

## **Instructions:**

1. write the codes in the logbooks
2. Draw the flowchart diagram

The project is designed to reinforce your understanding and application of:

1. Web scraping using requests and BeautifulSoup.
2. String and list manipulation to clean and filter textual data.
3. Custom Python modules to promote code modularity and reuse.
4. Exception handling to ensure robustness and reliability.
5. Project structuring using separate Python files (main.py, scraper.py, and text\_utils.py) to reflect real-world software organization.

## **TASK**

Create a Python script that scrapes news headlines from the website <https://www.theguardian.com/international>, cleans and processes the headlines using string and list methods, and handles errors gracefully using exceptions. The script will be modular, organized using custom modules and possibly packages.

## **Project Description:**

1. Use requests and BeautifulSoup (from bs4) to scrape headlines from a news site.
2. Use string methods to clean text: remove extra spaces, punctuation, convert to title case, etc.
3. Use list methods to process and store headlines, e.g., filtering duplicates or unwanted keywords.
4. Organize the project using:
  - A main script (main.py)
  - A custom module for scraping (scraper.py)

- A custom module for text processing (text\_utils.py)

5. Add exception handling for:

- Network errors
- HTML parsing issues

Empty or malformed data