Web-Scraping-Lab

April 17, 2023

1 Hands-on Lab : Web Scraping

Estimated time needed: **30 to 45** minutes

1.1 Objectives

In this lab you will perform the following:

- Extract information from a given web site
- Write the scraped data into a csv file.

1.2 Extract information from the given web site

You will extract the data from the below web site:

[1]:
```
#this url contains the data you need to scrape
url = "https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/
↪IBM-DA0321EN-SkillsNetwork/labs/datasets/Programming_Languages.html"
```

The data you need to scrape is the **name of the programming language** and **average annual salary**. It is a good idea to open the url in your web broswer and study the contents of the web page before you start to scrape.

Import the required libraries

[2]:
```
# Your code here
from bs4 import BeautifulSoup as bs
import requests
```

Download the webpage at the url

[3]:
```
#your code goes here
data = requests.get(url).text
```

Create a soup object

[7]:
```
#your code goes here
soup = bs(data)
```

Scrape the `Language name` and `annual average salary`.

[8]:
```
#your code goes here
table = soup.find("table")
popular_language =[]

for row in table.find_all('tr'):
    cols= row.find_all('td')
    language= cols[1].getText()
    salary = cols[3].getText()
    print(("{}--->{}").format(language,salary))
    popular_lan=[language,salary]
    popular_language.append(popular_lan)

print(popular_language)
```

```
Language--->Average Annual Salary
Python--->$114,383
Java--->$101,013
R--->$92,037
Javascript--->$110,981
Swift--->$130,801
C++--->$113,865
C#--->$88,726
PHP--->$84,727
SQL--->$84,793
Go--->$94,082
[['Language', 'Average Annual Salary'], ['Python', '$114,383'], ['Java',
'$101,013'], ['R', '$92,037'], ['Javascript', '$110,981'], ['Swift',
'$130,801'], ['C++', '$113,865'], ['C#', '$88,726'], ['PHP', '$84,727'], ['SQL',
'$84,793'], ['Go', '$94,082']]
```

Save the scrapped data into a file named *popular-languages.csv*

[ ]:
```
# your code goes here
import csv
with open('popular-languages.csv','w',newline='') as file:
    csvwriter = csv.writer(file)
    for row in popular_language:
        csvwriter.writerow(row)

import pandas as pd
df = pd.read_csv('popular-languages.csv')
df.head(50)

import matplotlib.pyplot as plt
import numpy as np

perfomance = [130801, 114383,113865,110981,101013,94082,92037,88726,84727,84793]
dfs = ['Swift','Python','JAVA','C++','JavaScript','GO','R','C#','SQL','PHP']
```

```
tks= np.arange(len(dfs))
plt.figure(figsize=(16,7))
plt.bar(tks, perfomance, align='center',alpha=0.5)
plt.xticks(tks,dfs)
plt.title('Programming Language')
plt.show()
```

1.3 Authors

Ramesh Sannareddy

1.3.1 Other Contributors

Rav Ahuja

1.4 Change Log

Date (YYYY-MM-DD)	Version	Changed By	Change Description
2020-10-17	0.1	Ramesh Sannareddy	Created initial version of the lab