

Nama : Imas Nabellia Venda
NIM : 1123102116
Matkul : Data Scraping

Scraping Image

Kode python

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

url = 'https://eksepsionline.com/category/sastra/cerpen/'
page = requests.get(url)
soup = BeautifulSoup(page.content, 'html.parser')
data = soup.find_all('img')

images = [ ]

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

print(images)

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)
```

Penjelasan kode

- Import library os, request, MainFungsi, serta BeautifulSoup.
- URL website yang ingin diambil datanya disimpan dalam variabel url.
- Lakukan request HTTP GET ke URL yang sudah ditentukan untuk mendapatkan isi dari halaman web.
- Isi HTML diproses menggunakan BeautifulSoup dengan html.parser. Lalu program akan mencari semua tag yang ada di halaman tersebut.
- Terdapat list kosong yang nantinya akan digunakan untuk menyimpan URL gambar yang valid.
- Iterasi setiap elemen dalam daftar data dimulai dan akan mengambil nilai src yang ada pada tag .
- Pastikan bahwa URL gambar tidak kosong / none. Filter URL gambar dengan format akhiran jpg, png, jpeg, dan webp.
- URL gambar yang memenuhi syarat akan disimpan dalam list images dan kemudian akan ditampilkan.
- Buat direktori baru dengan nama "Hasil Gambar" menggunakan MainFungsi.CreateDirectory.

- Iterasi setiap URL gambar dalam list images.
- Buat permintaan HTTP GET untuk mengambil data gambar dan mendapatkan nama file dari URL.
- Simpan gambar ke direktori yang sudah dibuat.
- Program menampilkan status HTTP (200 artinya berhasil) dan nama file gambar yang berhasil disimpan.