CSC 111aD – Fall 2013 Name _____
Lab_10
Due 5pm, Thursday, 11/14/13

Overview

The purpose of this lab is to develop a program for reading a text file and computing some basic statistics about the text, including the Flesch-Kincaid Grade Level (FKGL) statistic. Read about FKGL at http://en.wikipedia.org/wiki/Flesch-Kincaid_Readability Test before proceeding. Make notes about the values required to compute the FKGL index for a text.

Deliverables

Turn in this lab sheet with the table on page 2 completed. Submit your program using the file naming convention *Lab_10_lastname_initial*.py through the assignment link on Sakai.

Steps

1. To keep things simple you will want to have all your lab files in the same folder. Download the following four text files from Sakai->Resources->Labs->Lab_10 Materials: DonQuixote.txt, LittleWomen.txt, Marian.txt, SteamEngine.txt. Also download 'countWords.py'. You should use 'countWords.py' as an example to help with this lab. Your tasks will be slightly different but will have a lot in common with the countWords example.

2. Create a file named 'Lab_10_lastName_Initial.py' using your own name and initial.

3. Write your program incrementally. Your program should begin by prompting and reading the name of a text file typed by the user. If the file exists then your program should open it for reading. Your program should only read through the file once.

4. Your program should read the text file and compute and report the following statistics:
   - The number of words
   - The number of sentences
   - The average word length (rounded to one decimal place)
   - The standard deviation of the word lengths (rounded to one decimal place)
   - The Flesch-Kincaid Grade Level of the text (rounded to one decimal place)

   Example run:
   ```
   What .txt file would you like to analyze? Marian.txt

   Marian.txt contains 37759 words in 2794 sentences.
   Average word length:  4.1
   Std Dev of word length:  2.0
   Flesch-Kincaid Grade Level: 5.7
   ```

5. You must use a function to compute the standard deviation of the word lengths. I suggest writing a single function that will take any list of numbers as its parameter and will return a tuple consisting of the (sum, average, standard deviation) of the numbers in the list.

6. Modify the program so that it will compute and print the FKGL. Here are some of the questions you will need to answer as you proceed:
   - What are the values needed to compute the FKGL?

CSC 111aD – Fall 2013 Name _____
Lab_10
Due 5pm, Thursday, 11/14/13

- How are you going to determine the number of sentences? The reality is that it's very hard to do exactly. I suggest that you simply count the number of times that either a '.', ';', or '?' appear in the text and use that as an approximation to the number of sentences. That's close enough.
- How are you going to determine the total number of syllables in the text? This is actually a VERY challenging problem so we're going to use an approximation that works perfectly well for most English text. The total number of syllables in English text can be closely approximated by summing the total number of letters in all of the individual words and dividing by 3.

Your Results:

	Words	Sentences	Avg Word Length	StdDev of Word Length	FKGL
Marian.txt					
SteamEngine.txt					
LittleEngine.txt					
DonQuixote.txt					

Scoring (50 points total)

15pts – program is readable, uses meaningful variable names, and has useful and adequate comments
20pts – program computes the correct answers
5pts - program strips out punctuation marks correctly
10pts – program uses a function to compute and return the standard deviation of the word lengths; the function does not print out the value, it returns the value