

Text Extraction and Analysis

This project is based on Web Scraping and analyzing the articles which are web scrapped from different websites.

The article title and articles are extracted from webpages and saved into text file. From those text files, the article was analyzed and the following attributes were calculated:

1. Positive Score
2. Negative Score
3. Polarity
4. Subjectivity
5. Average sentence length
6. Percentage of complex words [words which have more than two syllables]
7. Fog index
8. Average number of words per sentence
9. Complex words
10. Word count
11. Syllable per word
12. Personal pronouns
13. Average word length

The output file is saved under the name ‘final_output.csv’.