

LAPORAN PRAKTIK PEMOGRAMAN PYTHON

SCRAPING



Disusun oleh :
Rifqy Rivaldi
(V3922040)

Dosen
Yusuf Fadila Rachman S.Kom,M.Kom

**PS D-III TEKNIK INFORMATIKA
SEKOLAH VOKASI
UNIVERSITAS SEBELAS MARET
2023**

SOAL

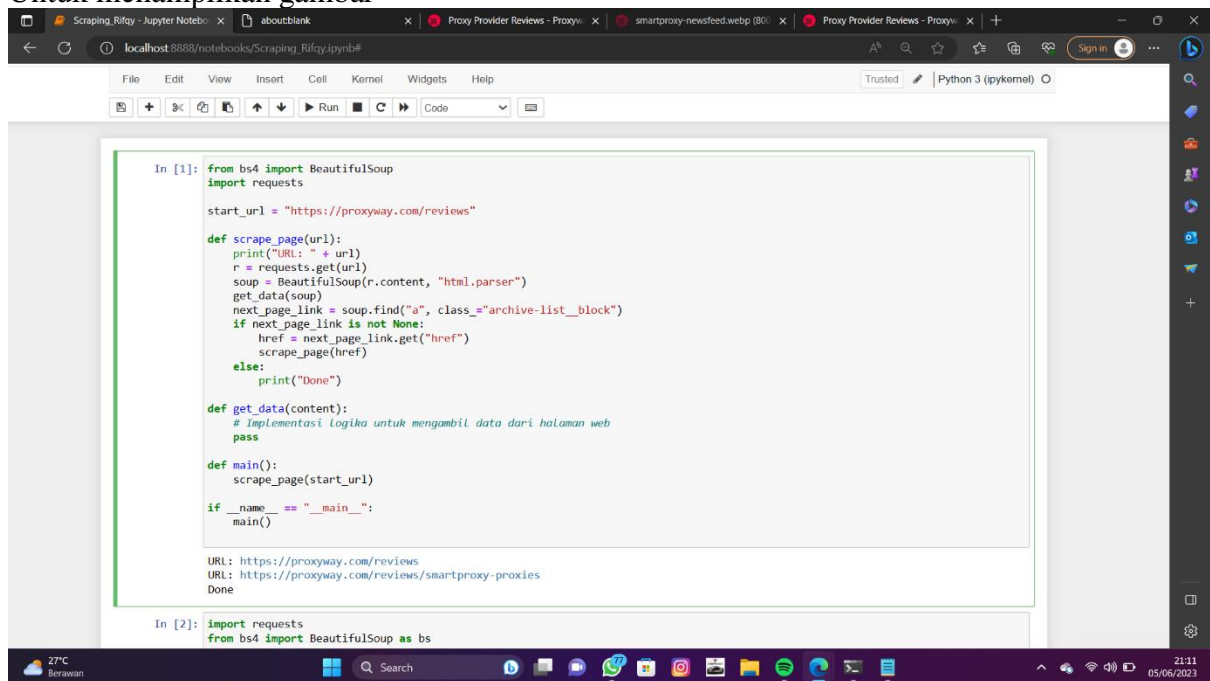
TUGAS LATIHAN SCRAPING WEBSITE

- Silahkan scraping data di website <https://proxyway.com/reviews> atau <https://proxyway.com/news>
- Ambil data judul/title di setiap halaman
- Simpan data ke dalam format CSV

JAWABAN

SCRIPT :

Untuk menampilkan gambar



```
In [1]: from bs4 import BeautifulSoup
import requests

start_url = "https://proxyway.com/reviews"

def scrape_page(url):
    print("URL: " + url)
    r = requests.get(url)
    soup = BeautifulSoup(r.content, "html.parser")
    get_data(soup)
    next_page_link = soup.find("a", class_="archive-list_block")
    if next_page_link is not None:
        href = next_page_link.get("href")
        scrape_page(href)
    else:
        print("Done")

def get_data(content):
    # Implementasi logika untuk mengambil data dari halaman web
    pass

def main():
    scrape_page(start_url)

if __name__ == "__main__":
    main()

URL: https://proxyway.com/reviews
URL: https://proxyway.com/reviews/smartproxy-proxies
Done

In [2]: import requests
from bs4 import BeautifulSoup as bs
```

Untuk Menampilkan Perhalaman dan hasilnya

The image shows a Jupyter Notebook interface with a Python script for web scraping and its execution output displayed in a browser window.

Jupyter Notebook Code:

```
In [7]: from bs4 import BeautifulSoup
import requests

start_url = "https://proxypool.com/reviews"

def scrape_page(url):
    print('URL: ' + url)
    r = requests.get(url)
    soup = BeautifulSoup(r.content, "html.parser")
    get_data(soup)
    next_page_link = soup.find("a", class_="next")
    if next_page_link is not None:
        href = next_page_link.get("href")
        scrape_page(href)
    else:
        print("Done")

def get_data(content):
    # Implementasi logika untuk mengambil data dari halaman web
    pass

def main():
    scrape_page(start_url)

if __name__ == "__main__":
    main()

import csv
import requests
from bs4 import BeautifulSoup as bs

URL = "https://proxypool.com/reviews"

data = []

for page in range(1, 4):
    print("ur")
    print("Sub Titles Page:", page, "ur")

    req = requests.get(URL + "/page/" + str(page))
    soup = bs(req.text, "html.parser")

    titles = soup.find_all("h3", class_="archive-list_title")

    for i, title in enumerate(titles):
        print(f'{i+1} {title.text}')
        data.append({
            'Page Number': f'Page {page}',
            'Title Number': f'Title {i+1}',
            'Title Name': title.text
        })

# Menyimpan data ke dalam file CSV
filename = "proxypooldata.csv"
fieldnames = ['Page Number', 'Title Number', 'Title Name']

with open(filename, 'w', newline='') as csvfile:
    writer = csv.DictWriter(csvfile, fieldnames=fieldnames)
    writer.writeheader()
    writer.writerows(data)

print("Data telah disimpan ke dalam file, Terima Kasih Untuk Semuanya", filename)

URL: https://proxypool.com/reviews
URL: https://proxypool.com/reviews/page/2
URL: https://proxypool.com/reviews/page/3
Done
```

Browser Output:

```
titname = proxypooldata.csv
fieldnames = ['Page Number', 'Title Number', 'Title Name']

with open(filename, 'w', newline='') as csvfile:
    writer = csv.DictWriter(csvfile, fieldnames=fieldnames)
    writer.writeheader()
    writer.writerows(data)

print("Data telah disimpan ke dalam file, Terima Kasih Untuk Semuanya", filename)

URL: https://proxypool.com/reviews
URL: https://proxypool.com/reviews/page/2
URL: https://proxypool.com/reviews/page/3
Done

Sub Titles Page: 1

1 Smartproxy
2 Oxylabs
3 Bright Data
4 SOAX
5 NetNut
6 Proxydrop
7 Rayobyte
8 Geosurf
9 Jyze (Crawlera)
10 IPRoyal
11 PacketStream
12 Shifter

Sub Titles Page: 2

1 Multilogin
2 Gologin
3 Infatica
4 Storm Proxies
5 SERPMaster
6 AstroProxy
7 InstantProxies
8 Rocks
9 Ioad Out Proxies
10 Brazy Kicks
11 The Proxy store
12 LineProxies

Sub Titles Page: 3

1 HighProxies
2 MyPrivateProxy
Data telah disimpan ke dalam file, Terima Kasih Untuk Semuanya proxypooldata.csv
```

The browser window shows the output of the script, including the URL being scraped, the sub-titles for each page, and the final message indicating that the data has been saved to the file.