# SPELL CHECKER


## BACKGROUND

Most publishing companies have sophisticated software programs to spell-check and grammar-check their documents. For this assignment, you must write a program to spell check text files

## FEATURES

You must write a program named **SpellCheck**. The program must accept command-line arguments where the first command-line argument is the name of a dictionary file, the second argument is the name of a source file and the third argument is a parameter K (a positive integer).

The dictionary file is a text file that alphabetically lists all properly spelled words in the English language. The source file is a text file that must be spell checked. Any word in the source file that is not in the dictionary (all words should be considered in a case-insensitive fashion) AND has a length that exceeds K is misspelled. All other words are spelled correctly. You are to print all of the misspelled words and the line numbers on which they occur in the source file. A word will only contain letters a-z, A-Z, dashes and apostrophes.

### *OUTPUT FORMAT*

The misspelled words must be printed in descending order of the number of times that each misspelled word occurs in the source file. If two words have the same number of occurrences they must be sorted in reverse-alphabetical order. Each word must be followed by a sorted list of line numbers where the misspelled word occurs. This format is generalized as shown below:

**`misspelled-word (number-of-times-misspelled),l1,l2,…,ln`**

where l1, l2, ..., ln are the line numbers of each occurrence of the misspelled word. No line number should be listed more than once for any misspelled word. The following lists a few lines of a sample output where K was given as 3. In this example, the word ‘laxity’ occurs in the source file four times (on lines 43, 46, 67, and 71) but the word ‘laxity’ does not appear in the dictionary.

```
...
horne (8),12,13,18,25,84,90
verner (4),54,67,69,87
laxity (4),43,46,67,71
distractedly (4),32,67,87,101
...
```

### *INPUT FORMAT*

The dictionary file contains one word per line and is sorted in a case-insensitive manner. Make no assumptions concerning the number of words in the dictionary since different dictionary files may be used at different times.

## TESTING

For this assignment you must run your program on the two source files 'source1.txt', 'source2.txt'and 'source3.txt' using the dictionaries 'words1.txt' and 'words2.txt' as found on the web site. For each combination of source and dictionary file you must run your program using K values of 2, 4, 8 and 12. For each test (there were will 24 total tests) you must provide the number of milliseconds that your code takes to execute. The timings must include the time taken for every aspect of your code, from reading the dictionary to scanning the source file to printing the report on misspelled words. You must then enter your timings into an Excel spreadsheet having the structure shown below.

| Source | Dictionary | K | Time (ms) |
|---|---|---|---|
| **source1.txt** | words1.txt | 2 | 419 |
| **source1.txt** | words1.txt | 4 | 395 |
| **…** | … | … | … |

## NOTE

The fastest submission will receive 15 extra credit points. The second-fastest solution will receive 10 extra-credit points.