

## Task: Web Scraping with BeautifulSoup - Extracting Structured Data from Static Websites

### Objective:

Develop a Python script that uses BeautifulSoup to scrape structured information (e.g., titles and links) from a static HTML website.

This exercise will help you understand the fundamentals of parsing HTML, navigating web structures, and extracting meaningful data using Python.

### Task Instructions:

#### 1. Select a Simple HTML Website

Choose any publicly accessible website that does not rely on JavaScript to display core content.

Suggestions:

- <http://quotes.toscrape.com>
- <http://books.toscrape.com>
- Any blog, archive, or portfolio with visible text and hyperlinks rendered in plain HTML

#### 2. Write a Python Script (scraper.py) That:

- Sends a GET request to the target website using requests
- Parses the HTML content using BeautifulSoup
- Extracts at least the following fields:
  - Title (of an article, quote, product, etc.)
  - Link (URL to the full content)
- (Optional) Add fields such as author, date, or category if available.

#### 3. Save the Output to a CSV File

- Format: Title, URL, [Other fields]

- Filename: output.csv

## **Repository Structure:**

.

- scraper.py      # Your main Python script
- output.csv      # Scraped results
- README.md      # Documentation of your project

## **README Requirements:**

Your README.md should include:

- Project Description: A brief explanation of what your scraper does
- Website Chosen: Link to the website you scraped and why you chose it
- How to Run the Script: Prerequisites and step-by-step instructions
- Challenges or Observations: Anything you found interesting or difficult

## **Submission Guidelines:**

- Upload your files to a public GitHub repository
- Ensure your code is clean, well-commented, and functional
- Include your GitHub repo link for review

## **Evaluation Criteria:**

- Clarity and readability of code
- Proper use of requests and BeautifulSoup
- Successful extraction of meaningful data
- Well-structured repository and clear documentation