# **Instruction**:

1. To analyze articles, download the "input.xlsx" file which contains a list of URLs.
2. Create a Python environment with the required dependencies, including BeautifulSoup, nltk, Pandas, and Requests.
3. Open the terminal or command prompt and navigate to the project directory.
4. Run the "web\_content\_extractor.py" Python script to extract the article text from the URLs and save it in text files with the URL\_ID as their filename:
5. Run the “main.py” Python scrip to extract article text from URLs and save it in the text files with the URL\_IS as their filename. Also this script will perform textual analysis on the extracted article text and compute the required variables.
6. The output will be saved in an Excel file named "output.xlsx" in the same directory as the input file.
7. Verify that the output file format matches the "Output Data Structure.xlsx" file provided.
8. Note that the "input.xlsx" file should be located in the parent directory, and the current working directory should be named "PythonFile."
9. A "textfile" folder will be created after the run of “main.py” python script, to store all the extracted text content with the URL\_ID as the base name.
10. A separate "LogfileFolder" will be created to store all the log files for each module.
11. Console handler and file handle is set to ERROR level, It can be changed from the logger module.