# Assignment 2
# Natural Language Processing

Kanika Saini (2016047)

Your task in this assignment is to write a python program that accepts as input a plain text newspaper article file and returns the following.

**Assumptions are as follows:**

1) Number of paragraphs, sentences, and words contained in the article.
   - Paragraphs
     - Two paragraphs have two or more newline character between them.
     - The last paragraph may have a new line at the end, not necessarily.
   - Words
     - A.B.C. is one word.
     - Anti-national is one word.
     - Google.com is one word.
     - Dr. Xyz are two words.
     - Any symbol separate from an alphabet or a digit is not a word. For instance, "-" is not a word.
   - Sentences
     - Sentences can end with ", ', ., !, ? or a period
     - Oh! I'm Kanika. - two sentences.

2) Given a word as input, number of sentences starting with the word.
   - The book was named butterfly. "The Hero" was the pen name of the author who said, "Life is precious."! The end was beautiful. - three occurrences of "the"

3) Given a word as input, number of sentences ending with the word.
   - With the same assumptions of finding the sentences, the word is just found before the ending of the sentence.

4) Given a word as input, count of that word in the input file.
   - B.P.M is one word.
   - Word can match when in between symbols. For instance, hello would match hello in "hey-hello".