**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

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.

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.

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:

A screenshot of a computer program

Description automatically generated

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/>