

## **CSE 222 HOMEWORK #9 Part 2**

**Orhan Aksoy**

**09104302**

### **TABLE OF CONTENTS**

<b>1.</b>	<b>PROBLEM DESCRIPTION .....</b>	<b>2</b>
1.1.	REQUIREMENTS .....	2
<b>2.</b>	<b>ANALYSIS.....</b>	<b>3</b>
<b>3.</b>	<b>DESIGN AND IMPLEMENTATION .....</b>	<b>3</b>
<b>4.</b>	<b>TESTING.....</b>	<b>4</b>
4.1.	TEST CASE 1 – NO SPELLING ERRORS .....	5
4.2.	TEST CASE 2 – THREE SPELLING ERRORS WITH EACH KIND. ....	6
4.3.	TEST CASE 3 – RANDOM CHARACTERS ON 50 LINES OF TEXT .....	7

## 1. Problem Description

A spell checker application is to be developed. The application will have the capability to load a text file of dictionary words. In the file, every word is placed in a single line. The application will also load a text file that it will check for spelling.

During spellcheck, if a word is found to be misspelled (not found in the dictionary), a set of alternative words are generated. This new set is comprised of the words that can be generated by adding a character to the word, removing a character from the word, or swapping two adjacent characters within the word. In the output, this list will also be presented as suggestions to the misspelled word.

### 1.1. Requirements

1.1.1. The application shall accept two arguments: The first one is a string representing the file name of the dictionary. The second is also a string representing the text file that is going to be checked for spelling errors.

1.1.2. The application shall load the dictionary from the dictionary file. The application shall accept only one word per line in the dictionary.

1.1.3. The application shall store the dictionary words in smallcase letters with leading and trailing white spaces removed. The text shall also be converted to trimmed lowercase words before spell checking.

1.1.4. Every word in the text shall be checked for a match in the dictionary. The matched words shall generate no output.

1.1.5. Every word in the text that do not match any entry in the dictionary shall be reported on the output as a *misspelled* word.

1.1.6. For every misspelled word, a list of words shall be reported as suggested alternatives from the dictionary. For a word in the dictionary to be suggested, one of the following conditions should be met:

- The word has only one character added to the misspelled word. The added character may appear at any position in the misspelled word.
- The word has only one missing character from the misspelled word.
- The word is exactly same as the misspelled word other than any two consecutive characters that are swapped.

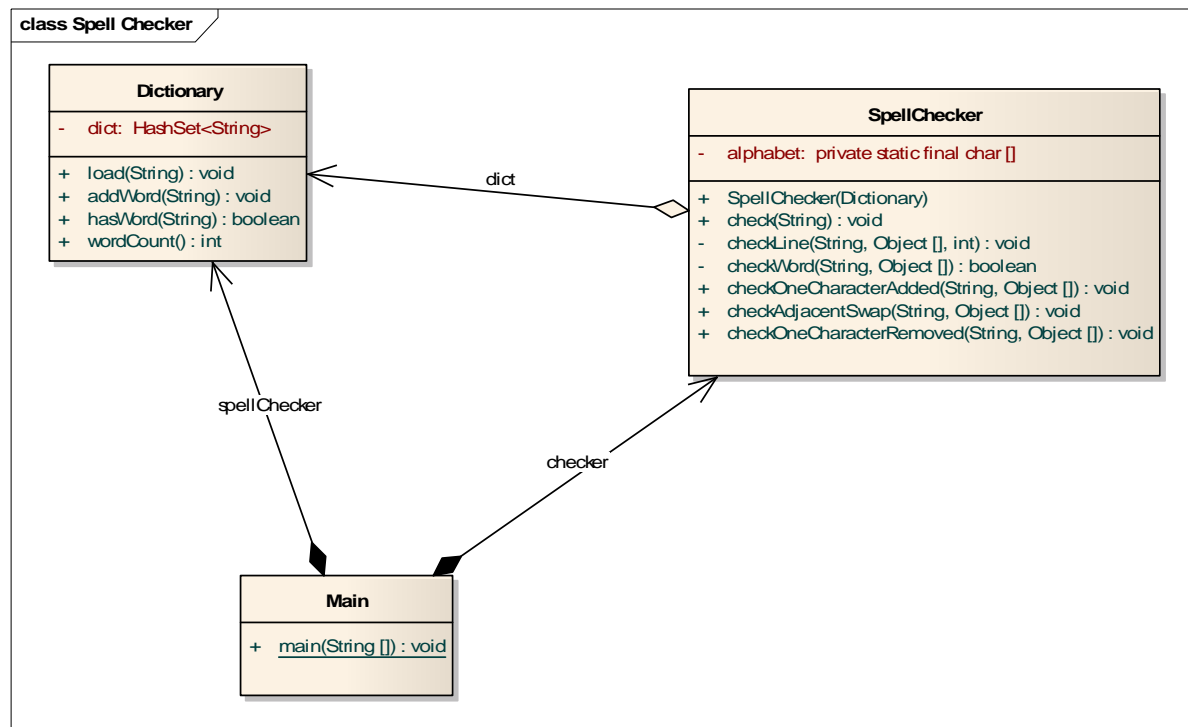
1.1.7. The line numbers of the misspelled words within the text file will also be reported on the output.

## 2. Analysis

For matching a word in a list of words, the Java class 'HashSet' is a convenient tool. This data structure is implemented using a hashmap. Hashmap implements the Set interface, so words can be added to the hashset using its 'add' method, and words can be checked within the dictionary with 'contains' method.

For words that do not match any entries in the dictionary will be checked for alternatives. The conditions for a word to be regarded as an alternative are explained in *Requirements* section 1.1.6. All of these three conditions shall be checked one by one for every misspelled word. These checks are explained in the following three paragraphs.

## 3. Design and Implementation



The spellchecker mechanism works with two classes:

The first class is the *Dictionary* class. The class loads the dictionary file and puts every word from the dictionary into a HashSet data structure. Before adding, every word is converted to small case letters and trimmed from leading and trailing whitespaces. This class presents a method named *hasWord* that accepts a String as a parameter and returns a Boolean. This method simply looks up the provided string in the HashSet. Returns true if a match is found, and false otherwise.

The second class is the *SpellChecker* class. This class has a method named *load*. This method accepts a string as a parameter. This parameter represents the name of the text file to be checked for spelling. The load method loads the text file into memory line by line. For every word in the line, a spell check is made by calling the *hasWord* function of the Dictionary.

If a word has no match in the dictionary, the following checks are made

- a. Into every position within the misspelled word, 26 characters are added one by one, and every time, the modified word is checked for a match in the dictionary. The matched words are appended to the list of alternatives. This computation requires  $(\text{number of characters in the word} + 1) * 26$  additional dictionary lookups.
- b. Every character in the misspelled word is removed one by one, and every time, the modified word is checked for a match in the dictionary. The matched words are appended to the list of alternatives. This computation requires  $(\text{number of characters in the misspelled word})$  additional dictionary lookups.
- c. For every two consecutive characters in the misspelled word, the characters are swapped and the modified word is checked for a match in the dictionary. The matched words are appended to the list of alternatives. This computation requires  $(\text{number of characters in the misspelled word} - 1)$  additional dictionary lookups.

## 4. Testing

In order to test the application, the following sources are necessary:

The sources:

- Main.java, Dictionary.java, and SpellChecker.java
- 

The dictionary:

- words.txt

The test texts:

- test1.txt, test2.txt and test3.txt

# javac Main.java

To run the application, use

# java Main <dictionary file> <test text>

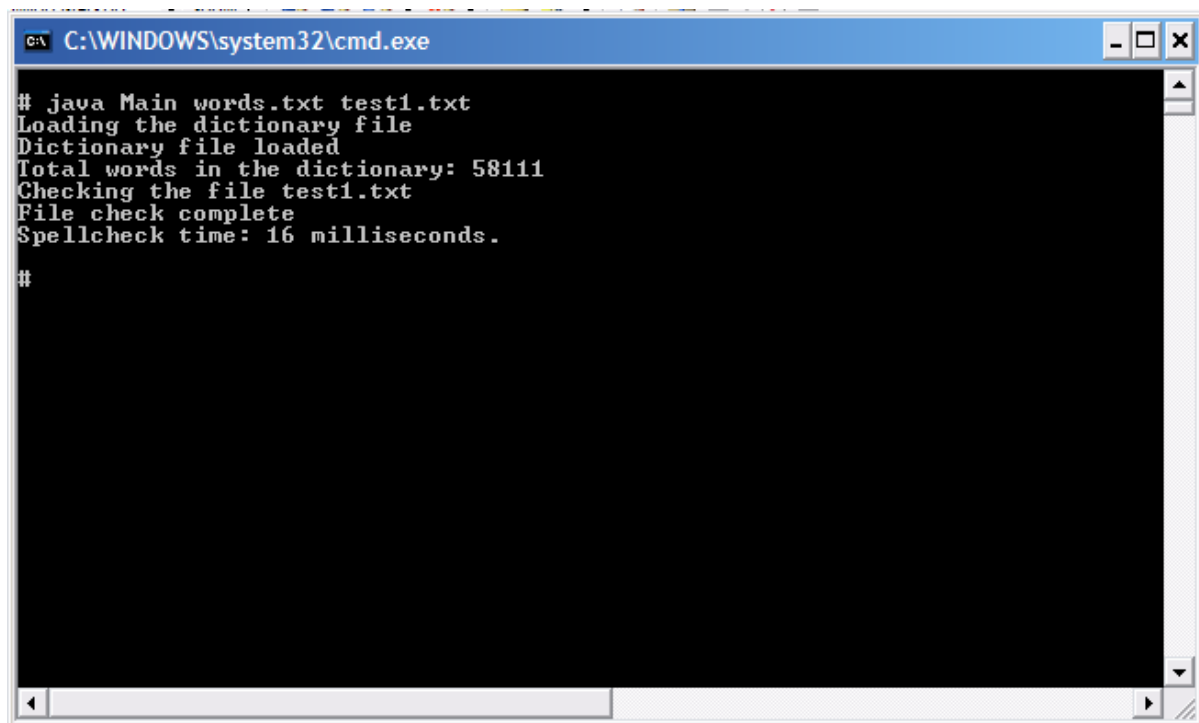
#### **4.1. Test Case 1 – No Spelling Errors**

A text file named 'test1.txt' with no spelling errors is used. The test text is:

*test file has  
no*

*misspelled words.*

The output of the program is:

A screenshot of a Windows command prompt window. The title bar reads "C:\WINDOWS\system32\cmd.exe". The command prompt shows the following text:

```
# java Main words.txt test1.txt
Loading the dictionary file
Dictionary file loaded
Total words in the dictionary: 58111
Checking the file test1.txt
File check complete
Spellcheck time: 16 milliseconds.
#
```

## 4.2. Test Case 2 – Three spelling errors with each kind.

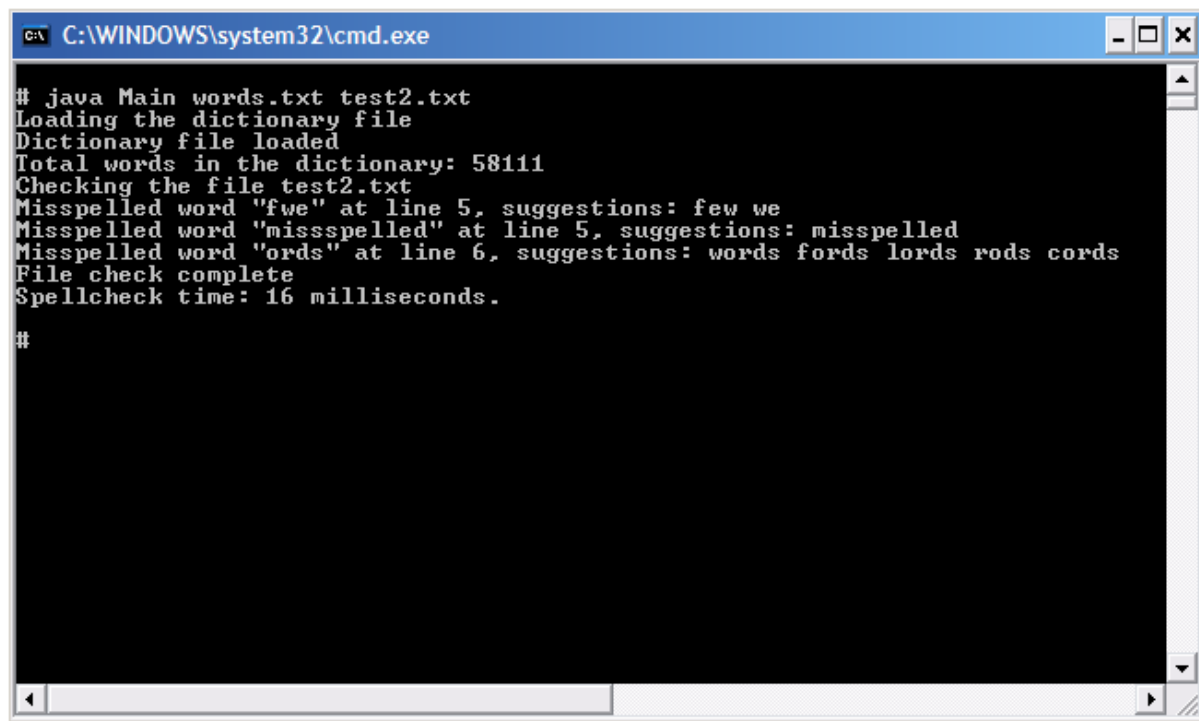
A text file named 'test2.txt' with spelling errors of all three kinds explained in requirements section 1.1.6 is used. The test text is:

*test file has*

*a*

*fwe missspelled*  
*ords.*

The output of the program is:



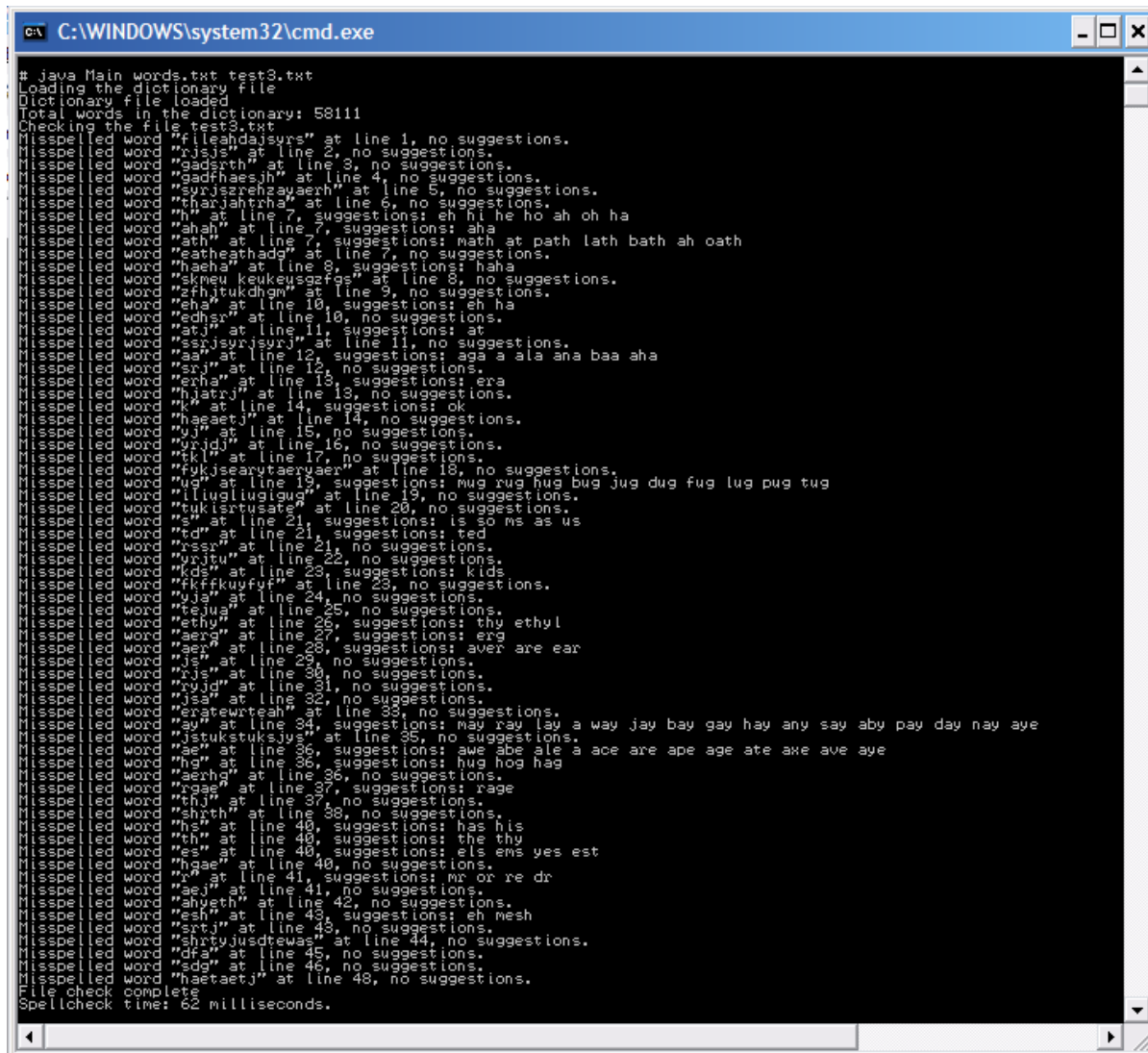
```
C:\WINDOWS\system32\cmd.exe

# java Main words.txt test2.txt
Loading the dictionary file
Dictionary file loaded
Total words in the dictionary: 58111
Checking the file test2.txt
Misspelled word "fwe" at line 5, suggestions: few we
Misspelled word "missspelled" at line 5, suggestions: misspelled
Misspelled word "ords" at line 6, suggestions: words fords lords rods cords
File check complete
Spellcheck time: 16 milliseconds.

#
```

### 4.3. Test Case 3 – Random characters on 50 lines of text

A text file named 'test3.txt' with 50 lines of random characters is used for testing. The output is shown below:



```
C:\WINDOWS\system32\cmd.exe

# java Main words.txt test3.txt
Loading the dictionary file
Dictionary file loaded
Total words in the dictionary: 58111
Checking the file test3.txt
Misspelled word "fileahdaisyr" at line 1, no suggestions.
Misspelled word "rjsjs" at line 2, no suggestions.
Misspelled word "gadsrth" at line 3, no suggestions.
Misspelled word "gadphaesjh" at line 4, no suggestions.
Misspelled word "syrjszrehzawaerh" at line 5, no suggestions.
Misspelled word "tharjsherha" at line 6, no suggestions.
Misspelled word "h" at line 7, suggestions: eh hi he ho ah oh ha
Misspelled word "ahah" at line 7, suggestions: aha
Misspelled word "ath" at line 7, suggestions: math at path lath bath ah oath
Misspelled word "eatheathadd" at line 7, no suggestions.
Misspelled word "aethd" at line 8, suggestions: haha
Misspelled word "skneu keuksuggerd" at line 8, no suggestions.
Misspelled word "zfhjtukdhgm" at line 9, no suggestions.
Misspelled word "eha" at line 10, suggestions: eh ha
Misspelled word "edhsr" at line 10, no suggestions.
Misspelled word "arj" at line 11, suggestions: at
Misspelled word "syrjszrehzawaerh" at line 11, no suggestions.
Misspelled word "aer" at line 12, suggestions: aga a ala ana baa aha
Misspelled word "erj" at line 12, no suggestions.
Misspelled word "erha" at line 13, suggestions: era
Misspelled word "hjatry" at line 13, no suggestions.
Misspelled word "k" at line 14, suggestions: ok
Misspelled word "agaej" at line 14, no suggestions.
Misspelled word "vj" at line 15, no suggestions.
Misspelled word "vjridj" at line 16, no suggestions.
Misspelled word "tkl" at line 17, no suggestions.
Misspelled word "fukjsearytaeryaer" at line 18, no suggestions.
Misspelled word "ug" at line 18, suggestions: mug tug hug bug jug dug fug lug pug tug
Misspelled word "liugliugug" at line 19, no suggestions.
Misspelled word "tuklertusafe" at line 20, no suggestions.
Misspelled word "s" at line 21, suggestions: is so ms as us
Misspelled word "td" at line 21, suggestions: ted
Misspelled word "rsxr" at line 21, no suggestions.
Misspelled word "vjrtu" at line 22, no suggestions.
Misspelled word "kds" at line 23, suggestions: kids
Misspelled word "kfkfkuyfyf" at line 23, no suggestions.
Misspelled word "vjs" at line 24, no suggestions.
Misspelled word "tejuar" at line 25, no suggestions.
Misspelled word "ethy" at line 26, suggestions: thy ethyl
Misspelled word "aerg" at line 27, suggestions: erg
Misspelled word "aer" at line 28, suggestions: aver are ear
Misspelled word "js" at line 29, no suggestions.
Misspelled word "rjs" at line 30, no suggestions.
Misspelled word "vjridj" at line 31, no suggestions.
Misspelled word "jsa" at line 32, no suggestions.
Misspelled word "egatwreah" at line 33, no suggestions.
Misspelled word "ay" at line 34, suggestions: may ray lay a way jay bay gay hay any say aby pay day nay aye
Misspelled word "jstuktstksjys" at line 35, no suggestions.
Misspelled word "ae" at line 36, suggestions: awe abe ale a ace are ape age ate axe ave aye
Misspelled word "hg" at line 36, suggestions: hug hog hag
Misspelled word "aerha" at line 36, no suggestions.
Misspelled word "rage" at line 37, suggestions: rage
Misspelled word "thj" at line 37, no suggestions.
Misspelled word "shyth" at line 38, no suggestions.
Misspelled word "hs" at line 40, suggestions: has his
Misspelled word "th" at line 40, suggestions: the thy
Misspelled word "es" at line 40, suggestions: els ems yes est
Misspelled word "hgae" at line 40, no suggestions.
Misspelled word "r" at line 41, suggestions: mr or re dr
Misspelled word "aej" at line 41, no suggestions.
Misspelled word "ahyeth" at line 42, no suggestions.
Misspelled word "esh" at line 43, suggestions: eh mesh
Misspelled word "srtj" at line 43, no suggestions.
Misspelled word "shrtvjusdtewas" at line 44, no suggestions.
Misspelled word "dfa" at line 45, no suggestions.
Misspelled word "sdg" at line 46, no suggestions.
Misspelled word "haetaetj" at line 48, no suggestions.
File check complete
Spellcheck time: 62 milliseconds.
```