

**LAPORAN PRAKTIKUM
PRAKTIK PEMROGRAMAN PYTHON**

**PRAKTIKUM
WEB SCARPPING**



Disusun Oleh:

Lia Fitriani (V3923013)

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



- Mendefinisikan sebuah fungsi bernama scrape_proxyway_images.
- Menyimpan URL halaman web yang akan di-scrape ke dalam variabel url.
- Memanggil fungsi fetch_content(url) untuk mendapatkan konten halaman. Fungsi fetch_content diharapkan mengembalikan isi halaman web jika permintaan berhasil atau None jika gagal.
- Memeriksa apakah konten berhasil diambil (content tidak None).
- Membuat objek BeautifulSoup dari konten HTML untuk memudahkan manipulasi dan pencarian elemen dalam HTML. Membuat list kosong images_list yang akan digunakan untuk menyimpan URL dan atribut alt dari gambar. Menggunakan metode select dari BeautifulSoup untuk menemukan semua elemen dalam halaman.
- Melakukan iterasi melalui setiap tag yang ditemukan:
- Mengambil nilai atribut src dari tag dan menyimpannya dalam variabel src.
- Mengambil nilai atribut alt dari tag dan menyimpannya dalam variabel alt.
- Menambahkan pasangan [src, alt] ke dalam images_list.
- Memanggil fungsi save_to_csv untuk menyimpan images_list ke file CSV dengan nama data_gambar.csv. Fungsi ini juga menulis header CSV ['src', 'alt'].
- Melakukan iterasi melalui images_list dan mencetak setiap elemen. Setiap elemen dalam images_list adalah list yang berisi URL gambar dan teks alt.

2. Data Subjudul



- Mendefinisikan sebuah fungsi bernama scrape_proxyway_subtitles.
- Menyimpan URL halaman web yang akan diambil kontennya dalam variabel url.
- Memanggil fungsi fetch_content(url) untuk mengirim permintaan HTTP ke URL dan mengambil kontennya. Fungsi fetch_content diharapkan mengembalikan isi halaman jika permintaan berhasil atau None jika gagal.

- Mengecek apakah content tidak bernilai None. Jika content bernilai None, berarti ada kegagalan dalam mengambil konten dari URL.
- Membuat objek BeautifulSoup dari konten HTML yang diambil sehingga dapat dengan mudah diolah dan diekstraksi.
- Menggunakan metode find_all dari BeautifulSoup untuk menemukan semua elemen <h2> dalam halaman web. Metode find_all mengembalikan daftar semua elemen yang cocok dengan tag yang diberikan
- Menggunakan list comprehension untuk membuat daftar subtitles yang berisi teks dari setiap elemen <h2>. Setiap teks disimpan sebagai elemen tunggal dalam sebuah list, sehingga subtitles adalah list of lists.
- Memanggil fungsi save_to_csv untuk menyimpan subtitles ke dalam file CSV dengan nama subjudul.csv. Fungsi ini menulis data dalam format CSV.
- Melakukan iterasi melalui subtitles dan mencetak setiap elemen. Karena setiap elemen dalam subtitles adalah list yang berisi satu elemen teks, kita mencetak elemen pertama dari list tersebut (subtitle[0]).

3. Data Keterangan



- Mendefinisikan sebuah fungsi bernama scrape_proxyway_descriptions.
- Menyimpan URL halaman web yang akan diambil kontennya dalam variabel url.
- Memanggil fungsi fetch_content(url) untuk mengirim permintaan HTTP ke URL dan mengambil kontennya. Fungsi fetch_content diharapkan mengembalikan isi halaman jika permintaan berhasil atau None jika gagal.
- Mengecek apakah content tidak bernilai None. Jika content bernilai None, berarti ada kegagalan dalam mengambil konten dari URL.
- Membuat objek BeautifulSoup dari konten HTML yang diambil sehingga dapat dengan mudah diolah dan diekstraksi.
- Menggunakan metode find_all dari BeautifulSoup untuk menemukan semua elemen <div> yang memiliki atribut data-widget_type dengan nilai theme-post-excerpt.default. Metode find_all mengembalikan daftar semua elemen yang cocok dengan atribut yang diberikan
- Membuat list kosong descriptions yang akan digunakan untuk menyimpan deskripsi yang ditemukan.
- Melakukan iterasi melalui setiap elemen <div> yang ditemukan:
- inner_div = div_element.find('div', class_='elementor-widget-container'): Mencari elemen <div> anak dengan kelas elementor-widget-container.
- if inner_div:: Memeriksa apakah inner_div ditemukan.
- text_content = inner_div.get_text(strip=True): Mengambil teks dari inner_div dan menghapus spasi kosong di awal dan akhir teks.
- descriptions.append([text_content]): Menambahkan teks deskripsi ke dalam descriptions sebagai elemen tunggal dalam sebuah list.

- `print(text_content)`: Mencetak teks deskripsi ke konsol.
- Memanggil fungsi `save_to_csv` untuk menyimpan descriptions ke dalam file CSV dengan nama keterangan.csv. Fungsi ini menulis data dalam format CSV.

Output

1. Data Gambar

```
[ 'https://proxyway.com/wp-content/uploads/2023/04/proxyway.svg?ver=1681290142', 'Proxyway' ]
[ 'https://proxyway.com/wp-content/uploads/2023/04/Newsfeed-image-min.png', 'Adam sitting in a chair reading a newspaper' ]
[ 'https://secure.gravatar.com/avatar/40de7ce27e119cfc9a04eef5e77cc6c2?s=96&r=g', 'Adam Dubois' ]
[ 'https://proxyway.com/wp-content/uploads/2020/07/oxylabs-logo.png.png?ver=1704718753', 'Oxylabs logo' ]
[ 'https://secure.gravatar.com/avatar/40de7ce27e119cfc9a04eef5e77cc6c2?s=96&r=g', 'Adam Dubois' ]
[ 'https://proxyway.com/wp-content/uploads/2022/05/bright-data-logo.png.png?ver=1704718964', 'Bright Data logo' ]
[ 'https://secure.gravatar.com/avatar/40de7ce27e119cfc9a04eef5e77cc6c2?s=96&r=g', 'Adam Dubois' ]
[ 'https://proxyway.com/wp-content/uploads/2022/07/rayobyte-logo.png.png?ver=1704718347', 'rayobyte logo' ]
[ 'https://secure.gravatar.com/avatar/40de7ce27e119cfc9a04eef5e77cc6c2?s=96&r=g', 'Adam Dubois' ]
[ 'https://proxyway.com/wp-content/uploads/2022/05/bright-data-logo.png.png?ver=1704718964', 'Bright Data logo' ]
[ 'https://secure.gravatar.com/avatar/40de7ce27e119cfc9a04eef5e77cc6c2?s=96&r=g', 'Adam Dubois' ]
[ 'https://proxyway.com/wp-content/uploads/2020/08/infatica-logo.png.png?ver=1704718655', 'infatica logo' ]
[ 'https://secure.gravatar.com/avatar/40de7ce27e119cfc9a04eef5e77cc6c2?s=96&r=g', 'Adam Dubois' ]
[ 'https://proxyway.com/wp-content/uploads/2022/07/rayobyte-logo.png.png?ver=1704718347', 'rayobyte logo' ]
[ 'https://secure.gravatar.com/avatar/40de7ce27e119cfc9a04eef5e77cc6c2?s=96&r=g', 'Adam Dubois' ]
[ 'https://proxyway.com/wp-content/uploads/2022/05/bright-data-logo.png.png?ver=1704718964', 'Bright Data logo' ]
[ 'https://secure.gravatar.com/avatar/40de7ce27e119cfc9a04eef5e77cc6c2?s=96&r=g', 'Adam Dubois' ]
[ 'https://proxyway.com/wp-content/uploads/2020/07/oxylabs-logo.png.png?ver=1704718753', 'Oxylabs logo' ]
[ 'https://secure.gravatar.com/avatar/40de7ce27e119cfc9a04eef5e77cc6c2?s=96&r=g', 'Adam Dubois' ]
[ 'https://proxyway.com/wp-content/uploads/2022/05/bright-data-logo.png.png?ver=1704718964', 'Bright Data logo' ]
[ 'https://secure.gravatar.com/avatar/40de7ce27e119cfc9a04eef5e77cc6c2?s=96&r=g', 'Adam Dubois' ]
[ 'https://proxyway.com/wp-content/uploads/2022/03/zyte-logo.png.png?ver=1704718975', 'Zyte logo' ]
[ 'https://secure.gravatar.com/avatar/40de7ce27e119cfc9a04eef5e77cc6c2?s=96&r=g', 'Adam Dubois' ]
```

```
[ 'https://secure.gravatar.com/avatar/40de7ce27e119cfc9a04eef5e77cc6c2?s=96&r=g', 'Adam Dubois' ]
[ 'https://proxyway.com/wp-content/uploads/2022/07/rayobyte-logo.png.png?ver=1704718347', 'rayobyte logo' ]
[ 'https://secure.gravatar.com/avatar/40de7ce27e119cfc9a04eef5e77cc6c2?s=96&r=g', 'Adam Dubois' ]
[ 'https://proxyway.com/wp-content/uploads/2022/05/bright-data-logo.png.png?ver=1704718964', 'Bright Data logo' ]
[ 'https://secure.gravatar.com/avatar/40de7ce27e119cfc9a04eef5e77cc6c2?s=96&r=g', 'Adam Dubois' ]
[ 'https://proxyway.com/wp-content/uploads/2020/07/oxylabs-logo.png.png?ver=1704718753', 'Oxylabs logo' ]
[ 'https://secure.gravatar.com/avatar/40de7ce27e119cfc9a04eef5e77cc6c2?s=96&r=g', 'Adam Dubois' ]
[ 'https://proxyway.com/wp-content/uploads/2022/05/bright-data-logo.png.png?ver=1704718964', 'Bright Data logo' ]
[ 'https://secure.gravatar.com/avatar/40de7ce27e119cfc9a04eef5e77cc6c2?s=96&r=g', 'Adam Dubois' ]
[ 'https://proxyway.com/wp-content/uploads/2022/03/zyte-logo.png.png?ver=1704718975', 'Zyte logo' ]
[ 'https://secure.gravatar.com/avatar/40de7ce27e119cfc9a04eef5e77cc6c2?s=96&r=g', 'Adam Dubois' ]
[ 'https://proxyway.com/wp-content/uploads/2019/09/Smartproxy-Logo.png.png?ver=1704719397', 'smartproxy-logo' ]
[ 'https://secure.gravatar.com/avatar/40de7ce27e119cfc9a04eef5e77cc6c2?s=96&r=g', 'Adam Dubois' ]
[ 'https://proxyway.com/wp-content/uploads/2022/05/bright-data-logo.png.png?ver=1704718964', 'Bright Data logo' ]
[ 'https://proxyway.com/wp-content/uploads/2023/04/proxyway.svg?ver=1681290142', 'Proxyway' ]
```

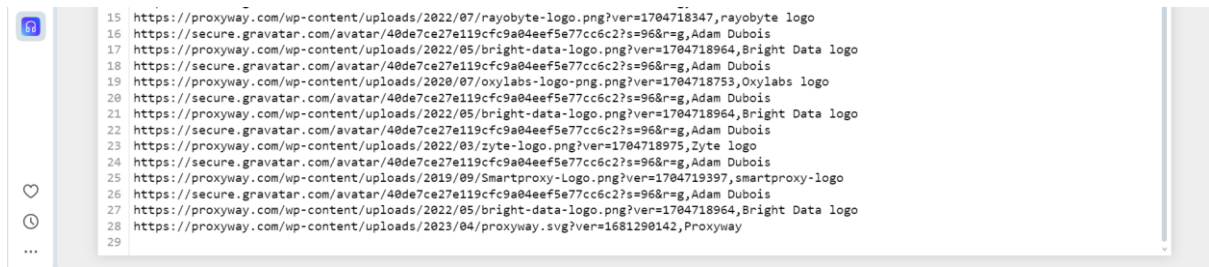
Dalam Bentuk CSV

UPDATE Read the migration plan to Notebook 7 to learn about the new features and the actions to take if you are using extensions - Please note that updating to Notebook 7 might break some of your extensions. Don't show anymore

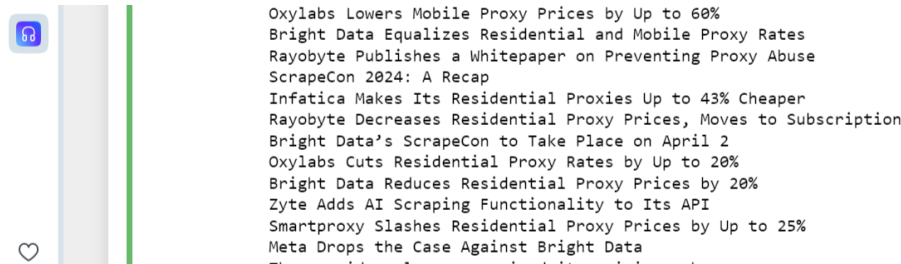
jupyter data_gambar.csv 32 minutes ago Logout

File Edit View Language current mode

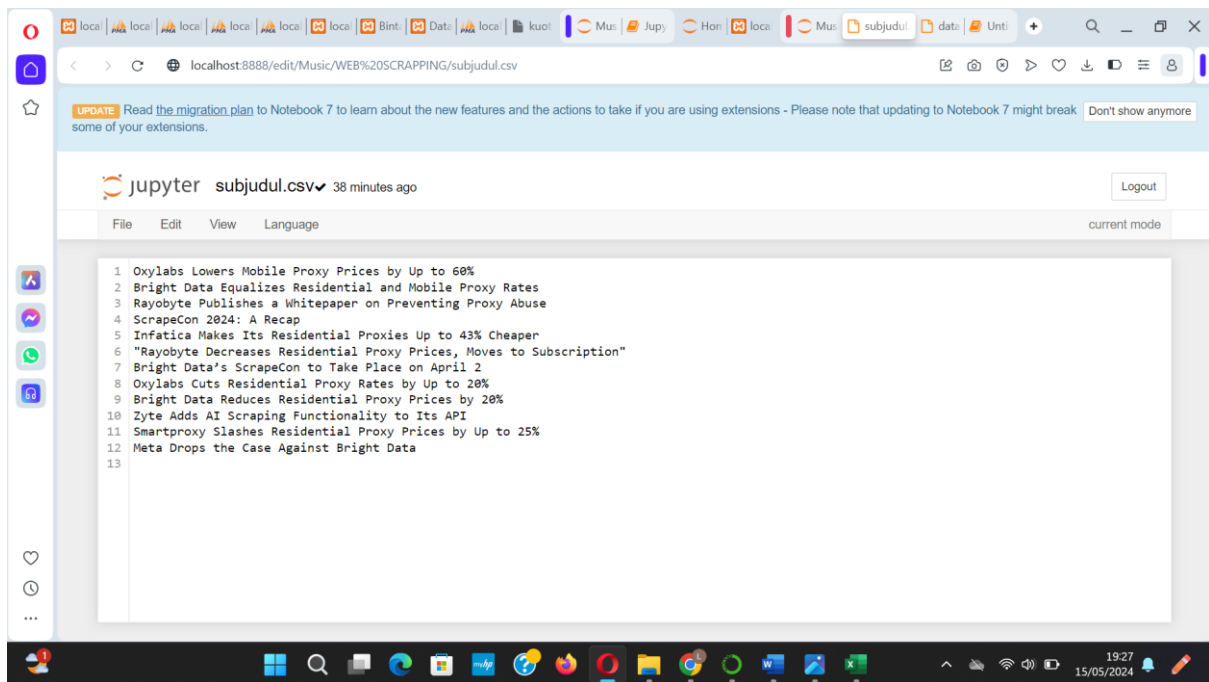
```
1 src,alt
2 https://proxyway.com/wp-content/uploads/2023/04/proxyway.svg?ver=1681290142,Proxyway
3 https://proxyway.com/wp-content/uploads/2023/04/Newsfeed-image-min.png,Adam sitting in a chair reading a newspaper
4 https://secure.gravatar.com/avatar/40de7ce27e119cfc9a04eef5e77cc6c2?s=96&r=g,Adam Dubois
5 https://proxyway.com/wp-content/uploads/2020/07/oxylabs-logo.png.png?ver=1704718753,Oxylabs logo
6 https://secure.gravatar.com/avatar/40de7ce27e119cfc9a04eef5e77cc6c2?s=96&r=g,Adam Dubois
7 https://proxyway.com/wp-content/uploads/2022/05/bright-data-logo.png.png?ver=1704718964,Bright Data logo
8 https://secure.gravatar.com/avatar/40de7ce27e119cfc9a04eef5e77cc6c2?s=96&r=g,Adam Dubois
9 https://proxyway.com/wp-content/uploads/2022/07/rayobyte-logo.png.png?ver=1704718347,rayobyte logo
10 https://secure.gravatar.com/avatar/40de7ce27e119cfc9a04eef5e77cc6c2?s=96&r=g,Adam Dubois
11 https://proxyway.com/wp-content/uploads/2022/05/bright-data-logo.png.png?ver=1704718964,Bright Data logo
12 https://secure.gravatar.com/avatar/40de7ce27e119cfc9a04eef5e77cc6c2?s=96&r=g,Adam Dubois
13 https://proxyway.com/wp-content/uploads/2020/08/infatica-logo.png.png?ver=1704718655,infatica logo
14 https://secure.gravatar.com/avatar/40de7ce27e119cfc9a04eef5e77cc6c2?s=96&r=g,Adam Dubois
15 https://proxyway.com/wp-content/uploads/2022/07/rayobyte-logo.png.png?ver=1704718347,rayobyte logo
16 https://secure.gravatar.com/avatar/40de7ce27e119cfc9a04eef5e77cc6c2?s=96&r=g,Adam Dubois
17 https://proxyway.com/wp-content/uploads/2022/05/bright-data-logo.png.png?ver=1704718964,Bright Data logo
18 https://secure.gravatar.com/avatar/40de7ce27e119cfc9a04eef5e77cc6c2?s=96&r=g,Adam Dubois
19 https://proxyway.com/wp-content/uploads/2020/07/oxylabs-logo.png.png?ver=1704718753,Oxylabs logo
20 https://secure.gravatar.com/avatar/40de7ce27e119cfc9a04eef5e77cc6c2?s=96&r=g,Adam Dubois
21 https://proxyway.com/wp-content/uploads/2022/05/bright-data-logo.png.png?ver=1704718964,Bright Data logo
22 https://secure.gravatar.com/avatar/40de7ce27e119cfc9a04eef5e77cc6c2?s=96&r=g,Adam Dubois
23 https://proxyway.com/wp-content/uploads/2022/03/zyte-logo.png.png?ver=1704718975,Zyte logo
```



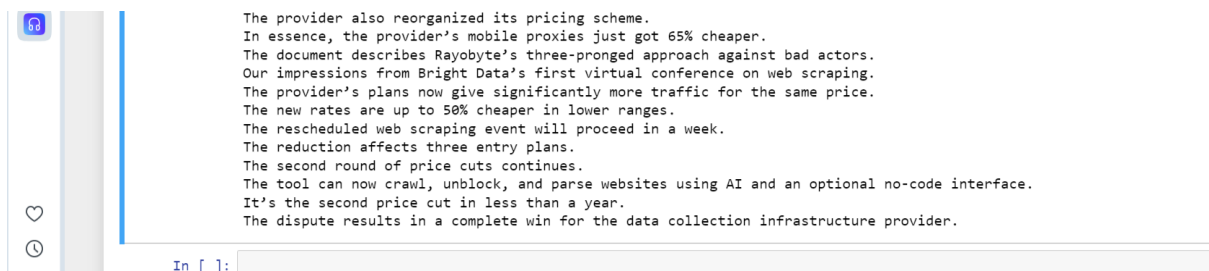
2. Data Subjudul



Dalam bentuk CSV



3. Data Keterangan



Dalam Bentuk CSV

local local local local local local local local local local local local Mu Jup Hor local Mu kete x sub dat Unt +

localhost:8888/edit/Music/WEB%20SCRAPPING/keterangan.csv

UPDATE Read the [migration plan](#) to Notebook 7 to learn about the new features and the actions to take if you are using extensions - Please note that updating to Notebook 7 might break some of your extensions. Don't show anymore

jupyter keterangan.csv 39 minutes ago Logout

File Edit View Language current mode

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

19:28 15/05/2024