

ASSIGNMENT – WEB SCRAPING

Aim

Implement the basics of HTML tags using BeautifulSoup library and use any html code from the Practical of Web Tech subject.

Implementation

File: scrap.html

```
<!DOCTYPE html>
```

```
<html>
```

```
<head>
```

```
<title>
```

```
Web_Scraping_Example
```

```
</title>
```

```
</head>
```

```
<body>
```

```
<h1>
```

```
Welcome to my Webpage!
```

```
</h1>
```

```
<p>
```

```
My name is Harsh Shah!
```

```
</p>
```

```
<p>
```

```
Roll no.- 21BCP359, Div-6 G12
```

```
</p>
```

```
<p>
```

```
Implementing Web Scraping using BeautifulSoup in Python!
```

```
</p>
```

```
<ol>
```

```
<li>
```

```
List-1
```

```
</li>
```

```
<li>
```

```
List-2
```

```
</li>
```

```
<li>
```

```
List-3
```

```
</li>
```

```
<li>
```

```
List-4
```

```
</li>
```

```
</ol>
```

```
<div>
```

Now creating a division element in the webpage.

```
</div>
```

```
</body>
```

```
</html>
```

File: scrap.html

```
from bs4 import BeautifulSoup
```

```
with open("UNIT 3\Assignment\scrap.html", "r") as htmlfile:
```

```
    content = htmlfile.read()
```

```
soup = BeautifulSoup(content, "html.parser")
```

```
title = soup.title
```

```
print("Title: ", title.text)
```

```
heading = soup.h1
```

```
print("Heading: ", heading.text)
```

```
paragraphs = soup.find_all("p")
```

```
print("Paragraphs: ")
for paragraph in paragraphs:
    print(paragraph.text)
```

```
lists = soup.find_all("li")
print("Lists: ")
for list in lists:
    print(list.text)
```

```
print()
print("New HTML File")
print()
```

```
new_list = soup.new_tag("li")
new_list.string = "List-4"
ol = soup.ol
ol.append(new_list)
```

```
lists = soup.find_all("li")
print("Lists: ")
for list in lists:
    print(list.text)
```

```
div = soup.new_tag("div")
div.string = "Now creating a division element in the webpage."
body = soup.body
body.append(div)
print(div.text)
```

```
with open(r"UNIT 3\Assignment\scrap.html", "w") as htmlfile:
    htmlfile.write(soup.prettify())
```

File: new_scrap.html

```
<!DOCTYPE html>

<html>

<head>

<title>
  Web_Scraping_Example
</title>
</head>
<body>

<h1>
  Welcome to my Webpage!
</h1>

<p>
  My name is Adit Shah!
</p>

<p>
  Roll no.- 21BCP419, Div-6 G12
</p>

<p>
  Implementing Web Scraping using BeautifulSoup in Python!
</p>

<ol>

<li>
  List-1
</li>

<li>
  List-2
</li>

<li>
  List-3
</li>
```

```
<li>
```

```
List-4
```

```
</li>
```

```
</ol>
```

```
<div>
```

Now creating a division element in the webpage.

```
</div>
```

```
</body>
```

```
</html>
```

Output

```
Title: Web_Scraping_Example
Heading: Welcome to my Webpage!
Paragraphs:
My name is Adit Shah!
Roll no.- 21BCP419, Div-6 G12
Implementing Web Scraping using BeautifulSoup in Python!
Lists:
List-1
List-2
List-3

New HTML File

Lists:
List-1
List-2
List-3
List-4
Now creating a division element in the webpage.
```