

LAPORAN PRAKTIKUM
PRAKTIK PEMROGRAMAN PYTHON

**PRAKTIKUM
WEB SCARPPING**



Disusun oleh :
Aulia Diva Sukmadevi (V3923004)

Dosen
Yusuf Fadila Rachman. S.Kom., M.Kom

**PS D-III TEKNIK INFORMATIKA
SEKOLAH VOKASI
UNIVERSITAS SEBELAS MARET
2024**

A. HASIL DAN PEMBAHASAN

INPUT

1. DATA GAMBAR

```
In [20]: 1 import requests
          2 from bs4 import BeautifulSoup
          3 import csv
          4
          5 url = 'https://proxyway.com/news'
          6 r = requests.get(url)
          7
          8 soup = BeautifulSoup(r.content, 'html.parser')
          9
         10 images_list = []
         11
         12 images = soup.select('img')
         13
         14 for image in images:
         15     src = image.get('src')
         16     alt = image.get('alt')
         17     images_list.append({"src": src, "alt": alt})
         18
         19 for image in images_list:
         20     print(image)
         21
         22 with open('data gambar.csv' , 'w', encoding='UTF8') as f:
         23     writer = csv.writer(f)
         24
         25     writer.writerow(images_list)
```

Kode menggunakan library requests dan BeautifulSoup untuk melakukan web scraping dari halaman web <https://proxyway.com/news>. Pertama, kode mengambil konten HTML dari halaman tersebut menggunakan requests.get(url), lalu menguraikan konten HTML tersebut menggunakan BeautifulSoup untuk memudahkan ekstraksi informasi. Kemudian, kode mencari semua elemen di halaman web dengan menggunakan soup.select('img'). Untuk setiap elemen gambar yang ditemukan, kode mengekstrak atribut src (URL sumber) dan alt (teks alternatif) dari elemen tersebut. Informasi gambar-gambar ini kemudian disimpan dalam bentuk dictionary dan ditambahkan ke dalam list images_list.

Selanjutnya, kode mencetak informasi setiap gambar (URL sumber dan teks alternatif) dari `images_list` menggunakan loop `for`. Terakhir, kode menyimpan informasi gambar-gambar ke dalam file CSV dengan nama 'data gambar.csv' menggunakan `csv.writer`

2. DATA SUBJUDUL

```
In [11]: 1 import requests
          2 from bs4 import BeautifulSoup
          3 import csv
          4
          5 url = 'https://proxyway.com/news'
          6 r = requests.get(url)
          7
          8 if r.status_code == 200:
          9     soup = BeautifulSoup(r.text, 'html.parser')
         10     paragraphs = soup.find_all('h2')
         11
         12     with open('subjudul.csv', 'w', newline='', encoding='UTF8') as f:
         13         writer = csv.writer(f)
         14
         15         for paragraph in paragraphs:
         16             writer.writerow([paragraph.text])
         17             print(paragraph.text)
         18
         19 else:
         20     print('Gagal mengambil halaman:', r.status_code)
```

Kode menggunakan library `requests` dan `BeautifulSoup` untuk melakukan web scraping dari halaman `https://proxyway.com/news`. Tujuannya adalah untuk mengekstrak teks dari semua elemen `<h2>` (yaitu subjudul) pada halaman tersebut, kemudian menyimpan teks subjudul ke dalam file CSV dengan nama 'subjudul.csv'. Pertama, kode melakukan permintaan HTTP GET ke URL 'https://proxyway.com/news' menggunakan `requests.get(url)`. Setelah mendapatkan respons dari permintaan tersebut, kode memeriksa status kode respons HTTP (`r.status_code`). Jika respons adalah 200 (artinya permintaan berhasil), maka kode menggunakan `BeautifulSoup` untuk menguraikan konten HTML halaman tersebut.

Selanjutnya, kode mencari semua elemen `<h2>` di dalam konten HTML menggunakan `soup.find_all('h2')`. Informasi teks dari setiap elemen `<h2>` kemudian diambil dan disimpan dalam list `paragraphs`. Kemudian, kode membuka file CSV dengan mode write ('w') menggunakan `csv.writer` untuk menulis teks subjudul dari `paragraphs` ke dalam file CSV. Setiap teks subjudul ditulis sebagai satu baris dalam file CSV. Terakhir, jika terjadi kesalahan dalam mengambil halaman (status kode respons HTTP bukan 200), kode mencetak pesan bahwa halaman tidak dapat diambil.

3. DATA KETERANGAN

```

In [1]: 1 import requests
        2 from bs4 import BeautifulSoup
        3 import csv
        4
        5 url = 'https://proxyway.com/news'
        6
        7 response = requests.get(url)
        8
        9 if response.status_code == 200:
        10     soup = BeautifulSoup(response.content, 'html.parser')
        11     div_elements = soup.find_all('div', attrs={'data-widget_type': 'theme-post-excerpt.default'})
        12
        13     if div_elements:
        14
        15         with open('keterangan.csv', 'w', newline='', encoding='utf-8') as csvfile:
        16             csvwriter = csv.writer(csvfile)
        17
        18             for div_element in div_elements:
        19                 inner_div = div_element.find('div', class_='elementor-widget-container')
        20                 if inner_div:
        21                     text_content = inner_div.get_text(strip=True)
        22                     csvwriter.writerow([text_content])
        23                     print(text_content)
        24     else:
        25         print(f'Gagal mengambil halaman. Status code: {response.status_code}')
        26

```

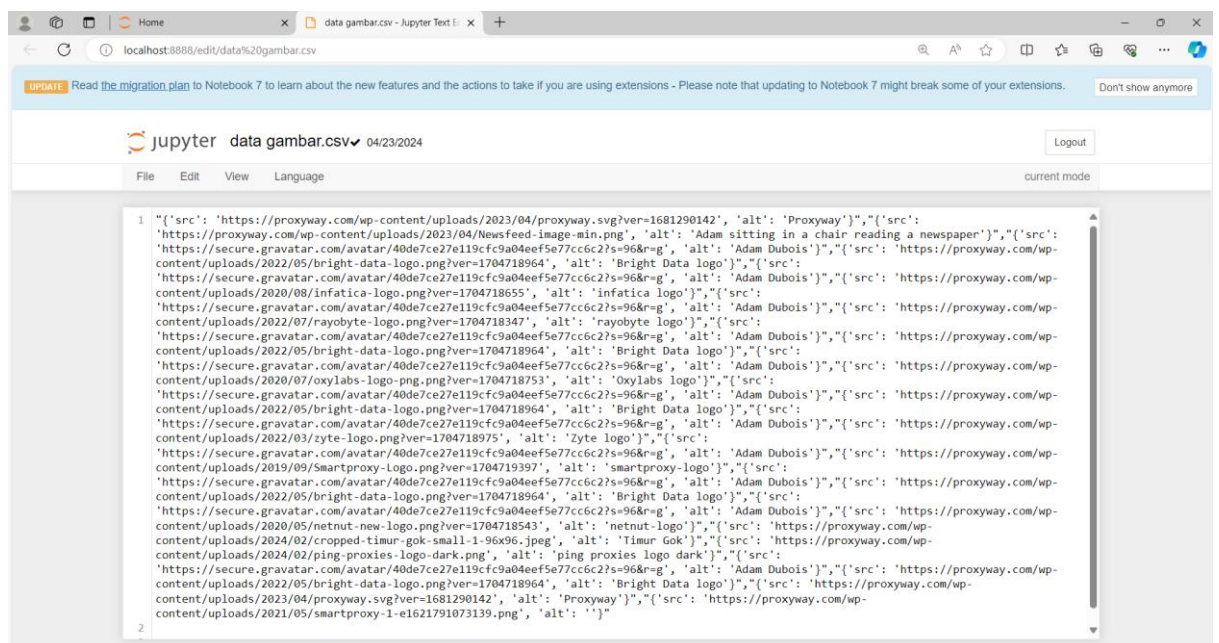
Kode digunakan untuk melakukan web scraping dari halaman <https://proxyway.com/news> dengan menggunakan library requests dan BeautifulSoup. Tujuan dari kode ini adalah untuk mengekstrak teks konten dari elemen <div> yang memiliki atribut khusus (data-widget_type='theme-post-excerpt.default') pada halaman tersebut, dan kemudian menyimpan teks tersebut ke dalam file CSV dengan nama 'keterangan.csv'. Prosesnya dimulai dengan mengirimkan permintaan HTTP GET ke URL target menggunakan requests.get(url). Jika permintaan berhasil (status kode 200), maka konten HTML dari halaman tersebut diuraikan menggunakan BeautifulSoup. Selanjutnya, kode mencari semua elemen <div> yang memiliki atribut data-widget_type='theme-post-excerpt.default' menggunakan soup.find_all('div', attrs={'data-widget_type': 'theme-post-excerpt.default'})). Setelah itu, untuk setiap elemen <div> yang ditemukan, kode mengekstrak teks konten dari elemen <div> dalamnya menggunakan .find('div', class_='elementor-widget-container'). Jika teks konten berhasil ditemukan, maka teks tersebut ditambahkan ke dalam file CSV dengan menggunakan csv.writer.writerow([text_content]). Selama proses ekstraksi dan penyimpanan, teks konten juga dicetak ke konsol dengan print(text_content). Jika ada kesalahan dalam mengambil halaman (misalnya status kode HTTP bukan 200), maka pesan kesalahan akan dicetak di konsol.

OUTPUT

1. DATA GAMBAR

```
{'src': 'https://proxyway.com/wp-content/uploads/2023/04/proxyway.svg?ver=1681290142', 'alt': 'Proxyway'}
{'src': 'https://proxyway.com/wp-content/uploads/2023/04/Newsfeed-image-min.png', 'alt': 'Adam sitting in a chair reading a newspaper'}
{'src': 'https://secure.gravatar.com/avatar/40de7ce27e119cfc9a04eef5e77cc6c2?s=96&r=g', 'alt': 'Adam Dubois'}
{'src': 'https://proxyway.com/wp-content/uploads/2022/05/bright-data-logo.png?ver=1704718964', 'alt': 'Bright Data logo'}
{'src': 'https://secure.gravatar.com/avatar/40de7ce27e119cfc9a04eef5e77cc6c2?s=96&r=g', 'alt': 'Adam Dubois'}
{'src': 'https://proxyway.com/wp-content/uploads/2020/08/infatica-logo.png?ver=1704718655', 'alt': 'infatica logo'}
{'src': 'https://secure.gravatar.com/avatar/40de7ce27e119cfc9a04eef5e77cc6c2?s=96&r=g', 'alt': 'Adam Dubois'}
{'src': 'https://proxyway.com/wp-content/uploads/2022/07/rayobyte-logo.png?ver=1704718347', 'alt': 'rayobyte logo'}
{'src': 'https://secure.gravatar.com/avatar/40de7ce27e119cfc9a04eef5e77cc6c2?s=96&r=g', 'alt': 'Adam Dubois'}
{'src': 'https://proxyway.com/wp-content/uploads/2022/05/bright-data-logo.png?ver=1704718964', 'alt': 'Bright Data logo'}
{'src': 'https://secure.gravatar.com/avatar/40de7ce27e119cfc9a04eef5e77cc6c2?s=96&r=g', 'alt': 'Adam Dubois'}
{'src': 'https://proxyway.com/wp-content/uploads/2020/07/oxylabs-logo.png.png?ver=1704718753', 'alt': 'Oxylabs logo'}
{'src': 'https://secure.gravatar.com/avatar/40de7ce27e119cfc9a04eef5e77cc6c2?s=96&r=g', 'alt': 'Adam Dubois'}
{'src': 'https://proxyway.com/wp-content/uploads/2022/05/bright-data-logo.png?ver=1704718964', 'alt': 'Bright Data logo'}
{'src': 'https://secure.gravatar.com/avatar/40de7ce27e119cfc9a04eef5e77cc6c2?s=96&r=g', 'alt': 'Adam Dubois'}
{'src': 'https://proxyway.com/wp-content/uploads/2022/03/zyte-logo.png?ver=1704718975', 'alt': 'Zyte logo'}
{'src': 'https://secure.gravatar.com/avatar/40de7ce27e119cfc9a04eef5e77cc6c2?s=96&r=g', 'alt': 'Adam Dubois'}
{'src': 'https://proxyway.com/wp-content/uploads/2019/09/Smartproxy-Logo.png?ver=1704719397', 'alt': 'smartproxy-logo'}
{'src': 'https://secure.gravatar.com/avatar/40de7ce27e119cfc9a04eef5e77cc6c2?s=96&r=g', 'alt': 'Adam Dubois'}
{'src': 'https://proxyway.com/wp-content/uploads/2022/05/bright-data-logo.png?ver=1704718964', 'alt': 'Bright Data logo'}
{'src': 'https://secure.gravatar.com/avatar/40de7ce27e119cfc9a04eef5e77cc6c2?s=96&r=g', 'alt': 'Adam Dubois'}
{'src': 'https://proxyway.com/wp-content/uploads/2020/05/netnut-new-logo.png?ver=1704718543', 'alt': 'netnut-logo'}
{'src': 'https://proxyway.com/wp-content/uploads/2024/02/cropped-timur-gok-small-1-96x96.jpeg', 'alt': 'Timur Gok'}
{'src': 'https://proxyway.com/wp-content/uploads/2024/02/ping-proxies-logo-dark.png', 'alt': 'ping proxies logo dark'}
{'src': 'https://secure.gravatar.com/avatar/40de7ce27e119cfc9a04eef5e77cc6c2?s=96&r=g', 'alt': 'Adam Dubois'}
{'src': 'https://proxyway.com/wp-content/uploads/2022/05/bright-data-logo.png?ver=1704718964', 'alt': 'Bright Data logo'}
{'src': 'https://proxyway.com/wp-content/uploads/2023/04/proxyway.svg?ver=1681290142', 'alt': 'Proxyway'}
{'src': 'https://proxyway.com/wp-content/uploads/2021/05/smartproxy-1-e1621791073139.png', 'alt': ''}
```

Dalam Bentuk CSV



2. DATA SUBJUDUL

Bright Data Equalizes Residential and Mobile Proxy Rates
Rayobyte Publishes a Whitepaper on Preventing Proxy Abuse
ScrapeCon 2024: A Recap
Infatica Makes Its Residential Proxies Up to 43% Cheaper
Rayobyte Decreases Residential Proxy Prices, Moves to Subscription
Bright Data's ScrapeCon to Take Place on April 2
Oxylabs Cuts Residential Proxy Rates by Up to 20%
Bright Data Reduces Residential Proxy Prices by 20%
Zyte Adds AI Scraping Functionality to Its API
Smartproxy Slashes Residential Proxy Prices by Up to 25%
Meta Drops the Case Against Bright Data
NetNut Launches Website Unblocker, SERP API, Professional Datasets

Dalam bentuk CSV

The screenshot shows a Jupyter Notebook interface with a browser window at the top displaying tabs for 'Home', '*subjudul.csv - Jupyter Text Editor', and 'Web_Scraping - Jupyter Notebook'. The address bar shows 'localhost:8888/edit/subjudul.csv'. Below the browser, a blue banner contains an 'UPDATE' message. The Jupyter logo and file name 'subjudul.csv' with a date '04/26/2024' are visible. A menu bar includes 'File', 'Edit', 'View', and 'Language'. The main area displays a list of 12 items, each with a line number and a title related to proxy services and web scraping.

```
1 Bright Data Equalizes Residential and Mobile Proxy Rates
2 Rayobyte Publishes a Whitepaper on Preventing Proxy Abuse
3 ScrapeCon 2024: A Recap
4 Infatica Makes Its Residential Proxies Up to 43% Cheaper
5 "Rayobyte Decreases Residential Proxy Prices, Moves to Subscription"
6 Bright Data's ScrapeCon to Take Place on April 2
7 Oxylabs Cuts Residential Proxy Rates by Up to 20%
8 Bright Data Reduces Residential Proxy Prices by 20%
9 Zyte Adds AI Scraping Functionality to Its API
10 Smartproxy Slashes Residential Proxy Prices by Up to 25%
11 Meta Drops the Case Against Bright Data
12 "NetNut Launches Website Unblocker, SERP API, Professional Datasets"
```

3. DATA KETERANGAN

26

In essence, the provider's mobile proxies just got 65% cheaper.
The document describes Rayobyte's three-pronged approach against bad actors.
Our impressions from Bright Data's first virtual conference on web scraping.
The provider's plans now give significantly more traffic for the same price.
The new rates are up to 50% cheaper in lower ranges.
The rescheduled web scraping event will proceed in a week.
The reduction affects three entry plans.
The second round of price cuts continues.
The tool can now crawl, unblock, and parse websites using AI and an optional no-code interface.
It's the second price cut in less than a year.
The dispute results in a complete win for the data collection infrastructure provider.
The Israeli proxy provider moves up the data extraction value chain.

Dalam Bentuk CSV

The screenshot shows a Jupyter Notebook interface with a browser window at the top displaying tabs for 'Home', 'keterangan.csv - Jupyter Text Editor', and '+'. The address bar shows 'localhost:8888/edit/keterangan.csv'. Below the browser, a blue banner contains an 'UPDATE' message. The Jupyter logo and file name 'keterangan.csv' with a date '04/29/2024' are visible. A menu bar includes 'File', 'Edit', 'View', and 'Language'. The main area displays the same 12 items as the previous screenshot, but each item is now enclosed in double quotes, indicating it is a CSV file.

```
1 "In essence, the provider's mobile proxies just got 65% cheaper."
2 "The document describes Rayobyte's three-pronged approach against bad actors."
3 "Our impressions from Bright Data's first virtual conference on web scraping."
4 "The provider's plans now give significantly more traffic for the same price."
5 "The new rates are up to 50% cheaper in lower ranges."
6 "The rescheduled web scraping event will proceed in a week."
7 "The reduction affects three entry plans."
8 "The second round of price cuts continues."
9 "The tool can now crawl, unblock, and parse websites using AI and an optional no-code interface."
10 "It's the second price cut in less than a year."
11 "The dispute results in a complete win for the data collection infrastructure provider."
12 "The Israeli proxy provider moves up the data extraction value chain."
13
```