

CSE102 – Computer Programming (Spring 2021)

Homework #7

Handed out: April 15, 2021.

Due: 11:55pm April 28, 2021.

Hand-in Policy: Via Moodle. No late submissions will be accepted. A student with number 20180000001 should hand in 20180000001.c for this homework.

Collaboration Policy: No collaboration is permitted.

Grading: This homework will be graded on the scale of 100.

Homework Description: In this homework, you will write a C program that will check and correct word spelling mistakes. 2 files will be given to you. These are "dictionary.txt" and "text.txt". Correct spelling of the words will be found in "dictionary.txt". The article that needs to be checked and corrected will be found in the "text.txt" file.

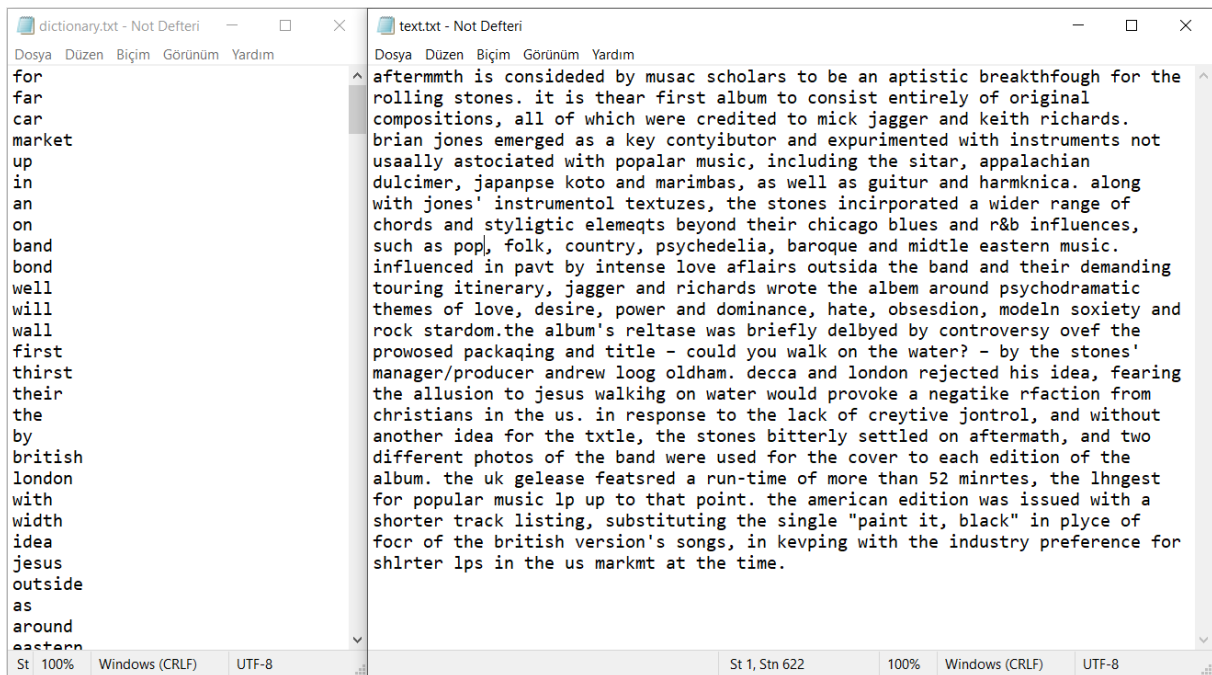


Figure. 1 Sample dictionary and text files.

You should take the words from the text found in "text.txt" and compare them with the correct words in the "dictionary.txt" file. Comparing the word taken from "text.txt", you need to find the closest word found in the "dictionary.txt" file. When the best matching word is found, it should be replaced with the misspelled word. It is assumed that there is a maximum 1 letter mistakes in misspelled words.

Figure 1 shows an example. The actual test will be done with another set of files with different text and dictionary words. You should take the following into account:

- Make sure that you extract the word such that white spaces and punctuation marks are properly handled.
- If there is an exact match to the given word, there is no need to change it.
- If there are more than one matching words (with one letter difference), you must take the word that changes from consonant to vowel as correct. For example, if the given word "ssarch", you have two candidates in the dictionary, i.e., "search" and "starch", you should pick the word "search" as it changes "s" to "e" when you replace "ssarch" to "search".
- It is possible that you might have more than one matches with one letter difference. In that case, you are expected to find the first three matching words in the "dictionary.txt" file and apply the previous rule to choose the answer. In case of a tie, you use the **second** occurrence.

You have to write a minimum of 2 functions. First, you should call `fix_spelling_error(...)` function from main function with the two file names. Then you need to call the `find_best_correction(...)` function that finds the best match to the given word in the given dictionary. You are not limited to open the files once (you may open and close them as many times as you want).

Bonus (50pt): You will be rewarded with bonus points if you write the program by assuming that a maximum of 2 letters of mistakes are made in misspelled words and the previous rules apply. Likewise, you need to correct the mistakes with 2 letters and find the correct word and replace the original. You will have to have `fix_spelling_error2(...)` function as the main part of your solution.

General Rules:

1. Obey and do not break the function prototypes that are shown on each part, otherwise, you will get zero from the related part.
2. The program must be developed on Linux based OS and must be compiled with GCC compiler, any problem which rises due to using another OS or compiler won't be tolerated.
3. Note that if any part of your program is not working as expected, then you can get zero from the related part, even it's working in some way.
4. You can ask any question about the homework by using the forum in the Teams page of the course.