CSCE100 Introduction to Informatics
Fall 2021

Programming Assignment 2: Sentence Analyzer

Points: 100 points. Assignment Date: September 14, 2021 Due Date: September 21, 2021

Objectives

1. To familiarize with writing and running Python programs and the Python environment
2. To practice computational thinking skills to develop the solution approach
3. To familiarize with the use of loops (e.g., the for and while loops)
4. To familiarize with data structures, particularly arrays/lists
5. To familiarize with file input/output in Python
6. To be exposed to the use of built-in functions
7. To be exposed to the use of built-in modules or packages (e.g., import math)
8. To familiarize with the use of online documentations on Python

Relevance to Informatics or Data Science

1. A user interface program for collecting data (e.g., data)
2. A program for pre-processing or processing data (e.g., texts)
3. A solution for parsing texts and computing certain properties automatically
4. (Bonus) A solution for parsing texts and extracting targeted words automatically

Problem

Write a program that will prompt the user to select between two options: (1) enter a sentence and (2) calculate some specific statistics of the sentences. The program will *keep prompting the user until they choose to select option 2*. Then the program should use loops in its implementation. It must print out the number of sentences entered by the user, and it must compute and print out the total number of vowels in the sentences. Here are some additional requirements:

- The program is required to display an explanation of the program (e.g., its expected range of input values) in the beginning before prompting the user for a number. (5 points)
- The program is required to use an array to store the sentences such that each sentence is an element of the array. (5 points)
- The program is required to use at least one loop structure. (10 points)
- The program should display proper error messages when an invalid input is entered. (5 points)
- Your program is ***not*** allowed to use Python's built-in functions that compute the occurrence of each letter (note: treat lowercase and uppercase of a letter the same).

- You must document your program (see https://devguide.python.org/documenting/).
    - Name, Date, Affiliation, a description of the program, what inputs does it need, what outputs does it generate (5 points)
    - Inline comments in the program (5 points)

IMPORTANT: This assignment is more open-ended compared to the first assignment. The solution approach is not described clearly so that you have a chance to practice computational thinking: how to break the problem down into smaller subproblems, and how to sequence your steps into an algorithm. Think about this: How many substantial subproblems are there?

Example Run Session (red texts are user response)

```
Welcome to the Sentence Analysis program!
This program finds the total number of vowels in sentences.

1. Enter a Sentence
2. Generate Statistics

Choose 1 or 2: 1
Enter your sentence here: To be or not to be. That's the
question.
Choose 1 or 2: 1
Enter your sentence here: Try not. Do. Or do not. There's no
try.
Choose 1 or 2: 2

The number of sentences entered: 2
The number of vowels: 20

Thank you for using this program. Bye!
```

Bonus (15 points)

Once you have a working solution, extend it such that the program also identifies and prints out the number of occurrences of each vowel to the screen.

Example Bonus Run Session (red texts are user response)

```
Welcome to the Sentence Analysis program!
This program finds the total number of vowels in sentences.

1. Enter a Sentence
2. Generate Statistics
```

```
Choose 1 or 2: 1
Enter your sentence here: To be or not to be. That's the
question.
Choose 1 or 2: 1
Enter your sentence here: Try not. Do. Or do not. There's no
try.
Choose 1 or 2: 2

The number of sentences entered: 2
The number of vowels is: 20
The number of 'a' or 'A' is: 1
The number of 'e' or 'E' is: 6
The number of 'i' or 'I' is: 1
The number of 'o' or 'O' is: 11
The number of 'u' or 'U' is: 1

Thank you for using this program. Bye!
```

Handin

1. The submission deadline for all handins is 11:00 AM September 21, 2021. **Late handins will *not* be accepted or graded.**
2. You are required to handin a screen capture of your "testing session" using your program. (10 points)
3. You are required to handin all program files. (10 points)
4. You are required to handin online the above files to Canvas under Programming Assignment #2.

Think About

Now, think about what if we want to build a system that computes statistics for thousands of sentences, or even, millions of numbers. Do we want to input the sentences one by one manually? What would be some common challenges or issues with that approach? Are there other ways to input the data? By the same token, what if we want to generate different types of statistics, for different subsets of sentences, and thus will produce many different tables? How should we store the tables of sentences? Do we want to copy the sentences down one by one and re-enter them, say, into an Excel spreadsheet? (Hint: Think about File Input/Output.) (Hint: Think about Big Data, Scalability, and Reliability, and how they relate to Informatics or Data Science.) Note also that looking for words that start with an uppercase letter is a very valid application in today's text processing, with the purpose of finding place names, pronouns, etc., to automatically metatag documents or texts.