SOCIAL MEDIA – WEBSCRAPING USING PYTHON

Assignment report submitted to

Graduate School at

The University of Arkansas at Little Rock

in partial fulfillment of

requirements for the graduate course of

202510 Spring: Data Science - Technologies (9U1)

April 2025

by

Naga Venkata Ramkiran Hota

Instructors

Dr. Elizabeth Pierce

and

Dr. Serhan Dagtas

SOCIAL MEDIA – WEBSCRAPING USING PYTHON

ABSTRACT

This assignment consists of implementing a specific technique listed in *Geeks for Geeks* tutorial to scrape a publicly available Web URL using python libraries. Check “List of References” section for the tutorial link.

For this report, the technique chosen to scrape the website using *BeautifulSoup* Library. This report also publishes the inputs and outputs (results).

The website which is used as source to be scraped is a Microsoft Publicly available *MSDN Assessments* and check “List of References” section for the link.

To help complete the assignment, Google Colab environment is used along with *ast, requests* and *bs4* modules.

SOCIAL MEDIA – WEBSCRAPING USING PYTHON

IMPLEMENTATIONT WITH INPUTS AND OUTPUTS

Main Method:

```
from ast import parse
import requests
from bs4 import BeautifulSoup

# establish the URL and call the URL using requests lib
microsoftLearnUrl='https://learn.microsoft.com/en-
us/assessments/browse/?page=1&pagesize=30'
responseData = requests.get(microsoftLearnUrl)

# parse the response html
parsedText = BeautifulSoup(responseData.text,'html.parser')
#print(parsedText)

# find a specific element using find_all function
titles = parsedText.find_all('a', class_='card-title margin-none padding-
top-xxs')

# loop thru the titles and format it with numbers.
index=0
outputText = ''
for title in titles:
    index+=1
    outputText += str(index) +') '+ title.string.strip() +'.\n'

# print the total titles scraped and the output text
print('Total titles scraped: ' + str(index) +'\n')
print(outputText)
```

In the above code snippet, we define the source url on which we try to fetch the response by making a *get* request. Then the python code scrapes the Microsoft site for assessment titles and prints them in a numbered list.

SOCIAL MEDIA – WEBSCRAPING USING PYTHON

The main method to scrape are *BeautifulSoup* which parses the HTML content of the response using *html.parser.* The next important piece of scraping is to find the useful information like the title of the assessments available on this site. The searching of this information is achieved by using *find_all* method which searched for all '*a*' hyperlink tags.

Output:

```
Total titles scraped: 21

    1) Azure Virtual Desktop | Microsoft Partner.
    2) Microsoft Cloud for Retail Adoption Guide | Microsoft Partners.
    3) Microsoft Cloud for Sustainability Adoption Guide | Microsoft Partners.
    4) Power Platform Solution Assessment (preview).
    5) AI Engineer Skill Assessment.
    6) AI Readiness Assessment.
    7) Analytics | Well-Architected Review.
    8) Analytics Journey Tracker.
    9) App and Data Modernization Readiness Tool.
    10) Azure Cognitive Search | Well-Architected Review.
    11) Azure Landing Zone Review.
    12) Azure Machine Learning.
    13) Azure Stack HCI | Microsoft Partners.
    14) Azure VMware Solution (AVS) | Microsoft Partner.
    15) Azure VMware Solution Landing Zone Assessment Review.
    16) Azure Well-Architected AI workload.
    17) Azure Well-Architected Azure Virtual Desktop workload.
    18) Azure Well-Architected Azure VMware Solution workload.
    19) Azure Well-Architected Oracle on Azure IaaS workload.
    20) Azure Well-Architected Review.
    21) Azure Well-Architected SaaS workload.
```

SOCIAL MEDIA – WEBSCRAPING USING PYTHON

LIST OF REFERENCES

Microsoft Assessments
https://learn.microsoft.com/en-us/assessments/browse/?page=1&pagesize=30

Geeks for Geeks Python Web Scraping Tutorial
https://www.geeksforgeeks.org/python-web-scraping-tutorial/