MODULE 1 : DATA EXPLORATION

WORKSHOP
WEB SCRAPING & AUTOMATION
AYOUB SARAB & YASSINE ERRAKKAS

Club Informatique & Télécom
Data Cell

Web Scraping & Automation

- What Is Web Scraping ?
- Pre-requisites
- Hands On Web Scraping

What is **Web Scraping** ?

Web Crawling vs Web Scraping

Common Libraries for Web Scraping

HTML in a nutshell

code

```html
<html>
    <head>
        <title>Title of your web page</title>
    </head>
    <body>HTML web page contents </body>
</html>
```
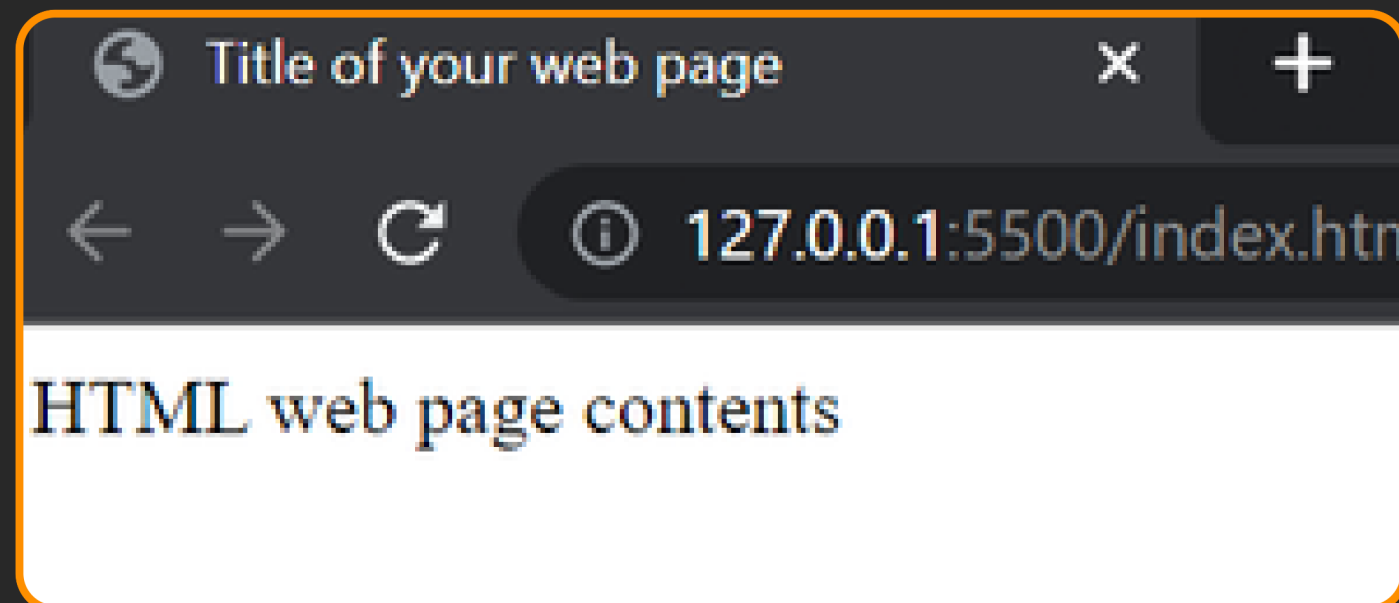
Output

HTML in a nutshell

code

```html
<html>
<head>
  <title>Data cell Web Page</title>
</head>
<body>
  <h1> title 1 </h1>
  <h2> title 2 </h2>
  <h3> title 3 </h3>
  <h4> title 4 </h4>
</body>
</html>
```

Output

title 1

title 2

title 3

title 4

HTML in a nutshell

Code

```html
<html>
<head>
  <title>Data cell Web Page</title>
</head>
<body>
  <h1>this is a title </h1>
  <p> this is a paragraph </p>
</body>
</html>
```

Output

this is a title

this is a paragraph

HTML in a nutshell

```html
<html>

<head>

  <title>Data cell Web Page</title>

</head>

<body>

  <p> visit our <a href="www.datacell.inpt">website</a> to learn more
tricks</p>

</body>

</html>
```

Output

visit our <u>website</u> to learn more tricks

HTML in a nutshell

code

```html
<html>
<head>
  <title>Data cell Web Page</title>
</head>
<body>
  <ul>
    <li> list item 1 </li>
    <li> list item 2 </li>
    <li> list item 3 </li>
    <li> list item 4 </li>
  </ul>
</body>
</html>
```

Output

- list item 1
- list item 2
- list item 3
- list item 4

HTML in a nutshell

code

```
<img src="an_image_of_your_choice.png" alt="a text of your choice">
```

Output

HTML in a nutshell

code

```
<body>
  <div > div tag   </div>
  <div > div tag   </div>
  <div > div tag   </div>
  <div > div tag   </div>


</body>
```

Output

div tag

div tag

div tag

div tag

HTML in a nutshell

```html
<body>
    <div class="div-class" > div tag  </div>
    <div class="div-class" > div tag  </div>
    <div class="div-class" > div tag  </div>
    <div class="div-class" > div tag  </div>
    <style>
        body{
            display: flex;
        }
        .div-class{
            background-color: ■red;
            text-align: center ;
            margin: 1rem;
            padding: 1rem;
            height: 1.5rem;
        }
    </style>
</body>
```

Output

HTML in a nutshell

code

```
<body>
    <div class="div-class" > div tag    </div>
    <div class="div-class" > div tag    </div>
    <div class="div-class" > div tag    </div>
    <div class="div-class" id="Y5E6S"> div tag    </div>
    <style>
        body{
            display: flex;
        }
        .div-class{
            background-color: red;
            text-align: center ;
            margin: 1rem;
            padding: 1rem;
            height: 1.5rem;
        }
        #Y5E6S{
            background-color: rgb(0, 255, 81);
        }
    </style>
</body>
```

Output

Time to practice !