

Social Media - Web Scraping Using Python

Assignment 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)

in

March 2025

by

Deepak Singla

Instructors

Dr. Elizabeth Pierce

and

Dr. Serhan Dagtas

Web Scraping Using Python

Assignment Instructions

For this assignment, check out this Python Web Scraping Tutorial - <https://www.geeksforgeeks.org/python-web-scraping-tutorial/>. Try implementing one of the techniques and submit your output (along with a brief description of the technique you used to do some simple web scraping). Submit your work to the Assignment Folder.

Web Scrapping Using Python

ABSTRACT

For the purpose of the assignment, I have chosen <https://cognitiveclass.ai/courses> page.

Following tools and libraries have been used to demonstrate the process of webscrapping from a public page.

- <https://colab.research.google.com/> as the platform to run the code
- Python
- Request Module from Python
- BeautifulSoup Module from Python

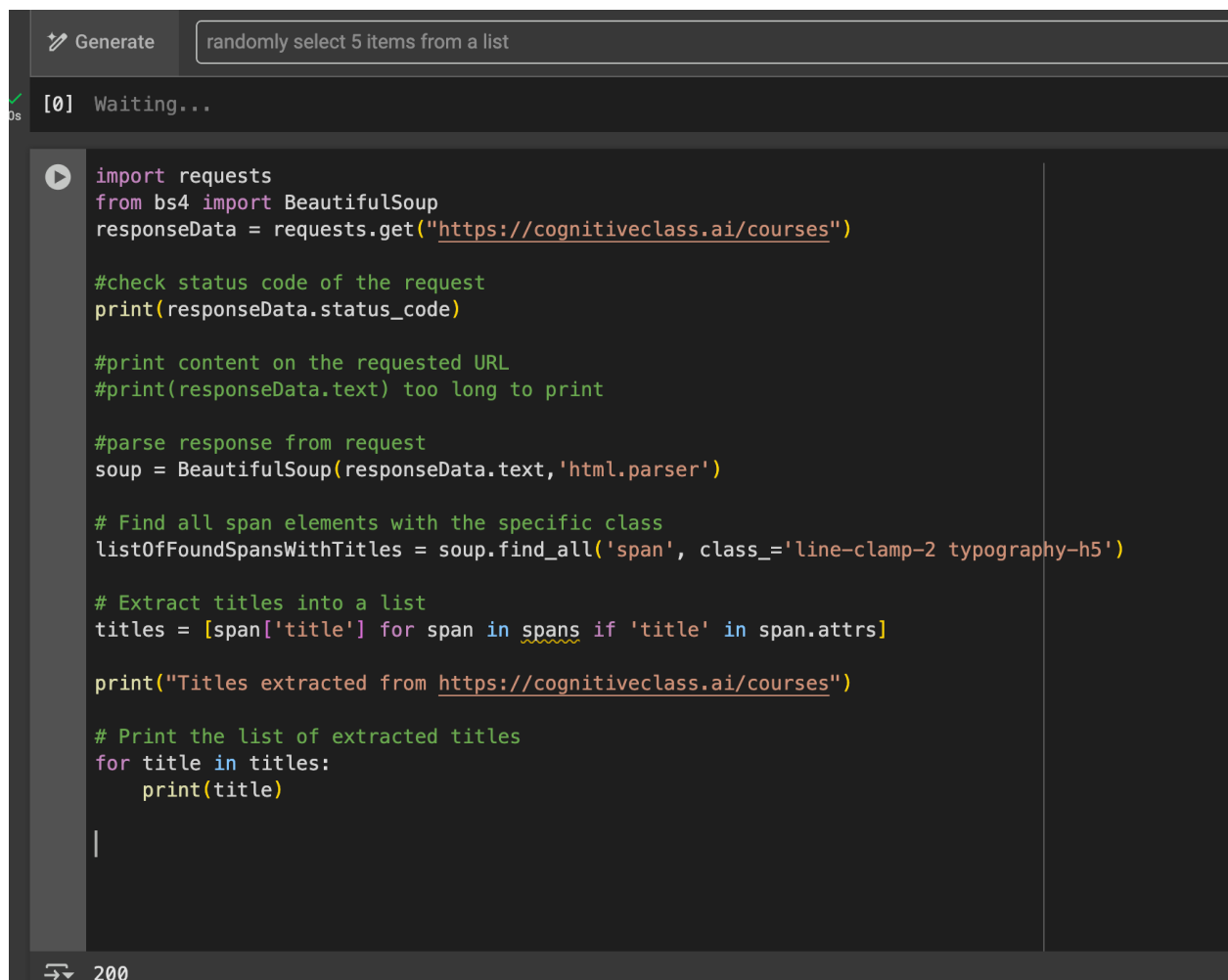
Web Scrapping Using Python

Steps followed

- Send a request using URL
- Parse the response from the requested URL
- Find the content web scrapping is done from parsed data by searching html document on tags.

Code Snippet

Following code snippet has been used to extract titles from the requested URL.

A screenshot of a code editor interface. At the top, there is a 'Generate' button and a text input field containing 'randomly select 5 items from a list'. Below this, a status bar shows '[0] Waiting...'. The main area contains a Python script with the following code:

```
import requests
from bs4 import BeautifulSoup
responseData = requests.get("https://cognitiveclass.ai/courses")

#check status code of the request
print(responseData.status_code)

#print content on the requested URL
#print(responseData.text) too long to print

#parse response from request
soup = BeautifulSoup(responseData.text, 'html.parser')

# Find all span elements with the specific class
listOfFoundSpansWithTitles = soup.find_all('span', class_='line-clamp-2 typography-h5')

# Extract titles into a list
titles = [span['title'] for span in spans if 'title' in span.attrs]

print("Titles extracted from https://cognitiveclass.ai/courses")

# Print the list of extracted titles
for title in titles:
    print(title)
```

The script is written in a dark-themed editor with syntax highlighting. At the bottom left, there is a small icon and the number '200'.

Using *requests* module, a get request is sent to the requested URL and response is saved in a variable *responseData*.

Using *BeautifulSoup* library html in *responseData* is parsed. Then the parsed is searched using method *find_all* on html tags and class name.

Web Scrapping Using Python

Output

Following is the final output of the code above.

```
➤ Titles extracted from https://cognitiveclass.ai/courses  
Python for Data Science  
SQL and Relational Databases 101  
Data Analysis with Python  
Prompt Engineering for Everyone  
Introduction to Cloud  
Data Science 101  
Big Data 101  
Docker Essentials: A Developer Introduction  
Data Visualization with Python  
Data Science Methodology  
Introducing AI  
Python for Data Science, AI & Development  
IBM Cloud Essentials  
Deep Learning Fundamentals  
Data Science Tools  
Introduction to Containers, Kubernetes, and OpenShift  
Build Your Own Chatbot  
Building AI Powered Chatbots Without Programming  
Introduction to DevOps  
Compare DeepSeek-R1 vs OpenAI o3-mini on Data Science tasks  
Hadoop 101  
Data Visualization with R  
R for Data Science  
Build an IoT Blockchain Network for a Supply Chain  
Spark Fundamentals I  
A Quick Introduction to Machine Learning  
Deep Learning with TensorFlow  
AI Ethics  
AI Concepts  
Agentic AI: Build a Multi-Agent App with CrewAI & Gradio
```

Web Scrapping Using Python

Credits

Report format has been copied from

<https://www.proquest.com/pqdtlocal1009983/docview/2902800999/fulltextPDF/C967C462E7DF4DF2PQ/1?%20Theses&accountid=14482&sourcetype=Dissertations%20>

Other References

<https://cognitiveclass.ai/courses>

<https://www.geeksforgeeks.org/python-web-scrapping-tutorial/#requests-module>