INSTRUCTIONS

1. Importing Libraries

Ensure you have the required libraries installed before running the script. Use the following commands to install them: *pip install pandas numpy seaborn matplotlib requests beautifulsoup4 nltk wordcloud*

2. Data Extraction

Make sure you have the input.xlsx file with the necessary URLs and URL_IDs. The script will create an 'Articles' folder to store the extracted text files.

3. Error Handling

If there are errors in extracting articles (e.g., no specific HTML tag found), the script will print an error message and skip that URL.

4. Variable Initialization

The script initializes stop words, positive words, and negative words lists. Ensure the 'StopWords' directory and the 'MasterDictionary' directory are in the correct location.

5. Defining Functions to Calculate Variables

The script defines various functions to calculate different variables required for sentiment analysis and readability analysis.

6. Output File

The script reads the structure from ‘Output Data Structure’ and text files from the 'Articles' folder and performs sentiment and readability analysis.
The results are saved in an Excel file named *'Sentiment_Analysis_Results.xlsx'*.

7. WordCloud

The generate_wordcloud function generates a WordCloud for a specific text file. You can customize the file name to visualize WordClouds for different articles.

8. Interpretation

The interpretation section provides guidelines on interpreting sentiment analysis results, including the meaning of polarity and subjectivity scores.

9. Execution

Place the script and the 'input.xlsx' file in the same directory.
Make sure the 'StopWords' and 'MasterDictionary' directories are available. Run the script using the following command:
python script_name.py

Note:
These 2 URL ids are not valid (Sorry, but the page you are looking for doesn't exist.)

- URL ID 11668
- URL ID 17671.4