

Assignment 1 report nlp

1. Paragraphs are separated by 2 or more '\n' characters.
2. Sentence ends with anyone of the following followed by a space and Capital letter or empty spaes (end of line):
 - a. .
 - b. ?
 - c. !
 - d. ?"
 - e. !"
 - f. ."
3. Words can be alphanumeric sequences, can also contain "-" or "." " " " " within themselves and have spaces on both sides or end with a ., ?,!,?","!,.,".
4. The first letter of the word to enter need not be capital but rest characters are case sensitive. Its a starting word if it is beginning word of the text, or is followed by the above 6 (2.(a-f)) tokens and spaces and after this word, we have a space.
5. The word followed by a 2(a-f) then space and then captial letter or end of text ie spaces till the end.
6. Removed all punctuations and counted number of times the word occurs.