

Assignment 3

Objective: Learn and apply web scraping techniques using Python to extract data from web pages, focusing on understanding and utilizing various tools and methods.

Scope: This assignment will involve scraping a simple, legally permissible website (e.g., a webpage displaying weather data, book listings, or movie reviews). Ensure the chosen website allows scraping by checking its **robots.txt** file and terms of service.

Tasks:

1. Choose a Website:
 - Identify a website to scrape. Verify that the website permits scraping.
 - Briefly describe the website and the data you intend to scrape.
2. Set Up Your Environment:
 - Install necessary Python libraries (requests, Scrapy, etc.).
 - Set up a Python script or Jupyter notebook for your scraping code.
3. Data Extraction:
 - Use the requests library to send an HTTP request and retrieve the content of the webpage.
 - Parse the webpage content using Scrapy to extract relevant data (e.g., titles, descriptions, ratings).
4. Data Cleaning:
 - Clean and organize the scraped data into a readable and usable format.
 - Handle any missing or inconsistent data entries.
5. Data Storage:
 - Store the extracted data in a structured format like a CSV file or a Pandas DataFrame.

Submission: A detailed report including the code (Jupyter Notebook is okay), and scraped data (or a sample of it). For your comments and headers, please use the *markdown* option.