

LAPORAN PRAKTIKUM
PRAKTIK PEMROGRAMAN PYTHON

PRAKTIKUM 8
WEB SCRAPPING DENGAN PYTHON



Disusun oleh :

Muhammad Farhan

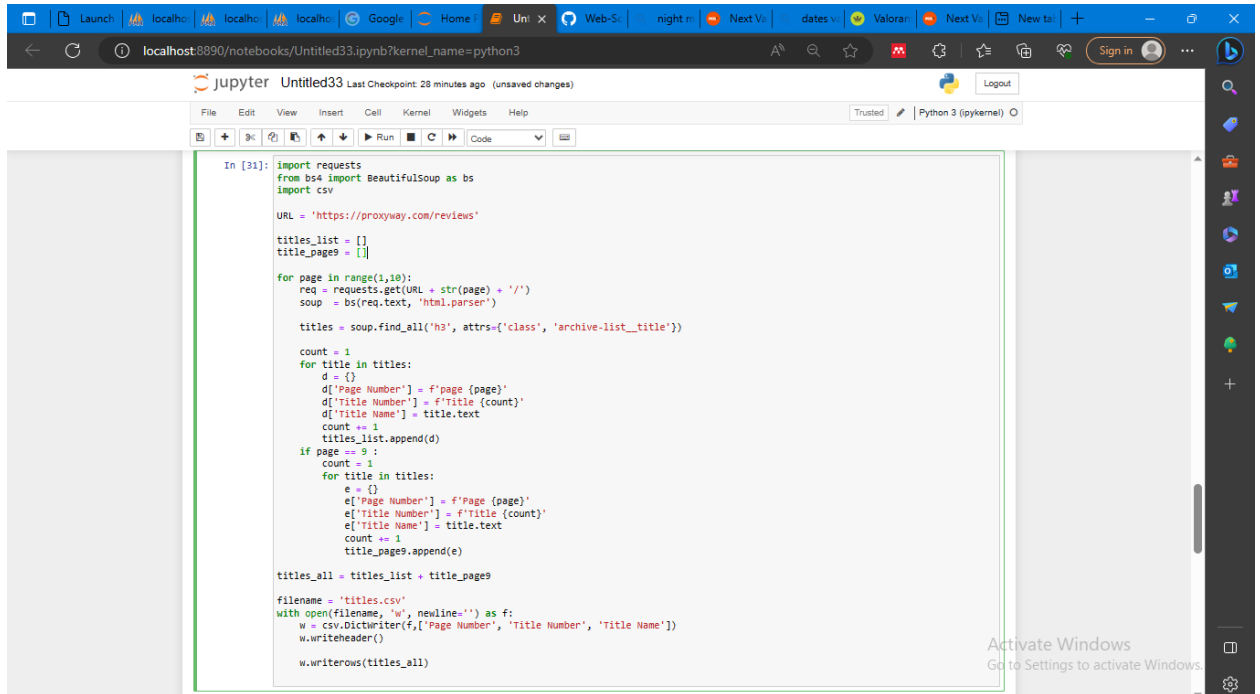
V3922033

Dosen

Yusuf Fadhillah S.T. M.T.

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

1. Mengambil data dari Web ke data Python



The screenshot shows a Jupyter Notebook running in a web browser at localhost:8890. The notebook is titled 'Untitled33' and has a Python 3 kernel. The code in the cell is as follows:

```
In [31]: import requests
from bs4 import BeautifulSoup as bs
import csv

URL = 'https://proxyway.com/reviews'

titles_list = []
title_page9 = []

for page in range(1,10):
    req = requests.get(URL + str(page) + '/')
    soup = bs(req.text, 'html.parser')

    titles = soup.find_all('h3', attrs={'class', 'archive-list__title'})

    count = 1
    for title in titles:
        d = {}
        d['Page Number'] = f'page {page}'
        d['Title Number'] = f'Title {count}'
        d['Title Name'] = title.text
        count += 1
        titles_list.append(d)

    if page == 9 :
        count = 1
        for title in titles:
            e = {}
            e['Page Number'] = f'Page {page}'
            e['Title Number'] = f'Title {count}'
            e['Title Name'] = title.text
            count += 1
            title_page9.append(e)

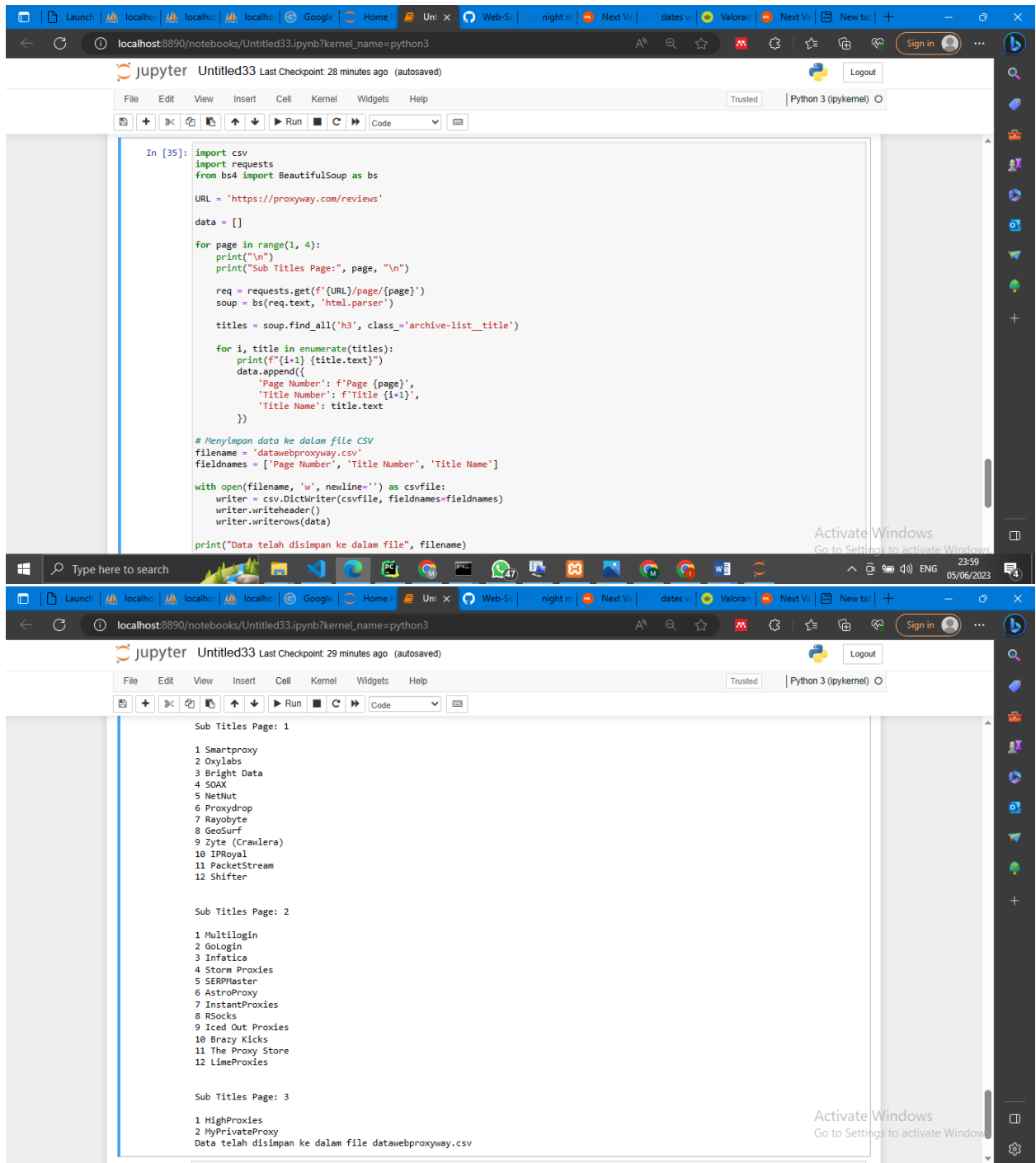
titles_all = titles_list + title_page9

filename = 'titles.csv'
with open(filename, 'w', newline='') as f:
    w = csv.DictWriter(f,['Page Number', 'Title Number', 'Title Name'])
    w.writeheader()

    w.writerows(titles_all)
```

An 'Activate Windows' watermark is visible in the bottom right corner of the notebook interface.

2. Memasukkan data python ke dalam csv



The image displays two screenshots of a Jupyter Notebook interface, showing the execution of a Python script that scrapes proxy data from a website and saves it to a CSV file.

Top Screenshot: The notebook is titled "Untitled33" and shows the code for scraping proxy data. The code imports `csv`, `requests`, and `BeautifulSoup`. It defines a URL and iterates through pages 1 to 4. For each page, it sends a GET request, parses the HTML, finds all `h3` tags with the class `archive-list_title`, and appends the page number, title number, and title text to a `data` list. Finally, it writes the data to a CSV file named `datawebproxyway.csv`.

```
In [35]: import csv
import requests
from bs4 import BeautifulSoup as bs

URL = 'https://proxyway.com/reviews'

data = []

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

    req = requests.get(f'{URL}/page/{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 = 'datawebproxyway.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", filename)
```

Bottom Screenshot: The notebook shows the output of the script. It displays the sub-titles for pages 1, 2, and 3, listing various proxy providers. The output for page 1 is:

```
Sub Titles Page: 1
1 Smartproxy
2 Oxylabs
3 Bright Data
4 SOAX
5 NetNut
6 Proxydrop
7 Raybyte
8 GeoSurf
9 Zyte (Crawlera)
10 IPRoyal
11 PacketStream
12 Shifter
```

The output for page 2 is:

```
Sub Titles Page: 2
1 Multilogin
2 GoLogin
3 Infatica
4 Storm Proxies
5 SERPMaster
6 AstroProxy
7 InstantProxies
8 RSocks
9 Iced Out Proxies
10 Brazy Kicks
11 The Proxy Store
12 LimeProxies
```

The output for page 3 is:

```
Sub Titles Page: 3
1 HighProxies
2 MyPrivateProxy
```

The final line of the output indicates that the data has been saved to the file `datawebproxyway.csv`.

```
Data telah disimpan ke dalam file datawebproxyway.csv
```

HASIL CSV

