

# Implementasi *Website Berita Online* Menggunakan Metode *Crawling* Data Dengan Bahasa Pemrograman *Python*

Sendi Algifari Rismawan<sup>1</sup>, Yuda Syahidin<sup>2</sup>

<sup>1,2</sup>Politeknik Piksi Ganesha Bandung; Jl. Gatot Subroto No.301, Maleer, Kec. Batununggal, Kota Bandung, Jawa Barat 40274, (022) 87340030

<sup>3</sup>Sistem Informasi, IT & Komputer, Bandung

e-mail: \*<sup>1</sup>[piksi.sendi.19402038@gmail.com](mailto:piksi.sendi.19402038@gmail.com), <sup>2</sup>[yudasy@gmail.com](mailto:yudasy@gmail.com)

## Abstrak

*Implementasi web berita online merupakan salah satu kebutuhan pada era digital saat ini. Hal ini karena semakin banyak orang yang mendapatkan informasi melalui internet, dan berita menyebar lebih cepat melalui media online. Salah satu cara untuk mendapatkan informasi dari berita online adalah dengan menggunakan teknik pengambilan data crawling. Data crawling adalah program komputer yang secara otomatis mengambil data dari web. Metode ini memungkinkan pengumpulan data yang efisien dan terstruktur dari berbagai sumber. Penelitian ini juga akan membahas secara mendalam konsep metode pengambilan data crawling dan penggunaan bahasa pemrograman Python serta PHP (PHP: Hypertext Pre-processor) untuk mengimplementasikan web berita online, hingga proses pengambilan dan penyimpanan data dari berbagai sumber. Metodologi penelitian yang digunakan adalah metode Waterfall atau metode air mancur, yang umum digunakan dalam pengembangan perangkat lunak. Metode ini menyediakan pendekatan terstruktur untuk pengembangan sistem, yang digunakan sebagai referensi oleh penulis dalam menentukan langkah-langkah penelitian. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi untuk pengembangan dan inspirasi bagi pengembangan web berita online menggunakan teknik pengambilan data crawling. pada penelitian ini akan menghasilkan sebuah website crawling data yang mengumpulkan beberapa artikel dari berbagai website untuk disajikan pada website yang dibuat.*

**Kata kunci**—Crawling, Berita, Website, PHP (PHP: Hypertext Pre-processor), Python

## Abstract

*The implementation of online news websites is one of the needs in the current digital era. This is because more and more people obtain information through the internet, and news spreads faster through online media. One way to gather information from online news is by using data crawling techniques. Data crawling is a computer program that automatically retrieves data from the web. This method enables efficient and structured data collection from various sources. This research will also deeply discuss the concept of data crawling methods and the use of programming languages such as Python and PHP (PHP: Hypertext Pre-processor) to implement an online news website, including the process of retrieving and storing data from various sources. The research methodology used is the Waterfall method, which is commonly used in software development. This method provides a structured approach to system development, serving as a reference for the author in determining the research steps. The expected outcome of this research is to contribute to the development and inspiration for the development of online news websites using data crawling techniques. The research will result in a data crawling website that collects several articles from various websites to be presented on the created website.*

**Keywords**—Crawling, News, Website, PHP (PHP: Hypertext Pre-processor), Python

## 1. PENDAHULUAN

Implementasi *website* berita *online* merupakan suatu kebutuhan yang semakin meningkat di era digital saat ini. Hal ini dikarenakan semakin banyaknya masyarakat yang memperoleh informasi melalui internet dan semakin cepatnya informasi tersebut tersebar melalui berita *online*. Salah satu cara untuk mendapatkan informasi dari berita *online* adalah dengan menggunakan metode *crawling* data.

Metode *crawling* data adalah suatu teknik pengambilan data dari suatu *website* secara otomatis dengan menggunakan program komputer. Metode ini memungkinkan pengambilan data dari berbagai sumber secara efisien dan terstruktur, sehingga dapat digunakan untuk berbagai keperluan seperti penelitian, analisis, dan pengembangan aplikasi.

Dalam penelitian ini, akan diimplementasikan sebuah *website* berita *online* menggunakan metode *crawling* data dengan bahasa pemrograman *Python*. Penelitian ini bertujuan untuk membuat *website* berita *online* yang dapat menampilkan berita dari berbagai sumber secara terstruktur dan efisien.

Dalam penelitian ini, penulis akan membahas secara mendalam mengenai konsep metode *crawling* data penggunaan bahasa pemrograman *Python* serta bahasa pemrograman *PHP* (*PHP: Hypertext Pre-processor*) untuk implementasi *website* berita *online*, hingga proses pengambilan dan penyimpanan data dari berbagai sumber. Diharapkan hasil penelitian ini dapat memberikan kontribusi dan inspirasi bagi pengembangan *website* berita *online* menggunakan metode *crawling* data.

Metode *Crawling* data ini telah dilakukan oleh beberapa penulis sebelumnya, Penelitian yang dijadikan pembanding yaitu penelitian yang dilakukan oleh Hanifah dan Nurhasanah dengan judul penelitian “Implementasi *Web Crawling* Untuk Mengumpulkan *Web Crawling Implementation for Collecting*”. Penelitian ini dilakukan dengan tujuan mengumpulkan informasi mengenai tempat-tempat kuliner di Bandar Lampung melalui implementasi *web crawling*, serta menghasilkan informasi tentang tempat kuliner yang paling populer di kota tersebut. Hasil dari penelitian ini mencakup 129 informasi tentang restoran atau tempat kuliner yang terdapat di Bandar Lampung. Informasi-informasi tersebut memiliki tingkat keakuratan sebesar 71.1% dan mencakup alamat, kode pos, nomor telepon, kota, *website*, rating, dan jumlah *reviewer*. Dalam penelitian ini, proses *web crawling* dilakukan dengan menggunakan bahasa pemrograman *R* [2].

Penelitian berikutnya yang dijadikan pembanding adalah penelitian yang dilakukan oleh Dandi dan Evangs dengan judul “Implementasi *Web Crawling* Untuk Pencarian Harga *Sparepart* Pada PT Asuransi Sinar Mas”. Dalam penelitian tersebut data yang diperoleh sesuai serta dapat digunakan oleh tim *procurement* sebagai acuan dalam proses bisnis. Melalui penelitian ini, dapat disimpulkan bahwa *web crawling* merupakan salah satu metode yang dapat digunakan untuk mengumpulkan dan mengambil informasi dari suatu *website* [3].

Penelitian yang dilakukan oleh Jian Budiarto dengan judul “Identifikasi Kebutuhan Masyarakat Nusa Tenggara Barat pada Pandemi *Covid-19* di media sosial dengan menggunakan metode *crawling*”. Untuk mengidentifikasi kebutuhan masyarakat di tengah-tengah wabah *Covid-19* Gubernur dan Wakil gubernur NTB memerlukan informasi tersebut secara tepat dan tepat, namun ketika identifikasi dilakukan secara manual maka membutuhkan waktu yang tidak sedikit. Maka Gubernur dan Wakil Gubernur membutuhkan sebuah aplikasi *artificial intelligence* yang dapat menyajikan informasi secara otomatis dan *realtime*. Metode *Crawling* digunakan sebagai jawaban untuk mengatasi masalah tersebut, dengan metode tersebut setiap aspirasi masyarakat dapat di tampung dan dijadikan sebagai grafik untuk menentukan tingkat kebutuhan masyarakat NTB [4].

Penelitian terakhir yang hampir mirip adalah penelitian yang dilakukan oleh Zul Indra dan Liza Trisnawati dengan judul penelitian “Pengembangan Intelegent Data Collector Untuk Analisis Big Data Artikel Berita Online”. Dalam penelitian tersebut Zul dan Liza berhasil membuat sistem aplikasi yang diberi nama IDC (Intelegent Data Collector) untuk mengumpulkan berbagai macam artikel berita online yang selanjutnya artikel tersebut akan disimpan kedalam database lokal, yang kemudian database tersebut akan digunakan untuk berbagai macam data mining, seperti klasifikasi artikel berdasarkan topik serta sentimental terhadap publik [1].

Berdasarkan hasil penelitian yang dilakukan dan membandingkan dengan penelitian-penelitian terdahulu terlebih pada penelitian zul dan liza bahwa aplikasi yang dibangun tersebut namun pada penelitian tersebut hasil dari *crawling* data hanya disimpan dalam *database* lokal. Pada penelitian ini setelah artikel-artikel pada *website online* di *crawling* oleh sistem maka sistem akan memasukkan data ke *database* dan menampilkan kembali hasil dari pencarian tersebut guna sebagai bahan analisis yang dapat digunakan khalayak umum. Sehingga penelitian ini mampi menjadi penelitian terbaru yang dapat digunakan sebagai acuan kedepannya sebagai penelitian pembuatan website berita online dengan metode *ccrawling data*.

## 2. METODE PENELITIAN

### 2.1 Kajian Pustaka

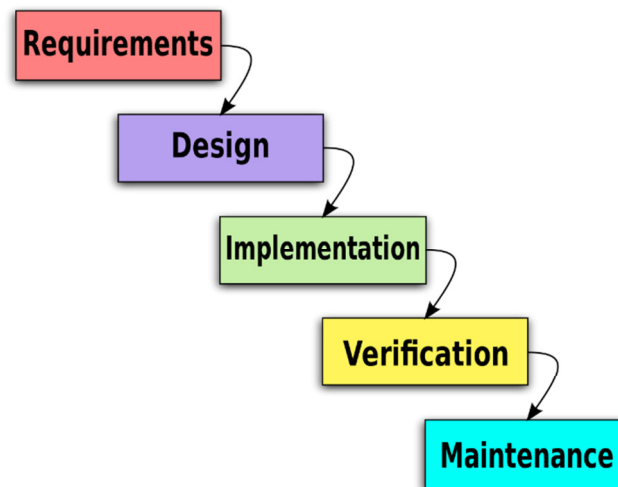
#### 2.1.1 Web Crawling

*Web Crawling* adalah program yang bekerja dengan metode tertentu dan secara otomatis mengumpulkan semua berita nasional yang ada dalam suatu *website*. Penggunaan teknologi *crawler* memungkinkan pengambilan informasi dilakukan secara otomatis. Untuk melakukan hal tersebut, diperlukan *crawler* yang dapat disesuaikan dengan karakteristik masing-masing media *online* [5].

Cara kerja *crawling data* yaitu web *crawler* akan menjelajahi setiap *URL* yang diberikan, dan mengunjungi *URL* tersebut secara satu per satu. Setiap halaman *URL* akan diidentifikasi apakah *URL* tersebut sudah masuk ke *database*. Jika ada *URL* yang belum masuk maka akan ditambahkan kedalam *database*. Siklus mengambil dan menambahkan list *URL* ini akan terus menerus dan berlanjut

#### 2.2 Metode Penelitian

Metode pengembangan yang digunakan dalam penelitian ini adalah metode waterfall atau metode air mancur, struktur metode ini menjadikannya metode yang umum dan banyak digunakan. Metode pengembangan inilah yang menjadi acuan penulis dalam menentukan langkah-langkah dalam penelitian. Metode Waterfall merupakan pengembangan sistem secara konseptual dan sistematis, dalam metode ini segala sesuatu harus dilakukan secara berurutan, tidak dapat menyelesaikan perancangan apabila belum dilakukan persyaratan, semua hal yang akan dilakukan harus berurutan dari mulai kebutuhan, perancangan, implementasi, verifikasi, dan pemeliharannya tentunya untuk meminimalisir terjadinya kesalahan yang mungkin terjadi [6].



Gambar 1. Metode Penelitian

### 2.2.1 Requirements

Untuk melakukan perencanaan yang baik, tahap analisis sistem menjadi hal yang sangat penting. Pada tahap ini, dilakukan analisis terhadap setiap data yang diperoleh guna menentukan sistem yang akan dibangun [7].

### 2.2.2 Design

Proses ini berfokus pada penentuan *UI/UX* yang akan diterapkan dalam aplikasi, penargetan *design* guna memperlihatkan gambaran terkait tampilan aplikasi yang akan dibangun serta ketersesuaian aplikasi dengan tujuan. Dalam proses ini penulis menggunakan *figma* sebagai media *design* penggambaran *design* [8].

### 2.2.3 Implementation

Pada proses ini penulis mengimplementasikan dari *design* yang sebelumnya telah dibuat menjadi suatu aplikasi menggunakan bahasa pemrograman *PHP* (*PHP: Hypertext Preprocessor*). *PHP: Hypertext Preprocessor* adalah suatu bahasa pemrograman *server-side scripting* yang digunakan bersama dengan *HTML* untuk membuat halaman *web* yang dinamis. Dikarenakan sifatnya sebagai *server-side scripting*, semua sintaks dan perintah dalam *PHP* akan dieksekusi di server terlebih dahulu, dan hasilnya akan dikirimkan ke *browser* dalam format *HTML* sehingga bahasa pemrograman *PHP* cocok untuk pembuatan *website* serta penggunaan bahasa pemrograman *python* sebagai pengelola data [9].

### 2.2.4 Verification

Setelah semua desain berhasil dibuat, maka akan diuji untuk setiap fiturnya. Pengujian dilakukan dengan menggunakan metode pengujian *Blackbox*. Pengujian ini bertujuan untuk mengetahui apakah setiap fitur dalam sistem dapat berfungsi dengan baik dan benar.

### 2.2.5 Maintenance

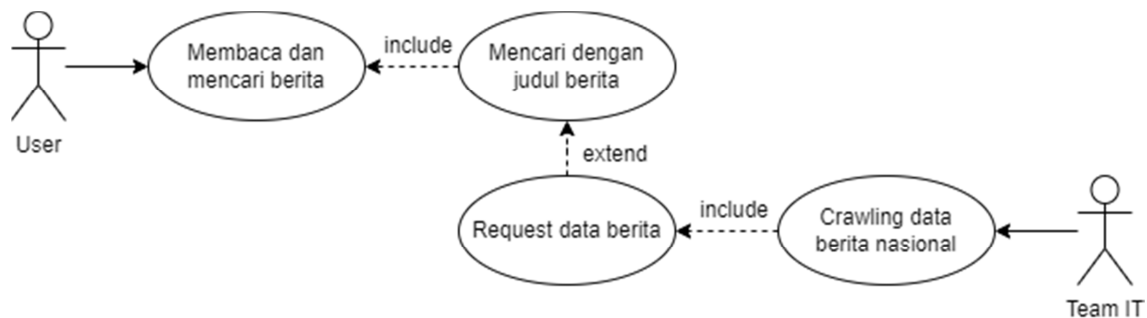
*Maintenance* adalah proses ketika aplikasi telah berhasil dibuat dan aplikasi akan diujikan sebelum aplikasi diserahkan kepada *user*.

### 2.3 Sistem Analis

Pada metode penelitian ini penulis menggunakan metode sistem analis *UML* (*Unified Modeling Language*). Dalam pengembangan sistem, sistem analis *UML* sangat penting untuk memastikan bahwa sistem yang dibangun dapat memenuhi kebutuhan pengguna dan berfungsi dengan baik. Terdapat beberapa jenis diagram *UML* yang biasa digunakan, diantaranya:

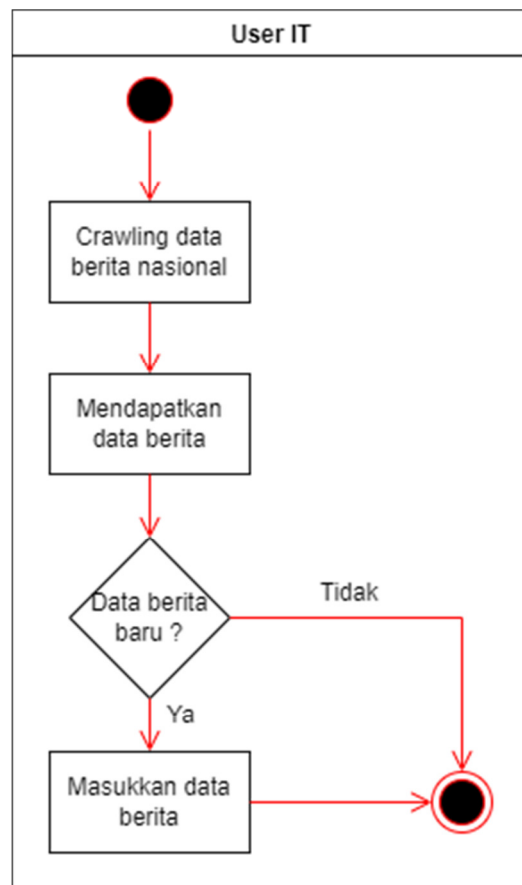
- Diagram *Use Case*: digunakan untuk memodelkan interaksi antara pengguna dan sistem.
- Diagram Aktivitas: digunakan untuk memodelkan alur kerja sistem dalam bentuk urutan kegiatan.
- Diagram Kelas: digunakan untuk memodelkan struktur dan hubungan antara kelas-kelas dalam sistem.
- Diagram Sekuens: digunakan untuk memodelkan urutan pesan atau komunikasi antar objek dalam sistem.
- Diagram Komponen: digunakan untuk memodelkan struktur dan hubungan antara komponen-komponen dalam sistem.

Dalam penelitian ini penulis hanya menggunakan jenis diagram *Usecase* dan *Activity*.



Gambar 1. Sistem Analis

Dalam penelitian ini, digunakan diagram *usecase* yang menunjukkan dua aktor, yaitu User dan *team IT*. User berperan sebagai aktor yang melakukan proses membaca dan mencari berita, sedangkan *team IT* berperan sebagai aktor yang melakukan proses *web crawling* pada *web* berita nasional dan menambahkan data berita yang belum ada di dalam *database*. Dalam proses pencarian berita, *team IT* selalu menggunakan judul berita yang ingin dicari.



Gambar 2. Activity Diagram

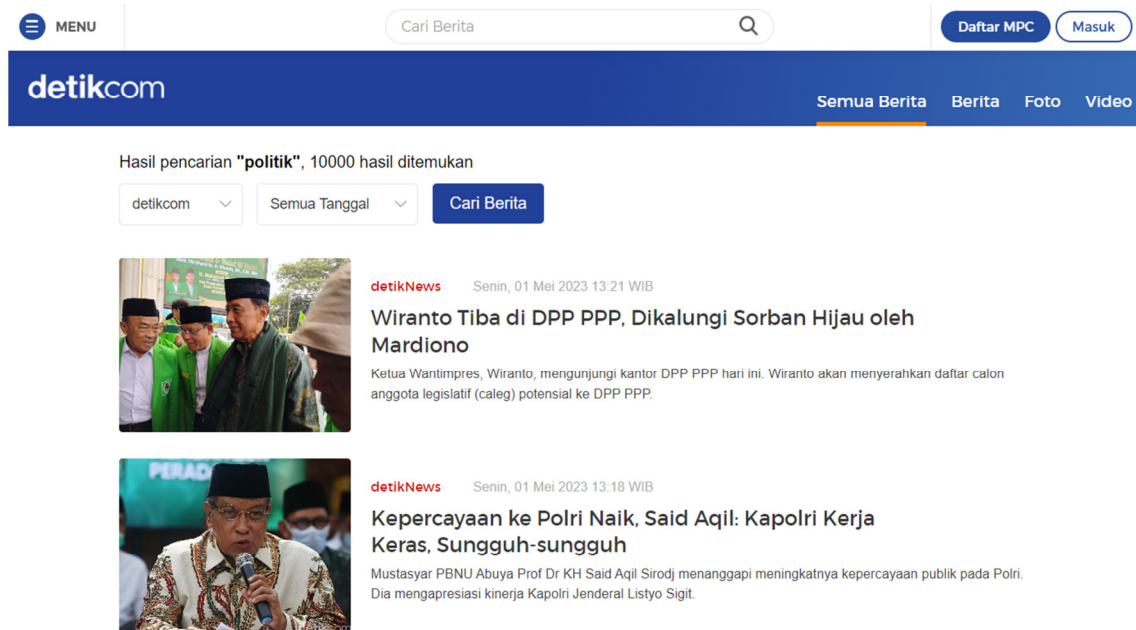
Diagram pada Gambar 2 menggambarkan proses *crawling* data berita. *User IT* akan memulai proses dengan melakukan *crawling* data berita pada *web* berita nasional. Setelah proses *crawling* selesai dan data berita berhasil didapatkan, *user IT* akan memeriksa apakah data berita tersebut merupakan data yang baru. Jika ya, maka data berita tersebut akan ditambahkan kedalam *database*. Namun, jika data tersebut sudah ada sebelumnya, maka data tersebut akan diabaikan dan *user IT* akan melanjutkan proses *crawling* data selanjutnya.

### 3. HASIL DAN PEMBAHASAN

Penelitian ini menghasilkan sebuah *website* yang dapat digunakan sebagai media membaca berita *online* menggunakan metode *crawling* data. Aplikasi ini mampu memberikan berita secara *uptodate* dari berbagai sumber berita yang ada [10].

#### 3.1 Crawling Data

Proses pertama yang dilakukan adalah *crawling* data dengan menggunakan fitur *search keyword* yang ada pada setiap berita *online* indonesia yang telah di definisikan. Berikut salah satu contoh fitur *search keyword* pada berita *online detik.com*.



Gambar 3. Halaman Pencarian

Gambar 3 merupakan salah satu contoh pencarian “politik” untuk kategori berita politik yang akan di *crawling*.



Gambar 4. Proses Data yang Berhasil di *Crawling* dari *Website*

Pada gambar 4 merupakan contoh hasil data yang berhasil di *crawling*. Apabila data tersebut terbaru maka akan masuk kedalam *database*.



### 3.2 Hasil Crawling Data

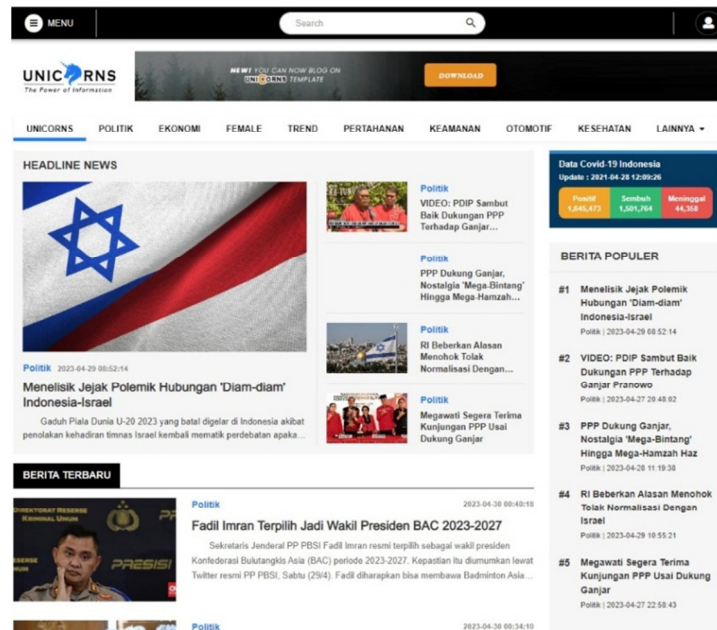
id	nama	s	judul	url_gambar	isi	tgl_publicasi	url_isi
a71779fc621fc9a732	Admin Okezone	3	Jenderal yang Dilengser	https://img.okezone.com/t	JAKARTA - Pada 1966 den	2023-04-03 00:00:00.000	https://nasional.okezone.com/read/2k
644dc7c81779fc621f	Admin Okezone	4	Jenderal TNI Berkari	https://img.okezone.com/t	JAKARTA Terdapat seju	2023-03-27 00:00:00.000	https://nasional.okezone.com/read/2k
644dc8381779fc621f	Admin Okezone	4	Pangkostrad Punya Karie	https://img.okezone.com/t	JENDERAL AH Nasution m	2023-03-06 00:00:00.000	https://nasional.okezone.com/read/2k
644dc7361779fc621f	Admin Okezone	5	Jenderal yang Diangkat J	https://img.okezone.com/t	PADA masa kepemimpinan	2023-04-28 00:00:00.000	https://nasional.okezone.com/read/2k
644dc7d31779fc621f	Admin Okezone	5	Kriteria Cawapres Anies:	https://img.okezone.com/t	JAKARTA - Bakal calon pre	2023-03-25 00:00:00.000	https://nasional.okezone.com/read/2k
64326bcc10b525062	Windy Wicaksono	5	Masjid di Jawa Barat, Fav	https://akcdn.detik.net.id/v	Bulan Ramadan menjadi w	2023-04-07 16:50:42.000	https://cnindonesia.com/gaya-hidup
644dc7851779fc621f	Admin Okezone	5	Negara yang Rela Merug	https://img.okezone.com/t	ADA 5 negara yang rela m	2023-04-08 00:00:00.000	https://bola.okezone.com/read/2023/
644dc7a11779fc621f	Admin Okezone	5	Profil Tokoh Golkar yang	https://img.okezone.com/t	GOLKAR atau Golongan Ki	2023-04-05 00:00:00.000	https://nasional.okezone.com/read/2k
644dc6ff1779fc621f	Pt. Viva Media Baru		Adian: Prabowo Belum Pur	https://thumb.viva.co.id/m	VIVA Politik - Elite PDI Perj	2023-04-29 12:15:31.000	https://www.viva.co.id/berita/politik/
64326bcc10b525062	Christina Andhika S		Ahli Ingatkan Risiko Wabal	https://akcdn.detik.net.id/v	Ahli peringatkan wabah ja	2023-04-06 08:03:36.000	https://cnindonesia.com/gaya-hidup
644dc42dc54373a9	Ahmad Bachrain		Airlangga Hartarto Bakal k	https://akcdn.detik.net.id/v	Ketua Umum Partai Golkar	2023-04-29 13:11:17.000	https://cnindonesia.com/nasional/2C
644dc77b1779fc621f	Admin Okezone		Airlangga Hartarto: Koalisi	https://img.okezone.com/t	JAKARTA - Ketua Umum D	2023-04-10 00:00:00.000	https://nasional.okezone.com/read/2k
644dc41ecb54373a9f	Dika Dania Kardi		Alasan PPP Usung Ganjar: f	https://akcdn.detik.net.id/v	Pit Ketua Umum Partai Per	2023-04-26 16:44:57.000	https://cnindonesia.com/nasional/2C
644dc6e91779fc621f	Pt. Viva Media Baru		Anak Buah Prabowo tegas	https://thumb.viva.co.id/m	VIVA Politik - Ketua Harian	2023-03-02 14:58:53.000	https://www.viva.co.id/berita/politik/
644dc6d91779fc621f	Pt. Viva Media Baru		Andre Rosiade: Gebrak Me	https://thumb.viva.co.id/m	VIVA Politik - Anggota Kor	2023-03-16 14:11:54.000	https://www.viva.co.id/berita/politik/
644dc77f1779fc621f	Admin Okezone		Asal Mula Status Daerah Is	https://img.okezone.com/t	Daerah Istimewa Yogyakarta	2023-04-08 00:00:00.000	https://nasional.okezone.com/read/2k
643289a31d0499604	Bimo Wiwoho		BPHN Kemenkumham Kon	https://akcdn.detik.net.id/v	Badan Pembinaan Hukum	2023-04-04 20:09:02.000	https://cnindonesia.com/nasional/2C
64326bcc10b525062	dzu		BPOM Cek Makanan Buka	https://akcdn.detik.net.id/v	Badan Pengawas Obat dar	2023-04-06 00:53:29.000	https://cnindonesia.com/nasional/2C
644dc79a1779fc621f	Admin Okezone		Berikan Pendidikan Politik	https://img.okezone.com/t	DEPOK - Indonesia akan m	2023-04-06 00:00:00.000	https://edukasi.okezone.com/read/20
644dc70c1779fc621f	Pt. Viva Media Baru		Cak Imin Disambut Hanga	https://thumb.viva.co.id/m	VIVA Politik - Ketua Umum	2023-04-28 18:23:19.000	https://www.viva.co.id/berita/politik/
644dc425cb54373a9f	Isnaini		Cerita Daso Saat Dittipi f	https://akcdn.detik.net.id/v	Ketua Harian DPP Partai G	2023-04-25 10:00:08.000	https://cnindonesia.com/nasional/2C
643289a31d0499604	Ahmad Bachrain		China Murka Presiden Taiw	https://akcdn.detik.net.id/v	China memprotes rencana	2023-04-04 19:14:09.000	https://cnindonesia.com/internasion
64326bcc10b525062	Dika Dania Kardi		Daftar Rute dan Keberang	https://akcdn.detik.net.id/v	Puncak arus mudik lebaran	2023-04-06 18:58:39.000	https://cnindonesia.com/nasional/2C
644dc447cb54373a9f	Pt. Viva Media Baru		Dani Ramdan Tak Lagi Diu	https://thumb.viva.co.id/m	VIVA Nasional - Pengamat	2023-04-12 16:28:47.000	https://www.viva.co.id/berita/nasiona
644dc7b01779fc621f	Admin Okezone		Dapet Pujian Elektabilitas	https://img.okezone.com/t	JAKARTA - Ketua Umum P	2023-04-02 00:00:00.000	https://nasional.okezone.com/read/2k
644dc6c11779fc621f	Pt. Viva Media Baru		Dasco Bocorkan Rencana f	https://thumb.viva.co.id/m	VIVA Politik - Ketua Harian	2023-04-13 13:21:18.000	https://www.viva.co.id/berita/politik/
644dc6c51779fc621f	Pt. Viva Media Baru		Dasco Buka-bukaan Sandi	https://thumb.viva.co.id/m	VIVA Politik - Menteri Pariw	2023-04-09 20:44:19.000	https://www.viva.co.id/berita/politik/
644dc6e11779fc621f	Pt. Viva Media Baru		Dasco Minta APH Tindak T	https://thumb.viva.co.id/m	VIVA Wakil Ketua DPR RI	2023-03-14 16:27:20.000	https://www.viva.co.id/siaran-pers/15
644dc76c1779fc621f	Admin Okezone		Demi Orde Baru, Ali Moer	https://img.okezone.com/t	DEMI berjayanya Orde Bar	2023-04-12 00:00:00.000	https://nasional.okezone.com/read/2k

Gambar 5. Contoh hasil *crawling* data

Pada gambar 5 merupakan contoh hasil *crawling* data yang telah masuk ke *database*. Pengujian *web crawling* ini menggunakan metode *Blackbox*. *Blackbox* adalah teknik pengujian program yang lebih menekankan pada evaluasi tampilan antarmuka dan fungsionalitas program, tanpa memperhatikan detail implementasi di dalam kode program tersebut. Pada penelitian ini, dilakukan pengujian *Blackbox* untuk memeriksa apakah data berita yang diambil dari hasil *web crawling* memiliki kesamaan dengan data berita pada *web* asli.

### 3.3 Menu Utama





Gambar 6. Menu Utama

Pada gambar 6 halaman menu utama bertujuan untuk memudahkan pengguna dalam melihat berita-berita yang tersedia, yang dapat dibagi berdasarkan popularitas, kategori, dan waktu terbaru. Dengan begitu, pengguna dapat dengan mudah menavigasi berita yang ingin mereka baca sesuai dengan preferensi masing-masing.

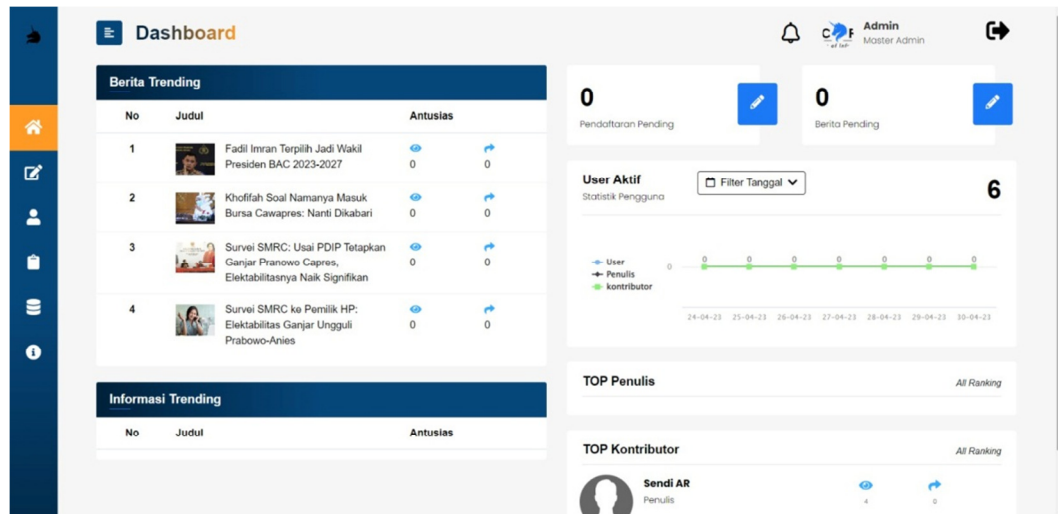
### 3.4 Login



Gambar 7. Login

Pada gambar 7 menampilkan halaman *login* yang dirancang untuk memungkinkan pengguna memberikan komentar pada sebuah berita atau membuat berita dan informasi baru. Dengan adanya halaman *login* ini, pengguna diharapkan dapat dengan mudah mengakses dan memanfaatkan fitur-fitur tertentu yang memerlukan otorisasi terlebih dahulu sebelum dapat digunakan.

### 3.5 Dashboard



Gambar 8. Dashboard

Pada gambar 8 menampilkan sebuah halaman *dashboard* yang berisi aktivitas-aktivitas dari para penulis, serta dapat membuat berita atau informasi baru. Halaman *dashboard* hanya dapat diakses oleh pengguna yang telah berhasil *login*.

## 4. KESIMPULAN

Implementasi *website* berita *online* dengan metode *crawling* data menjadi semakin penting di era digital saat ini. Metode ini memungkinkan pengambilan data dari berbagai sumber secara efisien dan terstruktur, sehingga dapat digunakan untuk berbagai keperluan seperti penelitian, analisis, dan pengembangan aplikasi. Dalam penelitian ini, akan diimplementasikan sebuah *website* berita *online* menggunakan metode *crawling* data dengan bahasa pemrograman *Python*. Metode pengembangan yang digunakan adalah metode *waterfall* atau metode air mancur.

Penelitian terdahulu telah membuktikan bahwa *web crawling* merupakan salah satu metode yang dapat digunakan untuk mengumpulkan dan mengambil informasi dari suatu *website* dengan hasil yang akurat. Beberapa penelitian sebelumnya menggunakan metode ini untuk mengumpulkan informasi tentang tempat kuliner, harga *sparepart*, dan kebutuhan masyarakat pada pandemi *Covid-19* di media sosial.

*Web crawling* adalah program yang bekerja dengan metode tertentu dan secara otomatis mengumpulkan semua berita nasional yang ada dalam suatu *website*. Cara kerja dari metode *crawling* data adalah *web crawling* akan menjelajahi setiap *URL* yang diberikan, akan mengunjungi *URL* tersebut satu per satu. Setiap page *URL* akan diidentifikasi apakah *URL*

tersebut sudah masuk ke *database*. Jika ada *URL* yang belum masuk maka akan ditambahkan kedalam *database*.

Penelitian ini diharapkan dapat memberikan kontribusi dan inspirasi bagi pengembangan *website* berita *online* menggunakan metode *crawling* data dengan bahasa pemrograman *Python*.

## 5. SARAN

Saran yang dapat penulis berikan berdasarkan kesimpulan dari pembuatan *website* Berita *Online* Menggunakan Metode *Crawling* Data yaitu:

1. Pada sistem *web crawling*, ada beberapa isi dari berita yang tidak *clean* karena setiap berita *online* memiliki karakteristik isi berita yang berbeda-beda dan terkadang berubah tampilan sehingga *web crawling* perlu terus direvisi.
2. Sistem *web crawling* akan terus dikembangkan apabila ada *website* berita *online* terbaru untuk memperkaya berita *online* yang telah tersebar di Indonesia.

## UCAPAN TERIMA KASIH

Pada kesempatan ini, penulis ingin menyampaikan ucapan terimakasih kepada pihak-pihak yang telah membantu serta memberikan dukungan sehingga Jurnal ini dapat terselesaikan dengan baik.

1. Dr. H. K. Prihartono A.H., Drs., S.Sos., S.Kom., M.M, MOS., CMA., MPM, selaku Direktur Politeknik Piksi Ganesha.
2. Yuda Syahidin, S.T., M.Kom., MTA., MOS, selaku Dosen Pembimbing Program Studi Manajemen Informatika yang telah meluangkan waktu serta memberikan bimbingan ilmu, arahan, dan motivasi kepada penulis untuk menyelesaikan Tugas Akhir ini.
3. Kedua orang tua tercinta yang selalu memberikan do'a serta dorongannya baik moral ataupun materil yang tidak terbatas.
4. Fitri Siti Nurjanah selaku kekasih tercinta penulis yang senantiasa membantu dan memberikan dorongan kepada penulis untuk menyelesaikan Tugas Akhir ini.
5. Keluarga besar yang selalu memberikan dukungan, nasehat, dan motivasi.
6. Rekan-Rekan seperjuangan yang selalu memberi kesan menarik, serta semangat warna yang indah selama perkuliahan di Politeknik Piksi Ganesha Bandung.
7. Serta seluruh pihak yang telah membantu pada saat pelaksanaan observasi maupun saat pembuatan laporan yang tidak bisa disebutkan satu per satu oleh penulis.

## DAFTAR PUSTAKA

- [1] Z. Indra and L. Trisnawati, "Pengembangan Intelligent Data Collector Untuk Analisis Big Data Artikel Berita Online," *Rabit J. Teknol. dan Sist. Inf. Univrab*, Vol. 3, No. 1, pp. 49–59, 2018, doi: 10.36341/rabit.v3i1.388.
- [2] R. Hanifah and I. S. Nurhasanah, "Implementasi Web Crawling Untuk Mengumpulkan Web Crawling Implementation for Collecting," *J. Teknol. Inf. dan Ilmu Komput.*, Vol. 5, No. 5, pp. 531–536, 2018, doi: 10.25126/jtiik20185842.

- 
- [3] D. Deddy, "Implementasi Web Crawling Untuk Pencarian Harga Sparepart pada PT Asuransi Sinar Mas," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, Vol. 7, No. 3, pp. 416–428, 2020, doi: 10.35957/jatisi.v7i3.505.
- [4] J. Budiarto, "Identifikasi Kebutuhan Masyarakat Nusa Tenggara Barat pada Pandemi Covid-19 di Media Sosial Dengan Metode Crawling (Requirements Identification for NTB People in Pandemic Covid-19 at Social Media Using Crawling Method)," Vol. 2, No. 4, pp. 244–250, 2021.
- [5] E. Zuliarso and K. Mustofa, "Crawling Web Berdasarkan Ontology," *J. Teknol. Inf. Din.*, Vol. XIV, No. 2, pp. 105–112, 2009, [Online]. Available: [https://repository.ugm.ac.id/32972/1/97-283-1-SM\\_\(1\).pdf](https://repository.ugm.ac.id/32972/1/97-283-1-SM_(1).pdf).
- [6] Syafnidawaty, "Metode Waterfall," 2020, [Online]. Available: <https://raharja.ac.id/2020/04/04/metode-waterfall/>.
- [7] S. Salamun and D. Arisandi, "Sistem Monitoring Kinerja Dosen pada Institusi Perguruan Tinggi Swasta (Studi Kasus: Universitas Abdurrab)," *Digit. Zo. J. Teknol. Inf. dan Komun.*, Vol. 11, No. 2, pp. 198–208, 2020, doi: 10.31849/digitalzone.v11i2.4304.
- [8] R. Refiani S, M. Hamdani, M. Muthmainnah, and Y. Syahidin, "Perancangan Sistem Informasi Akuntansi Penjualan Kain pada CV Vidi Sukses Berkah Menggunakan Microsoft Visual Basic 2010," *Fair Value J. Ilm. Akunt. dan Keuang.*, Vol. 5, No. 2, pp. 1159–1167, 2022, doi: 10.32670/fairvalue.v5i2.2441.
- [9] Y. Syahidin *et al.*, "Perancangan Sistem Informasi Pasien Appointment," *J. Cendikia*, Vol. XVIII, pp. 338–342, 2019.
- [10] M. S. Rohman, H. A. Santoso, G. W. Saraswati, N. Anisa, and S. Winarsih, "Pemanfaatan Topic-Focused Crawler Untuk Pembangunan Corpus Berita Bencana Menggunakan Teknik Scrapy CSS Selector," *Semin. Nas. APTIKOM*, p. 2019, 2019, [Online]. Available: <http://bpbdjateng.com>.