

Assignment1

Libraries are used: codecs, string, nltk, os

Implemented method:

1)read:

It is used to read file using filename and it's path.

2)sentence:

It is defined to count total number of string in an given file. I have used inbuilt function "sent_tokenize" to divide our text files data into tokenize array.

3)word_count:

Program passed tokenized array into this function and it used inbuilt "split()" function to calculate word count from it.

4)sent:

This method basically taking our tokenize array and word name to calculate – how many sentence starting with that words.

5)sent_end:

This method basically taking our tokenize array and word name to calculate – how many sentence ends with that words.

6)sent_mid:

This method basically taking our tokenize array and word name to calculate – how many times sentence contains that words.

7)email:

This method basically taking tokenize array and try to match with valid email id.

Assumption:

It is counting word including "@" and other punctuations symbols. It will count "dibyendu_roy" as a single words. It is also counting digits and combinations of digit and letters as word.