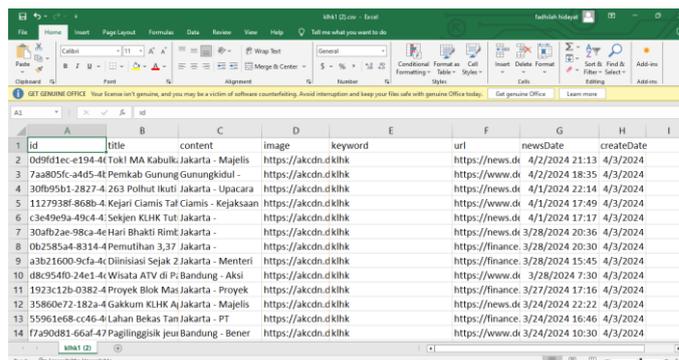
Berikut adalah pengembangan antarmuka pengguna dengan membuat sebuah *form* sederhana menggunakan HTML, CSS, dan *JavaScript*. *Form* ini memungkinkan pengguna untuk memasukkan parameter pencarian seperti tanggal, kata kunci, dan memastikan sumber berita. *Form* tersebut kemudian dikaitkan dengan *Backend* yang dibangun menggunakan *framework* seperti *Flask*. *Backend* bertanggung jawab untuk menerima input dari form, memvalidasi data yang diterima, dan memastikan bahwa input sesuai dengan format yang diharapkan sebelum diproses lebih lanjut.

4. Pengambilan Data Berdasarkan Parameter

Ketika parameter pencarian diterima di *Backend*, sistem akan membuat *query* berdasarkan input tanggal, kata kunci, dan sumber berita yang diberikan oleh pengguna. *Query* ini dijalankan terhadap *database* untuk mengambil data yang relevan. Hasil *query* kemudian diformat ke dalam bentuk JSON untuk dikirim kembali ke *frontend*. Tahap ini memastikan bahwa hanya data yang sesuai dengan kriteria pencarian yang di tampilkan kepada pengguna, meningkatkan relevansi dan kegunaan informasi yang disajikan.

Hasil dan Pembahasan

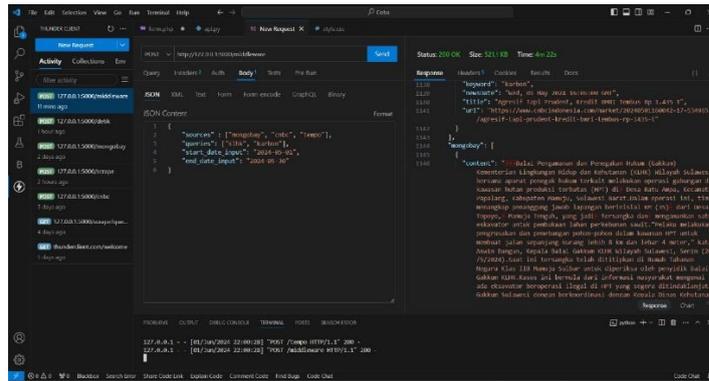
Proses pembuatan *website* pemantauan berita terpadu KLHK diawali dengan pengumpulan data dari portal berita *online* dengan menggunakan teknik *web scraping*. *Web scraping* adalah proses pengambilan sebuah dokumen semi-terstruktur dari internet, umumnya halaman-halaman web dalam bahasa *markup* seperti HTML atau XHTML, dan menganalisis dokumen tersebut untuk diambil data tertentu dari halaman tersebut untuk digunakan bagi kepentingan lain. *Web scraping* pada proyek ini dimanfaatkan untuk mengambil beberapa informasi yang diperlukan dari beberapa sumber media massa melalui beberapa portal berita. Portal berita yang digunakan untuk diambil informasinya yaitu CNBC Indonesia, Mongabay Indonesia, Detik, dan Tempo. Informasi yang diperlukan dari masing-masing portal berita yaitu judul berita, tanggal berita diterbitkan, link berita, isi berita, serta link gambar berita. *Scraping* pada 4 (empat) portal berita yang berbeda dengan struktur HTML yang berbeda pada masing-masing portal berita mengakibatkan *script* untuk melakukan *scraping* berbeda-beda juga. Penyesuaian untuk melakukan Parsing HTML serta penyesuaian terhadap format yang ada dan menyelaraskan hasil *output* data yang sama sehingga setiap portal berita memiliki bentuk proses yang berbeda dalam *script*nya.



A	B	C	D	E	F	G	H	
1	id	title	content	image	keyword	url	newsDate	createDate
2	0d9fd1ec-e194-4f11-8a4c-7a805fc-a4d5-4f	Tok1 MA Kabuki Jakarta - Majelis		https://akcdn.d.klkhk		https://news.dk	4/2/2024 21:13	4/3/2024
3	7aa805fc-a4d5-4f11-8a4c-30f895b1-2827-4	Pemkab Gunung Kidul -		https://akcdn.d.klkhk		https://www.dk	4/2/2024 18:35	4/3/2024
4	30f895b1-2827-4f11-8a4c-1127938f-868b-4	Polhut Iktu Jakarta - Upacara		https://akcdn.d.klkhk		https://news.dk	4/1/2024 22:14	4/3/2024
5	1127938f-868b-4f11-8a4c-c34f9e9a-49c4-4	Kejari Ciamis Tal Ciamis - Kejaksan		https://akcdn.d.klkhk		https://www.dk	4/1/2024 17:49	4/3/2024
6	c34f9e9a-49c4-4f11-8a4c-30f8f2ae-58ca-4e	Hari Bhakti Rintu Jakarta -		https://akcdn.d.klkhk		https://news.dk	4/1/2024 17:17	4/3/2024
7	30f8f2ae-58ca-4f11-8a4c-0b2585a4-8314-4	Pemutihan 3,37 Jakarta -		https://akcdn.d.klkhk		https://news.dk	3/28/2024 20:36	4/3/2024
8	0b2585a4-8314-4f11-8a4c-a3b21600-9cfa-4	Dinilasi Sejak 2 Jakarta - Menteri		https://akcdn.d.klkhk		https://finance	3/28/2024 20:30	4/3/2024
9	a3b21600-9cfa-4f11-8a4c-d8c954f0-24e1-4	Wisata ATV di P. Bandung - Aksi		https://akcdn.d.klkhk		https://www.dk	3/28/2024 7:30	4/3/2024
10	d8c954f0-24e1-4f11-8a4c-1923c12b-0382-4	Proyek Blok Mas Jakarta - Proyek		https://akcdn.d.klkhk		https://finance	3/27/2024 17:16	4/3/2024
11	1923c12b-0382-4f11-8a4c-35860e72-182a-4	Gakkum KLHK Aj Jakarta - Majelis		https://akcdn.d.klkhk		https://news.dk	3/24/2024 22:22	4/3/2024
12	35860e72-182a-4f11-8a4c-55961e68-cc46-4	Lahan Bekas Tan Jakarta - PT		https://akcdn.d.klkhk		https://finance	3/24/2024 16:46	4/3/2024
13	55961e68-cc46-4f11-8a4c-f7a90d81-66af-47	Paglinggisik jenu Bandung - Bener		https://akcdn.d.klkhk		https://www.dk	3/24/2024 10:30	4/3/2024
14	f7a90d81-66af-4f11-8a4c-							

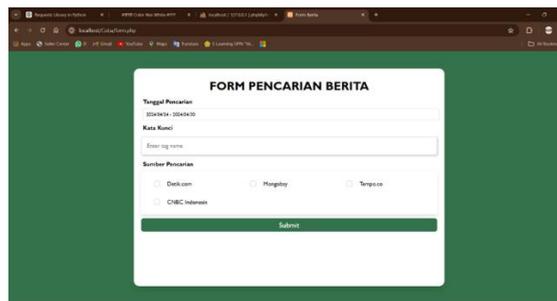
Sumber: Hasil Scraping (2024)
Gambar 2. Hasi Scraping Berita Online

Setelah pengumpulan data dari masing-masing portal berita *online* yang berhasil *disrape* selanjutnya yaitu penyimpanan data. Penyimpanan data yang telah *di-scrape* disimpan dalam basis data *MySQL* dengan skema yang telah ditentukan. Dalam menghubungkan hasil *scraping* kedalam *database* menggunakan sebuah API dengan mendefinisikan *endpoint* API menggunakan *Flask* untuk menerima hasil *scraping* dan menyimpannya ke dalam *database*.



Gambar 3. Uji Coba dan Validasi Sistem

Setelah basis data dipastikan telah berfungsi dengan baik dan memenuhi kebutuhan yang telah ditetapkan, pada gambar diatas merupakan proses uji coba serta validasi data terhadap input yang masuk pada *Simple Form* dengan *database*. Hasil *scraping* berdasarkan inputan tersebut dapat dilihat pada bagian “*response*” yang berisikan konten berita. Selanjutnya untuk mendukung hal tersebut dalam pemantauan berita, sebuah antarmuka pengguna (UI) yang sederhana namun fungsional dirancang.



Gambar 4. Tampilan UI/UX untuk *Simple Form*

UI ini berupa sebuah formulir sederhana (*Simple Form*) yang memudahkan pengguna untuk memasukkan parameter pencarian berita. *Simple Form* berisikan tanggal pencarian, kata kunci, dan sumber pencarian. Tanggal pencarian merupakan rentang tanggal mulai dan akhir dari periode pencarian berita. Kata kunci berfungsi untuk memasukkan banyak kata kunci yang akan digunakan dalam pencarian berita dimana dipisahkan dengan koma. Sumber pencarian berisikan *checkbox* list untuk memilih beberapa portal berita dari Detik, Mongabay, Tempo, dan CNBC. *Simple Form* juga dilengkapi dengan tombol submit untuk mengirimkan permintaan *scraping* berdasarkan parameter yang telah dimasukkan. Saat tombol ini diklik, *form* akan memvalidasi input dan mengirimkan data ke *Backendd* untuk diproses. Dengan desain UI/UX yang dirancang dengan baik, *Simple Form* ini diharapkan dapat membantu KLHK khususnya Biro Humas dalam mengajukan permintaan *scraping* berita secara mudah dan efisien, serta memastikan bahwa data yang dihasilkan relevan dan berguna untuk keperluan pemantauan media massa.

Simpulan

Penerapan teknik *web scraping* untuk membangun *website* terpadu pemantauan berita bagi Kementerian Lingkungan Hidup dan Kehutanan (KLHK) meningkatkan efisiensi dan efektivitas pemantauan informasi. Penggunaan bahasa pemrograman *Python* dan *library BeautifulSoup* memungkinkan pengumpulan data dari berbagai portal berita daring secara

otomatis, menggantikan metode manual yang tidak efisien. Proyek ini mencakup pengumpulan data, penyimpanan dalam *database*, pengembangan antarmuka pengguna, dan pengolahan data berdasarkan parameter pencarian. Setiap tahap dirancang untuk memastikan data yang dikumpulkan relevan dan mudah diakses. Hasilnya adalah *website* terpadu yang memungkinkan biro Humas KLHK memantau berita secara *real-time*, memperoleh wawasan yang lebih dalam, dan mengambil keputusan yang lebih tepat. Proyek ini menunjukkan komitmen KLHK dalam memanfaatkan teknologi digital untuk mendukung tugas dan fungsinya secara optimal.

Ucapan Terimakasih

Penulis mengucapkan terima kasih kepada seluruh pihak yang telah ikut berkontribusi dalam pembuatan *website* ini, sehingga *website* dapat diselesaikan dengan tepat waktu. Penulis telah mendapatkan berbagai macam pengalaman dan ilmu selama proses pembuatan aplikasi ini.

Daftar Pustaka

- Batubara, F. A. (2012). Perancangan *Website* Pada PT. Ratu Enim Palembang. *Jurnal Ilmu Pengetahuan dan Teknologi Terapan (REINTEK)*, 15-27. Retrieved from <https://jurnal.umsu.ac.id/index.php/reintek/article/view/252/196>
- Josi, A., Abdillah, L. A., & Suryayusra. (2014). Penerapan Teknik *Web Scraping* pada Mesin Pencari Artikel Ilmiah. Cornell University, 159-164.
- Mitra, V., Sujaini, H., & Negara, A. B. (2017). Rancang Bangun Aplikasi *Web scraping* untuk Korpus Paralel Indonesia - Inggris dengan Metode HTML DOM. *Jurnal Sistem dan Teknologi Informasi (JUSTIN)*, 36-41.
- Rizquina, A. Z., & Ratnasari, C. I. (2023). Implementasi *Web scraping* untuk Pengambilan Data Pada *Website E-Commerce*. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, 377-383. Retrieved from <http://jurnal.unidha.ac.id/index.php/jteksis/article/view/913/652>.