**Installed the required packages using pip:**

pip install pandas

pip install requests

pip install beautifulsoup4

pip install nltk

Also, downloaded **punkt** and **stopwords** package from nltk using the following command.

nltk.download('punkt')

nltk.download('stopwords')

**To execute the script:**

python text\_analysis.py

**Script Explanation:**

Script for web scraping and text analysis is written in text\_analysis.py

1. Scraping the article content and saving that in the text file is written in **save\_scarped\_data\_in\_txt\_file** function**.**
2. All the seven steps mentioned in the Text Analysis is written in the **text\_analysis** function.
3. Output CSV is written is written in the final\_output\_csv function.
4. Finally, all the functions are called in the main function

And text files are saved in the “Extracted text files folder” automatically.

**Notes:**

1. Formula is not mentioned to calculate the "Syllable Count Per Word". So calculated using formula,

**total count of syllable / total no of words**

1. In few places, it is not mentioned that which word count to proceed with. So, just used the cleaned word count.

Cleaned word count – count of words which are not in the StopWords files