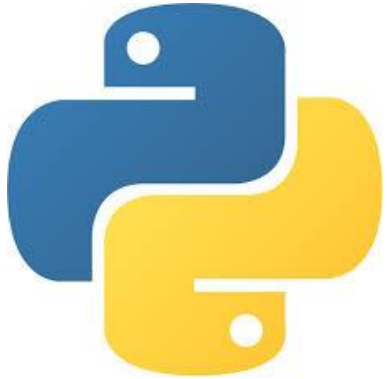
Python –
HTML and BeautifulSoup



Shahriar's Intelligence Academy

HTML

Definition:

- HTML stands for HyperText Markup Language.
- It is the standard language for creating web pages.
- HTML structures web content and defines elements like headings, paragraphs, links, and tables.

Key Features:

- Easy to learn and use.
- Compatible with all browsers.
- Allows embedding multimedia like images, audio, and video.

HTML Tags

Definition:

- Tags are the building blocks of HTML.
- Tags typically come in pairs: opening **(<tag>)** and closing **(</tag>).**

Syntax:

```
<!DOCTYPE html>
<html>
  <head>
    <title>Page Title</title>
  </head>
  <body>
    <h1>Welcome to HTML</h1>
    <p>This is a simple paragraph.</p>
  </body>
</html>
```

Common tags:

- **<html>:** Root tag.
- **<head>:** Metadata.
- **<body>:** Visible content.

HTML <p> and <h>

Paragraph:

- Use the **<p>** tag to define a paragraph.
- Text automatically flows within the browser window.

Example:

```
<p>This is a paragraph in HTML.</p>
```

Headings:

- Use **<h1>** to **<h6>** for headings, where **<h1>** is the largest and **<h6>** is the smallest.

Example:

```
<h1>Main Heading</h1>
<h2>Subheading</h2>
<h3>Sub-subheading</h3>
```

HTML <a>

Definition:

- Hyperlinks connect web pages using the **<a>** tag.

Attributes:

- ***href:*** Specifies the URL of the link.
- ***target:*** Defines how to open the link (e.g., _blank for a new tab).

Example:

```
<a href="https://www.example.com" target="_blank">Visit Example</a>
```

HTML <table>

Definition:

- Tables organize data into rows and columns using the **<table>** tag..

Structure:

- **<tr>:** Table row.
- **<th>:** Table header.
- **<td>:** Table data.

Example:

```
<table border="1">
  <tr>
    <th>Name</th>
    <th>Age</th>
  </tr>
  <tr>
    <td>John</td>
    <td>25</td>
  </tr>
  <tr>
    <td>Jane</td>
    <td>30</td>
  </tr>
</table>
```

Name	Age
John	25
Jane	30

BeautifulSoup

Definition:

- BeautifulSoup is a Python library for **parsing HTML and XML** documents.
- It is widely used for web scraping.

Key Features:

- Extracts and navigates content from web pages.
- Supports multiple parsers like **html.parser and lxml.**

Syntax:

```
from bs4 import BeautifulSoup
soup = BeautifulSoup(html_content, 'html.parser')
```

Extracting Link

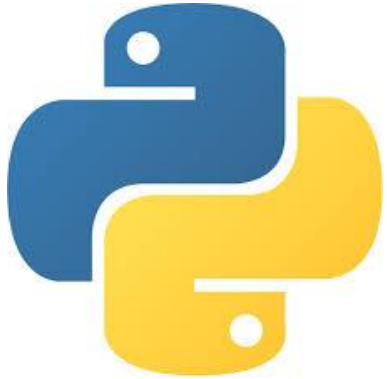
Example:

```
import requests
from bs4 import BeautifulSoup

url = "https://example.com"
response = requests.get(url)
soup = BeautifulSoup(response.text, "html.parser")

# Extract all links
links = soup.find_all('a')
for link in links:
    print(link.get('href'))
```

Thank You!



Shahriar's Intelligence Academy