

MODUL 2

2.1 Capaian Praktikum Pertemuan 2

- Mahasiswa mampu melakukan instalasi BeautifulSoup.
- Mahasiswa mengimplementasikan BeautifulSoup untuk melakukan scraping.
- Mahasiswa mengimplementasikan request untuk melakukan scraping.

2.2 Indikator Capaian

Mahasiswa mampu melakukan instalasi dan konfigurasi modul eksternal seperti BeautifulSoup dan requests pada Python. Selain itu, mahasiswa mampu mengimplementasikan BeautifulSoup untuk melakukan web scraping dengan mengekstrak data dari halaman HTML secara terstruktur. Mahasiswa juga menunjukkan pemahaman dalam penggunaan modul requests untuk mengambil konten dari halaman web secara otomatis. Melalui praktikum ini, mahasiswa dapat memproses dan mengambil informasi spesifik dari struktur HTML berdasarkan tag dan atribut yang sesuai.

2.3 Landasan Teori

Web scraping merupakan teknik otomatisasi untuk mengambil data dari situs web secara terstruktur. Menurut Mitchell (2018), web scraping adalah metode untuk mengekstrak informasi dari halaman web dengan cara memanfaatkan program yang dapat mengakses, membaca, dan memproses konten HTML. Teknik ini sangat bermanfaat dalam pengumpulan data secara cepat dan efisien, terutama ketika data tidak disediakan dalam bentuk API atau file unduhan resmi.

Menurut Sweigart (2019), BeautifulSoup memungkinkan pengembang untuk menavigasi dan memanipulasi elemen HTML secara sederhana menggunakan pemahaman dasar struktur dokumen HTML dan CSS. Sementara itu, modul requests digunakan untuk mengirim permintaan HTTP dan mendapatkan respon halaman web dalam bentuk teks atau data lainnya.

Menurut Yulianto (2020), kemampuan mahasiswa dalam memahami dan mengimplementasikan modul-modul tersebut mencerminkan penguasaan dasar pemrograman Python serta kemampuan untuk memecahkan masalah berbasis data secara otomatis. Web scraping juga memberikan pemahaman praktis tentang bagaimana data dapat diakses, diambil, dan digunakan kembali secara sah untuk tujuan analisis, riset, maupun pengembangan aplikasi.

2.4 Pelaksanaan Praktikum

2.4.1 Percobaan Pertama

Mahasiswa diminta untuk melakukan instalasi dan implementasi modul BeautifulSoup dan requests pada Python dengan tujuan memahami dasar-dasar teknik web scraping, cara mengambil data dari halaman web secara otomatis, serta mengolah struktur HTML untuk mengekstrak informasi yang relevan.

a. Script / Setting Program

```
import requests
from bs4 import BeautifulSoup

url = 'https://www.antaranews.com/'

response = requests.get(url)

if response.status_code == 200:

    soup = BeautifulSoup(response.content, 'html.parser')

    links = soup.find_all('a', href=True)
    for link in links:

        if 'berita' in link['href']:
            print(f'Tautan Berita: https://www.antaranews.com{link["href"]}')

    categories = soup.find_all('li', class_='list-inline-item')
    for category in categories:
        tag = category.find('a')
        if tag:
            print(f'Kategori Berita: {tag.get_text(strip=True)}')

    times = soup.find_all('li', class_='list-inline-item')
    for time in times:
        span = time.find('span', class_='text-secondary')
        if span:
            print(f'Waktu: {span.get_text(strip=True)}')
    footer_links = soup.find_all('div', class_='widget__footer')
    for footer in footer_links:

        footer_items = footer.find_all('a', href=True)
```

```

        for item in footer_items:
            print(f"URL Footer: {item['href']}")

    section_headers = soup.find_all('h4', class_='border_section')
    for section_header in section_headers:

        links = section_header.find_all('a')
        for link in links:
            title = link.get('title') # Judul dari atribut title
            if title:
                print(f"Title: {title}")
    else:
        print('Gagal mengakses website.')

```

b. Langkah Uji Coba

- **import requests**, untuk mengambil konten dari website.
- **BeautifulSoup**, dari **bs4** untuk parsing HTML.
- Menentukan URL target (**<https://www.antaranews.com/>**) dan mengirim permintaan HTTP GET menggunakan **requests.get()**.
- **if response.status_code == 200**, mengecek apakah permintaan berhasil dengan status kode 200 (artinya halaman berhasil diakses).
- **soup = BeautifulSoup(response.content, 'html.parser')**, mengubah konten HTML dari respon menjadi objek BeautifulSoup agar mudah ditelusuri.
- **links = soup.find_all('a', href=True)**, mengambil semua <tag> yang memiliki atribut href.
- **for link in links: if 'berita' in link['href']**, menyaring kata yang mengandung 'berita' dalam tautannya.
- **categories = soup.find_all('li', class_='list-inline-item')**, mencari elemen dengan class 'list-inline-item'. Menampilkan teks dari link <a> di dalamnya sebagai kategori berita.
- **times = soup.find_all('li', class_='list-inline-item')**, masih dari elemen yang sama, kali ini dicari dengan class 'text-secondary' untuk mengambil informasi waktu publikasi.
- **footer_links = soup.find_all('div', class_='widget__footer')**, mencari semua elemen div dengan class 'widget__footer'. Mengambil semua link <a> di dalamnya dan mencetak URL-nya.

- `section_headers = soup.find_all('h4', class_='border_section')`, menelusuri elemen h4 dengan class `'border_section'`. Menampilkan isi atribut title dari setiap link `<a>` yang ditemukan, yang menggambarkan judul atau topik berita.
- Jika `response.status_code` bukan 200, maka mencetak: `print('Gagal mengakses website.')`

c. Hasil Uji Coba

```

PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS

tm_campaign=category_home
Tautan Berita: https://www.antaranews.comhttps://www.antaranews.com/berita/474471
tm_campaign=category_home
Tautan Berita: https://www.antaranews.comhttps://www.antaranews.com/berita/474462
sktop&utm_campaign=category_home
Tautan Berita: https://www.antaranews.comhttps://www.antaranews.com/berita/474462
sktop&utm_campaign=category_home
Tautan Berita: https://www.antaranews.comhttps://www.antaranews.com/berita/474441
s&utm_medium=desktop&utm_campaign=category_home
Tautan Berita: https://www.antaranews.comhttps://www.antaranews.com/berita/474441
s&utm_medium=desktop&utm_campaign=category_home
Tautan Berita: https://www.antaranews.comhttps://www.antaranews.com/berita/474414
rce=antaranews&utm_medium=desktop&utm_campaign=category_home
Tautan Berita: https://www.antaranews.comhttps://www.antaranews.com/berita/474414
rce=antaranews&utm_medium=desktop&utm_campaign=category_home
Tautan Berita: https://www.antaranews.comhttps://www.antaranews.com/berita/474289
desktop&utm_campaign=category_home

```

```

Tautan Berita: https://www.antaranews.comhttps://www.antaranews.com/berita/474289
Kategori Berita: Haji 2025
Kategori Berita: Judi Online
Kategori Berita: Konflik Iran Israel
Kategori Berita: Liga Champions
Kategori Berita: MBG
Kategori Berita: IKN Nusantara
Kategori Berita: Palestina
Kategori Berita: Konflik Rusia - Ukraina
Kategori Berita: Arus Balik
Kategori Berita: Arus Mudik
Kategori Berita: Piala AFF U-19
Kategori Berita: Kebakaran Los Angeles

```

```

Waktu: 1 jam lalu
Waktu: 2 jam lalu
Waktu: 2 jam lalu
Waktu: 2 jam lalu
Waktu: 5 jam lalu
Waktu: 5 jam lalu
Waktu: 6 jam lalu
Waktu: 9 jam lalu
Waktu: 10 jam lalu
Waktu: 11 jam lalu
Waktu: 5 menit lalu
Waktu: 11 menit lalu
Waktu: 11 menit lalu
Waktu: 13 menit lalu

```

```

URL Footer: https://www.antaranews.com/terkini
URL Footer: https://www.antaranews.com/top-news
URL Footer: https://www.antaranews.com/terpopuler
URL Footer: https://www.antaranews.com/pilihan-editor
URL Footer: https://www.antaranews.com/foto
URL Footer: https://www.antaranews.com/video
URL Footer: https://www.antaranews.com/infografik
URL Footer: https://www.antaranews.com/politik
URL Footer: https://www.antaranews.com/hukum
URL Footer: https://www.antaranews.com/ekonomi
URL Footer: https://www.antaranews.com/metro
URL Footer: https://www.antaranews.com/sepakbola
URL Footer: https://www.antaranews.com/olahraga

Title: Info Haji 2025
Title: Pilihan Editor
Title: Indeks Pilihan Editor
Title: Berita Terkini
Title: Indeks Berita Terkini
Title: Berita Terpopuler
Title: Telaah
Title: Indeks Telaah
Title: Politik
Title: Indeks Politik
Title: Hukum
Title: Indeks Hukum

```

Gambar 2.1 Hasil Uji Coba Scraping dari web antaresnews.com

d. Analisa Hasil

Program berhasil melakukan scraping pada berbagai elemen penting seperti tautan berita, kategori, waktu publikasi, link footer, dan judul. Beberapa data seperti tautan berita mengalami pengulangan, perlu difilter agar tidak mencetak data yang sama. Secara umum, struktur scraping berjalan sesuai dengan target dan informasi yang ditampilkan cukup lengkap untuk membentuk berita ringkas berbasis data dari situs.

2.5 Kesimpulan

2.5.1 Kesimpulan Percobaan 1

Pada percobaan ini menunjukkan bahwa mahasiswa telah berhasil mengimplementasikan teknik dasar web scraping menggunakan Python dengan memanfaatkan modul requests dan BeautifulSoup. Mahasiswa mampu mengakses

halaman situs berita antaranews.com, mengekstrak berbagai elemen penting seperti tautan berita, kategori, waktu publikasi, tautan pada bagian footer, dan judul.

Scraper berhasil menangkap informasi secara otomatis dari struktur HTML yang kompleks, membuktikan pemahaman mahasiswa terhadap teknik navigasi elemen web serta pengolahan data teks. Walaupun terdapat sedikit duplikasi pada hasil tautan berita, hal tersebut menjadi bagian dari pembelajaran dalam penyaringan dan pembersihan data (data cleaning).

Secara keseluruhan, percobaan ini mencerminkan keberhasilan mahasiswa dalam memahami konsep dasar *web scraping*, parsing HTML, serta penerapan logika pemrograman Python untuk otomatisasi pengambilan data dari web.

Mengetahui:

Dosen Pengampu Mata Kuliah

Arif Hadi Sumitro , M.Kom

NIKP. xxx

DAFTAR PUSTAKA

1. Mitchell, R. (2018). *Web Scraping with Python: Collecting More Data from the Modern Web* (2nd ed.). O'Reilly Media.
2. Sweigart, A. (2019). *Automate the Boring Stuff with Python: Practical Programming for Total Beginners* (2nd ed.). No Starch Press.
3. Yulianto, B. (2020). *Pemrograman Python untuk Data Sains dan Otomatisasi*. Deepublish.