

Approach to the Solution:

1. Fetching Web Data:

- Utilizing the `requests` library to retrieve the HTML content of the web page.
- Employing BeautifulSoup (`bs4`) to parse the HTML and extract relevant information such as the title and main content.

2. Stop Words:

- Creating a function to extract stop words from files in a directory, organizing them into different categories.
- Merging all stop words into a single list to be used for text cleaning.

3. Cleaning Stop Words:

- Developing a function to remove stop words from the extracted text.
- Optionally removing personal pronouns using NLTK's English stop words list.

4. Sentiment Analysis:

- Defining a function to calculate sentiment scores (positive and negative counts, polarity score, and subjectivity score) based on predefined positive and negative word lists.

5. Readability Analysis:

- Creating a function to compute various readability metrics such as average sentence length, percentage of complex words, and Fog Index.

6. Personal Pronouns:

- Implementing a function to calculate the number of personal pronouns and the average word length.

7. Main Execution:

- Iterating through a list of URLs.
- Applying the defined functions to extract features and metrics from the web content.
- Constructing a DataFrame containing the analysis results.

How to Run the .py File to Generate Output:

1. Install Dependencies:

- Ensuring that the necessary dependencies are installed. These typically include:
 - `requests`: For making HTTP requests.
 - `BeautifulSoup` (`bs4`): For parsing HTML.
 - `nltk`: For natural language processing tasks.
 - `syllapy`: For estimating syllables in words.

2. Prepare Input:

- Ensuring that there are a list of URLs or a method to obtain URLs for web content analysis.

3. Run the Script:

- Executing the Python script containing the defined functions and main execution logic.

4. View Output:

- Once the script completes execution, it should generate an output DataFrame containing the analysis results.

Dependencies Required:

- ``requests``
- ``beautifulsoup4`` (or simply ``bs4``)
- ``nltk`` (Natural Language Toolkit)
- ``syllapy`` (for estimating syllables)