CPSC 217 Assignment 4

Due: Friday December 9, 2016 at 5:00pm
Weight: 8% of the overall grade

Individual Work

All assignments in this course are to be completed individually. Students are advised to read the guidelines for avoiding plagiarism located on the course website. Students are also advised that electronic tools may be used to detect plagiarism.

Late Penalty

Late assignments will not be accepted.

Questions and Marking

If you have any questions regarding this assignment, please ask your TA because it is your TA who will mark your assignment. In addition, your TA has full control over the assignment (i.e., they may have some specific requirements that your should follow).

Submission Instructions

Submit **only** your Python program (a .py file) to the Assignment 4 drop box in D2L.

Description

Write a Python program that reports the number of paragraphs, sentences, and words in a text file. Your program must meet the following requirements:

- Ask the user for an input file name with *.in* extension. The input file is a text file.
- The output file will have the same name as the input file but with *.out* extension. For example, if the user input *a4testfile.in* as the input file, *a4testfile.out* will be the output file.
- Report the number of paragraphs in the input file. For each paragraph, report the number of sentences and words.
- Report the total number of words in the whole input file.
- Report the word(s) that occurs the most number of times.
- The output must be in the format shown below. Additional sample input and output files can be found on D2L.

```
# of paragraphs: 3

Paragraph 1:
  # of sentences: 3
  # of words: 38

Paragraph 2:
  # of sentences: 3
  # of words: 55

Paragraph 3:
  # of sentences: 3
```

# of words: 65

Total # of words: 158
"a" occurs 20 times
"the" occurs 20 times

Additional Details

- You can assume the following conditions in your input file.
    - Each paragraph has an empty line (without space) at the end. Check the sample input files to see what it means.
    - There are only five different punctuations in the file. They are:
        - Period (.), exclamation mark (!), question mark (?), comma (,) and semicolon(;)
    - Always have at least one word.
    - No numbers with a decimal point.
    - Text follows correct punctuation rules.
- You should ignore cases when counting. For example, the sentence "I put an apple on an Apple laptop." has two occurrences of the word *apple*. When you report the result, you can write *"Apple" occurs 2 times*, *"apple" occurs 2 times*, or *"APPLE" occurs 2 times*.
- Depending on the number of occurrence, use either *time* or *times*.
- Numbers are considered as words.
- In the terminal, your program should ask for user input (an input file name with extension). After that, your program should tell the user the name of the output file with extension (also in terminal). This is how we usually deal with input and output.
- The actual report should be written in the output file.
- You should make proper use of functions. You need to decide how to properly break down your program into functions.
- Your program must include appropriate comments, including a header that includes your name and student number and describes the purpose of your program. There should also be comments within the program. Comments for each function must be provided: purpose of the function, parameter(s), and return value(s).
- You must not use any global variables.
- Your program must use good programming style. This includes things like appropriate variable names, good comments, minimizing the use of magic numbers, etc.

Grading

This assignment will be graded on a combination of functionality and style. A base grade will be determined from the general level of functionality of the program (Does it read and write files successfully? Does it count correctly? Does it report the counts in the correct format?) The base grade will be recorded as a **mark out of 14**.

Style will be marked on a subtractive scale from 0 to -3. For example, an assignment that receives a base mark of 12, but has several stylistic problems such as improper use of functions or insufficient comments resulting in a -2 adjustment will receive an overall mark of 10.