Laporan Program
Scraping Data Gambar dari Website

Nama : Muhammad Alfin Rahmadani
Nim : 1123102125
Kelas : SP. 3.1

1. Deskripsi Program

Program ini dirancang untuk melakukan scraping data gambar dari sebuah situs web, dalam hal ini https://news.detik.com/. Program membaca konten halaman web, memindai elemen gambar (tag <img>), lalu mengunduh gambar-gambar tersebut ke direktori lokal.

2. Langkah-Langkah Implementasi

2.1 Mengirim Permintaan HTTP

Kode mengirimkan permintaan HTTP GET menggunakan pustaka requests untuk mengunduh konten HTML dari URL target:

```
response = requests.get(url)
```

2.2 Parsing Konten HTML

HTML yang diterima kemudian diparsing menggunakan pustaka BeautifulSoup, sehingga memungkinkan analisis struktur dokumen HTML:

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

2.3 Membuat Direktori untuk Menyimpan Gambar

Program memeriksa apakah direktori untuk menyimpan gambar (scraped_images) sudah ada. Jika belum, direktori tersebut dibuat:

```
if not os.path.exists(output_dir):
    os.makedirs(output_dir)
```

2.4 Mencari Elemen Gambar
Program mencari semua elemen <img> di halaman web dengan perintah:
images = soup.find_all('img')

2.5 Menentukan URL Gambar
Untuk setiap elemen <img>, program mencoba menemukan URL gambar menggunakan atribut berikut secara berurutan:

- data-src
- data-lazy-src
- srcset (memilih resolusi tertinggi jika ditemukan)
- src

Jika URL ditemukan, program menambahkan protokol (https:) jika URL bersifat relatif:

```
if img_url.startswith('//'):
    img_url = 'https:' + img_url
elif img_url.startswith('/'):
    img_url = url + img_url
```

2.6 Unduhan dan Penyimpanan Gambar
Gambar diunduh menggunakan requests.get(img_url).content dan disimpan dalam direktori lokal:

```
with open(filename, 'wb') as f:
    f.write(img_data)
```

3. Hasil Eksekusi

- Semua gambar yang ditemukan disimpan di direktori scraped_images.
- Nama file gambar diambil langsung dari URL sumber.
- Jika terjadi kesalahan selama proses pengunduhan, program akan mencetak pesan error.

4. Analisis Masalah

4.1 Gambar yang Diunduh Hanya Logo atau Thumbnail

Beberapa website menggunakan teknik pemuatan gambar dinamis (lazy loading) atau memiliki atribut berbeda untuk gambar utama. Hal ini menyebabkan program hanya mengambil gambar kecil seperti logo.
Solusi:

- Tambahkan logika untuk menangkap atribut lain seperti data-srcset atau elemen <source> di dalam <picture>.

- Gunakan Selenium untuk memuat halaman sepenuhnya jika gambar dimuat oleh JavaScript.

4.2 URL Gambar Tidak Lengkap

Beberapa URL gambar mungkin bersifat relatif, sehingga perlu menambahkan domain utama atau protokol untuk mendapatkan URL lengkap.

5. Saran Peningkatan

- Gunakan Selenium: Untuk memuat halaman web sepenuhnya, terutama jika gambar dimuat secara dinamis dengan JavaScript.

- Tambahkan Logika Atribut: Perluas pencarian atribut untuk mencakup elemen seperti <source> atau atribut khusus lain.

- Penyaringan URL: Pastikan hanya gambar yang relevan (bukan ikon kecil) yang diunduh dengan memfilter berdasarkan dimensi atau ukuran file.

- Penggunaan API Resmi: Jika tersedia, gunakan API situs web untuk akses data yang lebih aman dan legal.

6. Kesimpulan

Program ini berhasil melakukan scraping gambar dari website target, namun ada keterbatasan jika gambar dimuat secara dinamis atau memiliki atribut khusus. Dengan penyesuaian tambahan seperti menggunakan Selenium atau memperluas pencarian atribut, program dapat diperbaiki untuk mencakup lebih banyak jenis gambar.