Instruction

1.Import Libraries:

You import necessary libraries such as numpy, re, os, nltk, pandas, BeautifulSoup, and others.

2.Define File Paths:

You define the paths for stop words, positive words, and negative words files.

excel_file_path is set to the path of the Excel file containing URLs.

3.Read Input Excel File:

Read the input Excel file into a DataFrame named input.

4.Tokenization and Text Cleaning:

You define a function tokenizer to tokenize text and remove stop words and punctuation.

5.Sentiment Analysis Dictionaries:

You load lists of positive words, negative words, and stop words.

6.Define Sentiment Analysis Functions:

Functions like positive_score, negative_score, subjectivity_score, polarity_score, etc., are defined for sentiment analysis.

7.Web Scraping:

You use the requests library to scrape content from a list of URLs.

The content is cleaned and added to the corps list.

8.Delete Rows from Excel:

A function delete_rows_from_excel is defined to delete rows from the input Excel file based on failed URLs.

9.Data Analysis and DataFrame Creation:

You create a DataFrame df with the scraped content and corresponding titles.

Various text analysis features are calculated and added as new columns in the DataFrame.

10.Data Cleaning and Output:

Unnecessary columns are dropped from the DataFrame.

The index is set to start from 1.

The final DataFrame is saved to a new Excel file named 'Output Data Structure.xlsx'.

11.Install -pip install numpy re nltk pandas beautifulsoup4 fake_useragent