Detailed Project Description: Web Scraper for News Headlines

Introduction:

The Web Scraper for News Headlines is a Python-based project designed to automate the collection of top headlines from a public news website. The scraper utilizes the requests library to fetch HTML content and BeautifulSoup to parse and extract relevant headline elements, such as or tags. The extracted headlines are then stored in a .txt file for offline reference, analysis, or reporting.

Objective:

The main objective of this project is to provide learners with practical experience in web scraping, HTML parsing, and automated data collection. By completing this project, students will understand how to use Python libraries to interact with websites, extract meaningful information, and save it persistently.

Tools & Technologies:

- Python: Core programming language for implementation.
- Requests: To fetch HTML content from the news website.
- BeautifulSoup (bs4): To parse and extract headline elements.
- Text File Handling: To save the collected headlines persistently.

Implementation Guide:

- Use the requests library to send an HTTP GET request to the target news website.
- Parse the fetched HTML content using BeautifulSoup.
- Locate headline elements (commonly inside or tags).
- Extract the text content of these tags and store them in a Python list.
- Open a .txt file in write mode and save each headline for persistent storage.
- Run the script periodically or on demand to collect the latest headlines.

Features:

- Automatically fetches the latest headlines from a chosen news site.
- Extracts headlines using HTML parsing techniques.
- Saves headlines to a text file for future reference.
- Simple, reusable, and extendable script.
- Can be scheduled with cron/Task Scheduler for automation.

Outcome:

By completing this project, learners will gain practical knowledge of web scraping using Python, including sending HTTP requests, parsing HTML structures, and handling persistent data storage. The final deliverables will include a Python script that fetches headlines and a .txt file containing the extracted data. This demonstrates automation in data collection and lays the foundation for

advanced projects such as news analysis, sentiment analysis, or building real-time dashboards.