Image ScrapperAssignment Questions

Assignment

Go to this given URL and solve the following questions

URL: https://www.youtube.com/@PW-Foundation/videos

Q1. Write a python program to extract the video URL of the first five videos.

Q2. Write a python program to extract the URL of the video thumbnails of the first five videos. Q3. Write a python program to extract the title of the first five videos.

Q4. Write a python program to extract the number of views of the first five videos. Q5. Write a python program to extract the time of posting of video for the first five videos. Note: Save all the data scraped in the above questions in a CSV file.

Create a simple UI with all functionalities mentioned above and deploy it in AWS.

Note:  Create your assignment in Jupyter notebook and upload it to GitHub & share that uploaded assignment file link through your dashboard. Make sure the repository is public. Submit your deployment link as well.

1.

Code:

import requests

from bs4 import BeautifulSoup

def extract\_video\_urls(channel\_url, num\_videos):

try:

response = requests.get(channel\_url)

response.raise\_for\_status()

soup = BeautifulSoup(response.content, 'html.parser')

video\_urls = []

video\_elements = soup.select('a.yt-simple-endpoint.style-scope.ytd-grid-video-renderer')

for i in range(min(num\_videos, len(video\_elements))):

video\_url = 'https://www.youtube.com' + video\_elements[i]['href']

video\_urls.append(video\_url)

return video\_urls

except requests.exceptions.RequestException as e:

print(f"Error occurred: {e}")

# Example usage

channel\_url = 'https://www.youtube.com/@PW-Foundation/videos'

num\_videos = 5

video\_urls = extract\_video\_urls(channel\_url, num\_videos)

if video\_urls:

print(f"The first {num\_videos} video URLs are:")

for url in video\_urls:

print(url)

2.

Code:

import requests

from bs4 import BeautifulSoup

def extract\_thumbnail\_urls(channel\_url, num\_videos):

try:

response = requests.get(channel\_url)

response.raise\_for\_status()

soup = BeautifulSoup(response.content, 'html.parser')

thumbnail\_urls = []

thumbnail\_elements = soup.select('a#thumbnail')

for i in range(min(num\_videos, len(thumbnail\_elements))):

thumbnail\_url = thumbnail\_elements[i].find('img')['src']

thumbnail\_urls.append(thumbnail\_url)

return thumbnail\_urls

except requests.exceptions.RequestException as e:

print(f"Error occurred: {e}")

# Example usage

channel\_url = 'https://www.youtube.com/@PW-Foundation/videos'

num\_videos = 5

thumbnail\_urls = extract\_thumbnail\_urls(channel\_url, num\_videos)

if thumbnail\_urls:

print(f"The thumbnail URLs of the first {num\_videos} videos are:")

for url in thumbnail\_urls:

print(url)

3.

Code:

import requests

from bs4 import BeautifulSoup

def extract\_video\_titles(channel\_url, num\_videos):

try:

response = requests.get(channel\_url)

response.raise\_for\_status()

soup = BeautifulSoup(response.content, 'html.parser')

video\_titles = []

title\_elements = soup.select('a#video-title')

for i in range(min(num\_videos, len(title\_elements))):

video\_title = title\_elements[i].text.strip()

video\_titles.append(video\_title)

return video\_titles

except requests.exceptions.RequestException as e:

print(f"Error occurred: {e}")

# Example usage

channel\_url = 'https://www.youtube.com/@PW-Foundation/videos'

num\_videos = 5

video\_titles = extract\_video\_titles(channel\_url, num\_videos)

if video\_titles:

print(f"The titles of the first {num\_videos} videos are:")

for title in video\_titles:

print(title)

4.

Code:

import requests

from bs4 import BeautifulSoup

def extract\_view\_counts(channel\_url, num\_videos):

try:

response = requests.get(channel\_url)

response.raise\_for\_status()

soup = BeautifulSoup(response.content, 'html.parser')

view\_counts = []

count\_elements = soup.select('span.style-scope.ytd-grid-video-renderer')

for i in range(min(num\_videos, len(count\_elements))):

view\_count = count\_elements[i].text.strip()

view\_counts.append(view\_count)

return view\_counts

except requests.exceptions.RequestException as e:

print(f"Error occurred: {e}")

# Example usage

channel\_url = 'https://www.youtube.com/@PW-Foundation/videos'

num\_videos = 5

view\_counts = extract\_view\_counts(channel\_url, num\_videos)

if view\_counts:

print(f"The number of views for the first {num\_videos} videos are:")

for count in view\_counts:

print(count)

5.

Code:

import requests

from bs4 import BeautifulSoup

def extract\_post\_times(channel\_url, num\_videos):

try:

response = requests.get(channel\_url)

response.raise\_for\_status()

soup = BeautifulSoup(response.content, 'html.parser')

post\_times = []

time\_elements = soup.select('a.yt-simple-endpoint.style-scope.ytd-grid-video-renderer span.style-scope.ytd-thumbnail-overlay-time-status-renderer')

for i in range(min(num\_videos, len(time\_elements))):

post\_time = time\_elements[i].text.strip()

post\_times.append(post\_time)

return post\_times

except requests.exceptions.RequestException as e:

print(f"Error occurred: {e}")

# Example usage

channel\_url = 'https://www.youtube.com/@PW-Foundation/videos'

num\_videos = 5

post\_times = extract\_post\_times(channel\_url, num\_videos)

if post\_times:

print(f"The posting times for the first {num\_videos} videos are:")

for time in post\_times:

print(time)