

LAPORAN DATA SCRAPING

Disusun untuk Memenuhi Matakuliah Data Scraping
Dibimbing oleh Pelsri Ramadar N.S., M.Kom



Oleh:

FAVIAN NURRUDDIN

1123102138

**PROGRAM STUDI S1 TEKNIK INFORMATIKA
SEKOLAH TINGGI ILMU KOMPUTER PGRI BANYUWANGI
2024**

LAPORAN MATERI 8

Percobaan Code

Code Python ini dirancang untuk mengunduh gambar dari situs web tertentu (dalam hal ini, "https://www.jagatreview.com/") dan menyimpannya ke dalam direktori lokal.

a. Script / Setting Program

```
import os
import requests
import MainFungsi
from bs4 import BeautifulSoup

url = 'https://www.jagatreview.com/'
page = requests.get(url)
soup = BeautifulSoup(page.content, 'html.parser')
datax = soup.find_all('img')

images = []

for img in datax:
    img_url = img.get('data-src')
    if img_url is not None and img_url.endswith((''.jpg', '.png', '.gif')):
        images.append(img_url)

#print(datax)
direktori = "Hasil Gambar"
MainFungsi.CreateDirectory(direktori)

for gmb in images:
    response = requests.get(gmb)
    fileName = os.path.basename(gmb)
    MainFungsi.WriteToFile2(direktori, fileName, response)
    print(response)
    print(fileName)

print(images)
```

b. Langkah Uji Coba

Impor Modul

- **Os** : Mengelola interaksi dengan sistem file (misalnya, membuat folder).
- **Requests** : Mengirim permintaan HTTP untuk mengambil data dari web.

- **MainFungsi** : Modul kustom untuk fungsi tambahan, seperti membuat folder dan menyimpan file.
- **BeautifulSoup** : Mem-parsing HTML dan XML untuk mengekstrak data dengan mudah.

Mengambil Halaman Web

- **url = 'https://www.jagatreview.com/'** : Menetapkan URL yang akan diambil.
- **page = requests.get(url)** : Mengirim permintaan GET ke URL dan menyimpan responsnya.
- **soup = BeautifulSoup(page.content, 'html.parser')** : Menginisialisasi BeautifulSoup untuk mem-parsing konten HTML.

Mencari Gambar

- **datax = soup.find_all('img')** : Menemukan semua tag dalam halaman dan menyimpannya.
- **images = []** : Membuat daftar kosong untuk menyimpan URL gambar yang valid.

Mengambil URL Gambar

- **Loop** : Mengiterasi setiap tag gambar yang ditemukan.
- **img_url = img.get('data-src')** : Mengambil URL dari atribut data-src.
- **if img_url is not None and img_url.endswith(('jpg', 'png', 'gif')):** : Memeriksa apakah URL ada dan memiliki ekstensi gambar yang valid.
- **images.append(img_url)** : Menambahkan URL gambar ke dalam daftar jika valid.

Membuat Direktori untuk Menyimpan Gambar

- **direktori = "Hasil Gambar"** : Menetapkan nama folder untuk menyimpan gambar.
- **MainFungsi.CreateDirectory(direktori)** : Membuat folder jika belum ada.

Mengunduh dan Menyimpan Gambar

- **Loop kedua** : Mengiterasi daftar URL gambar yang valid.
- **response = requests.get(gmb)** : Mengambil gambar dari URL.

- **fileName = os.path.basename(gmb)** : Mengambil nama file dari URL gambar.
- **MainFungsi.WriteToFile2(direktori, fileName, response)** : Menyimpan gambar ke folder dengan nama file yang sesuai.
- **print(response)** : Mencetak respons untuk debugging.
- **print(fileName)** : Mencetak nama file gambar yang disimpan.

Output Akhir

- **print(images)** : Mencetak daftar URL gambar yang telah diproses.

c. Hasil Uji Coba



d. Analisa Hasil

Tujuan Kode

Kode ini bertujuan untuk mengambil gambar dari situs web <https://www.jagatreview.com/>, mengunduhnya, dan menyimpannya ke dalam direktori lokal.

Proses Utama

1. **Ambil Halaman Web** : Menggunakan “requests” untuk mendapatkan konten dari URL.
2. **Parsing HTML** : Menggunakan “BeautifulSoup” untuk memudahkan pencarian elemen HTML.
3. **Cari Gambar** : Mencari semua tag “” dan mengambil URL gambar dari atribut “data-src”.
4. **Validasi URL** : Memastikan URL gambar berakhir dengan ekstensi yang valid (jpg, png, gif).
5. **Buat Direktori** : Membuat direktori baru untuk menyimpan gambar jika belum ada.
6. **Unduh Gambar** : Mengunduh setiap gambar dan menyimpannya ke dalam direktori yang telah dibuat.

Mengetahui:

Dosen Pengampu Mata Kuliah

Pelsri Ramadar N.S., M.Kom

NIKP. 0704068408

PENUTUP

Kesimpulan

Dari percobaan Scraping diatas, kode ini memungkinkan untuk mengambil dan mengunduh gambar dari situs web, serta memahami struktur HTML dengan lebih baik. Saya menyadari pentingnya penanganan kesalahan dan etika dalam pengambilan data tanpa izin. Percobaan ini mengajarkan tentang pengorganisasian file dan tantangan yang ada dalam pengembangan software.