

Data Extraction:

Use Python programming to extract article text from the provided URLs.

Libraries such as BeautifulSoup, requests, or Selenium can be utilized for web scraping.

Ensure that only the article title and text are extracted, excluding any website headers, footers, or other irrelevant content.

Save the extracted text into text files, with each file named after its URL_ID.

Text Analysis:

Utilize NLTK and TextBlob libraries for text analysis.

Implement functions to compute the required variables mentioned in the "Text Analysis" document, such as positive score, negative score, polarity score, subjectivity score, etc.

Ensure that the computation is accurate and follows the defined formulas.

Create a DataFrame containing the computed variables for each extracted text.

Save the DataFrame to an Excel file, following the specified output format.

Running the .py file:

Ensure that the Python script is well-documented with comments explaining each part of the code.

Include all necessary dependencies in the script or provide instructions on how to install them.

The script should be executable from the command line or any Python IDE.

Provide clear instructions on how to run the script, including any command-line arguments or input files required.

After running the script, the output should be generated in the specified Excel file format, containing the computed variables for each article.

Dependencies:

Include a list of all dependencies required to run the Python script.

Specify the version numbers of the dependencies if necessary.

Ensure that the dependencies are easily installable using pip or conda.

By following these instructions, users should be able to run the .py file provided, extract data from the URLs, perform text analysis, and generate the output Excel file as specified in the assignment.