

INSTRUCTIONS:

Extracting text from website using scrapping

- For this approach I have used BEATIFULSOUP. The code is provided in text_extraction.py file and the execution time outputs with respective to code is given in text_extraction.html.
- The entire text extraction task is done in iterative approach. By analyzing the tags and classes in given URL links I have made generalization process to retrieve the data from the web using web.
- The extracted data is written into text files by using given URL id's in the input file.
- Then task of exaction was completed.

Text Analysis:

Semantic analysis:

- Semantic analysis is performed with respective given conditions like using stop words list and master dictionary.
- First cleaning process is done with respective to stop words.
- Then, Dictionary preparation is done with master dictionary and stop words.
- Then all the variables are derived using functions.
- Polarity ratio is adjusted to -1 to +1 using normalization {min max scaler}
- Subjective ratio is also adjusted to 0 to 1 using normalization {min max scalar}

Analysis of Readability:

- In this approach the previous tokenized text is used to find the asked values with respective to formulas.
- The complex words are found with syllables more than 2.
- With the help of function calls entire things will be done.

Remaining variables like:

- AVG SENTENCE LENGTH
- FOG INDEX
- AVG NUMBER OF WORDS PER SENTENCE
- COMPLEX WORD COUNT
- WORD COUNT
- SYLLABLE PER WORD
- PERSONAL PRONOUNS
- AVG WORD LENGTH
- All are found with respective to function calls.
- Every instruction is given with respective to function block in the code itself
- Entire text analysis was done in the Text_Analysis.py and output with respective to code is give using Text_Analysis.html.

Note: Entire outputs are given in the output data structure file that was provided in the assignment.

- Entire Text Analysis is done in the recursive manner, In such way that during each iteration the program will take text file and perform all the functional operations.

- Those functions will give the output needed for every text file.
- Finally, the entire outputs are loaded into write list then I created the program to write the contents of the write list into the output data structure csv file.
- In this way the assignment was finished.
- Python and its libraries are used heavily to achieve the task.

Thank You