**Web Scraping Project Documentation**

**1. Extract Text Data**

**Objective**

Extract the following text elements from the web page:

* All headings (<h1> and <h2>).
* All text content inside <p> and <li> tags.

**Implementation Steps**

1. Locate all <h1> and <h2> tags using a selector like soup.find\_all().
2. Locate all <p> and <li> tags and extract their text content.
3. Save the extracted data into a CSV file named Extract\_Text\_Data.CSV.

**Output**

A CSV file with the following columns:

* Heading Type ( h1, h2)
* Text Content

**2. Extract Table Data**

**Objective**

Extract data from the table, including:

* Product Name
* Price
* Stock Status

**Implementation Steps**

1. Identify the table structure and locate rows using soup.find('table').
2. Extract the data for each cell (<td> tags) under respective headers (<th> tags).
3. Save the data into a CSV file named Extract\_Table\_Data.CSV.

**Output**

A CSV file with the following columns:

* Product Name
* Price
* Stock Status

**3. Extract Product Information (Cards Section)**

**Objective**

Extract data from book cards at the bottom of the page, including:

* Book Title
* Price
* Stock Availability
* Button text ("Add to basket")

**Implementation Steps**

1. Locate all product card elements.
2. Extract the text from respective tags (e.g., <strong> for title, <p> for price and stock, <button> for button text).
3. Save the data into a JSON file named Product\_Information.JSON.

**Output**

A JSON file with the following structure:

[

{

"title": "Book Title",

"price": "£20.00",

"stock": "In stock",

"button\_text": "Add to basket"

}

]

**4. Extract Form Details**

**Objective**

Extract all input fields from the form, including:

* Field name (e.g., username, password)
* Input type (e.g., text, password, checkbox, etc.)
* Default values, if any

**Implementation Steps**

1. Locate all <input> elements using soup.find\_all('input').
2. Extract attributes like name, type, and value (default value).
3. Save the data into a JSON file named Form\_Details.JSON.

**Output**

A JSON file with the following structure:

[

{

"name": "username",

"type": "text",

"default\_value": ""

}

]

**5. Extract Links and Multimedia**

**Objective**

Extract the following:

* Hyperlink (<a> tag) and its href value.
* Video link from <iframe> tag.

**Implementation Steps**

1. Locate all <a> tags and extract their href attributes.
2. Locate the <iframe> tag and extract the src attribute.
3. Save the data into a JSON file named Links\_and\_Multimedia.JSON.

**Output**

A JSON file with the following structure:

{

"links": [

{"text": "Link Text",

"href": "URL"},

],

"videos": [

{"src": "Video URL"}

]

}

**6. Scraping Challenge: Featured Products Section**

**Objective**

Extract data from the Featured Products section with the following details:

* **Product Name**: Located within <span class="name">.
* **Hidden Price**: Located within <span class="price"> (with style="display: none;").
* **Available Colors**: Located within <span class="colors">.
* **Product ID**: Value of the data-id attribute.

**Implementation Steps**

1. Locate all product cards using soup.find\_all('div', class\_='product-card').
2. Extract the required details from child tags:
   * Product Name: <span class="name">.
   * Hidden Price: <span class="price">.
   * Available Colors: <span class="colors">.
   * Product ID: data-id attribute.
3. Save the data into a JSON file with the following structure:

**Output**

A JSON file named Featured\_Products.JSON with the structure:

[

{

"id": "101",

"name": "Wireless Headphones",

"price": "$49.99",

"colors": "Black, White, Blue"

}

]