COP4338 Systems Programming

Programming Assignment 4:
Hash Table and I/O

Knight Foundation School of Computing & Info. Sciences
Florida International University

In this assignment, you are asked to write a C program that checks the spelling of words entered via standard input stream. The program must output each misspelled word, its line and column of occurrence in the input, as well as possible corrections for the word.

Command-line Arguments

All command-line arguments are optional for this program. Here is the list of arguments and their meaning:

- *-d dictionary_filename*: the argument that comes after '-d' specifies the name (and possibly path) of the file containing all English words listed with correct spelling. The default file name is "words.txt" which refers to the file given on Canvas.

- *-r*: means that the program must suggest possible corrections for any misspelled word by replacing one of its letters by another letter. For example, a possible correction for the misspelled word "mistape" is "mistake" constructed by replacing 'p' by 'k'.

- *-s*: means that the program must suggest possible corrections for any misspelled word by swapping two adjacent letters of the word. For example, a possible correction for the misspelled word "pinaepple" is "pineapple" constructed by swapping 'a' and 'e'.

- *-e*: means that the program must suggest possible corrections for any misspelled word by erasing one letter of the word. For example, a possible correction for the misspelled word "earthqueake" is "earthquake" constructed by erasing the second 'e' letter from the misspelled word.

Data Structure

To store the set of all words with correct spelling, you need to use the hashtable whose C implementation is given on Canvas (hashset.c). In this implementation, function *set_init* can be used to create an empty set, function *insert* can be used to add a new element to the set, and function *search* can be used to search for a given element in the set.

Input and Output Format

You need to assume that the user can input up to 1000 lines made of many words, each with length less than 60 characters. It is the responsibility of your program to remove all characters but a-z, A-Z, and apostrophe (e.g.punctuation marks, special characters, and digits) when checking the spelling of words.

The output of the program must be in the comma-separated value (CSV) format. For example, if the input has two misspelled words "miistake" (Line 3, Column 5), and "aap" (Line 6, Column 1) and the program is executed with flags "-rse", the generated output must be this:

"word,Ln,Col,Possible Corrections\nmiistake,3,5,mistake\naap,6,1,bap,cap,gap,hap, jap,lap,map,nap,pap,rap,sap,tap,yap,zap,adp,alp,amp,app,asp,atp,aaa,aah,ap"

Submission

You need to submit a *.zip* file compressing the makefile, C source file(s), and possibly any header file(s) related to the assignment (*.c* and *.h* files). Please **use "spell_checker" as the name of your program's executable file in the makefile**.