DS-Assignment 1

> Data Science Assignment-1

> Instructor : Shahzeb Khan

Assignment-1: Web Scraping Tasks on Fast PWR Website.

Objective:
In this assignment, you will perform various web scraping tasks on your Fast pwr website-
http://pwr.nu.edu.pk/ using Python. These tasks will help you understand how to navigate and extract specific information from webpages.

Instructions:

1. **Ask for the website URL**: The user (you) will be prompted to input the URL of your university's website.
2. **Tasks**: The assignment consists of multiple short tasks where you will extract different types of data from the webpage.

Task 1: Print all `h3` tags and their content

- Extract and print all the `h3` tags on the page, displaying the content inside them.

Task 2: Extract heading tags (`h1` to `h6`), title, and meta description

- Print all heading tags (`h1` to `h6`), the page's title, and the meta description.

Task 3: Extract all `ALT` attributes

- Extract and print all `ALT` attributes of images on the webpage.

Task 4: Count the words on the webpage

- Count the number of words in the body of the webpage.

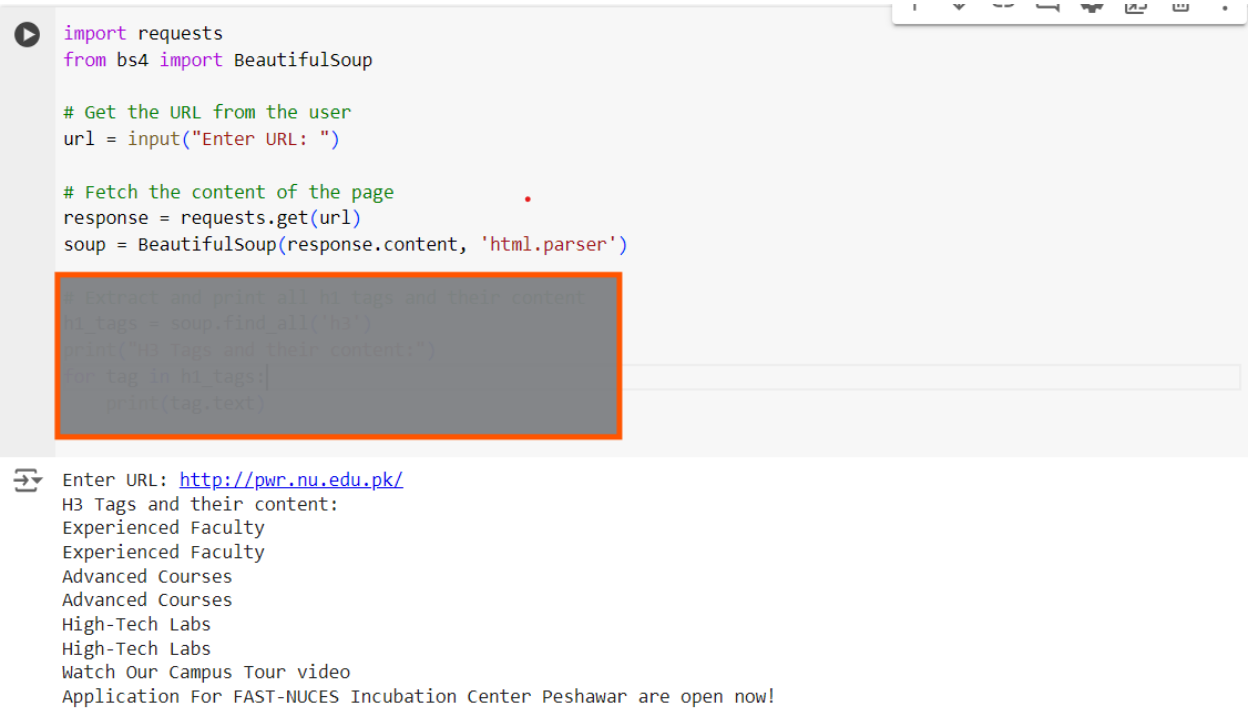
Task 5: Inspect for broken links

- Identify all links on the page and check if any of them are broken (404 errors).

Task 6: Count the number of images

[redacted]

Sample Output Task 1 :



```
import requests
from bs4 import BeautifulSoup

# Get the URL from the user
url = input("Enter URL: ")

# Fetch the content of the page
response = requests.get(url)
soup = BeautifulSoup(response.content, 'html.parser')

# Extract and print all h3 tags and their content
h3_tags = soup.find_all('h3')
print("H3 Tags and their content:")
for tag in h3_tags:
    print(tag.text)
```

```
Enter URL: http://pwr.nu.edu.pk/
H3 Tags and their content:
Experienced Faculty
Experienced Faculty
Advanced Courses
Advanced Courses
High-Tech Labs
High-Tech Labs
Watch Our Campus Tour video
Application For FAST-NUCES Incubation Center Peshawar are open now!
```

Submission:

Submit the following:

- .py or .ipynb file.
- pdf of python notebook
- CSV files if required for any task.

Ensure your solutions are clear and handle edge cases appropriately.