

**CHALLENGE :-5**

**PROJECT PROPOSAL: A BASIC SPELL CHECK**

**BY:**

**SHIKHA BARAD (002253634)**

**FOUAD ASSADI (002241438)**

**TO:**

**DR. LAYACHI BENTAB**

**HEAD OF COMPUTER SCIENCE DEPARTMENT**



**BISHOP'S UNIVERSITY**

**QUEBEC, CANADA**

## INTRODUCTION

Mistakes that are done by the person while typing in search box or while writing any document the red highlighted line indicates that there is mistake in the word or the sentence. So, to reduce the mistake spell check is one of the best tools on PC or laptop and on the internet.

## PROCEDURE AND SOLUTION

To solve this problem, we break it into five tasks. First is to load string pattern file where it will read the file and all the text in it. It will split the sentence to each word and count frequency using dictionary data type. Second, if the frequency. Txt file doesn't exit then it will create the frequency.txt file which contain the frequency of the words in that file. Thirdly, we are creating the Input and Output files for answers. Input.txt will contain the number of words we have to search and the words that need to be searched. And after the function that correct the words it will generate output file with will contain the correct words from the list of the words. Fourthly, making the function that get a string as its parameter and return the correct answers. If the string exists in the dictionary, then it will return the answers. And each value in the dictionary type variable must be stored by dictionary order. And it will check each word in the dictionary with the string by the following conditions:-

1. Only one letter is inserted.
2. Only one letter is deleted
3. Only one letter is replaced
4. Only two consecutive letters are changed

If any of these conditions met it will return the string . and if none of these conditions met then then t will return the string and a message "This word is not in the list". Finally, the Input and output answers. Like it will read each word in the input file. For the word, it will calculate the correct word by the function that we made. After that write the calculated word into the Output file.

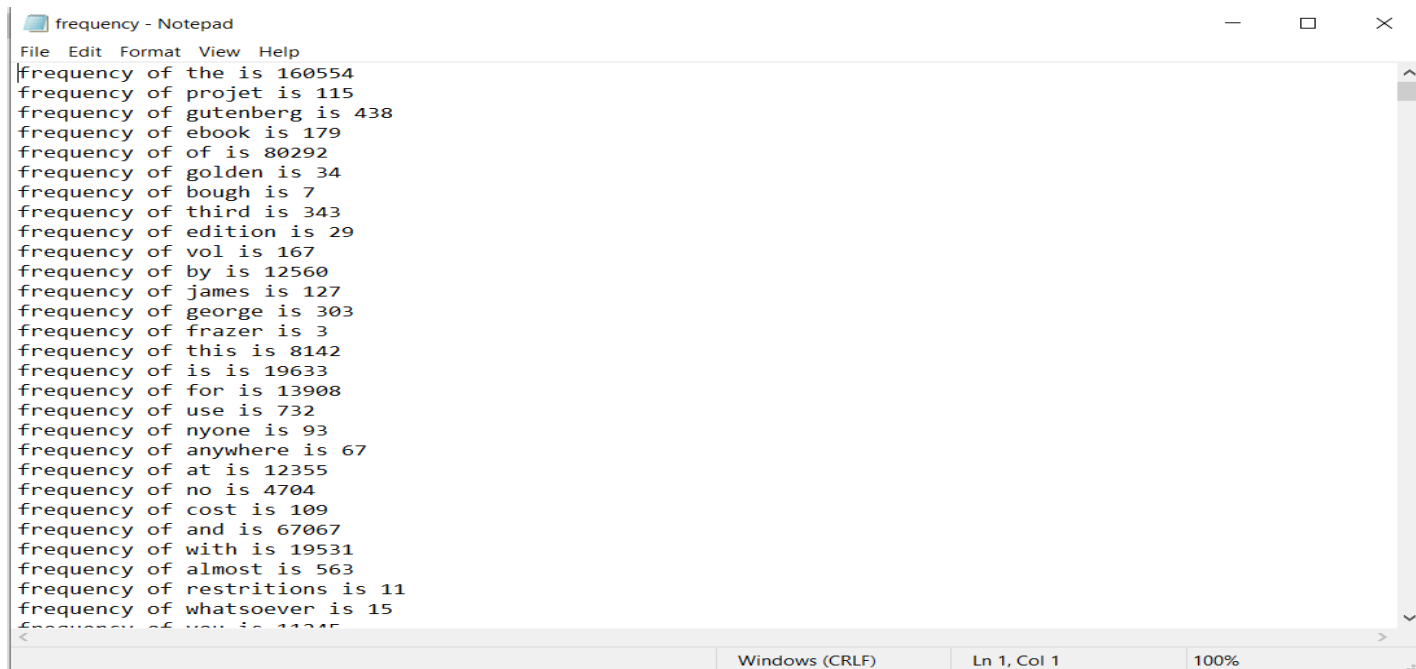
## IMPLEMENTATION

The solution of this problem is implemented using the language Python. Technique used to solve this problem is insertion , deletion, replacement of letter, and swapping the letter as mentioned in the challenge. No inbuilt spell check libraries used for implementation.

## RESULTS

Here is the input file the output file and the frequency file.

## Frequency.txt file

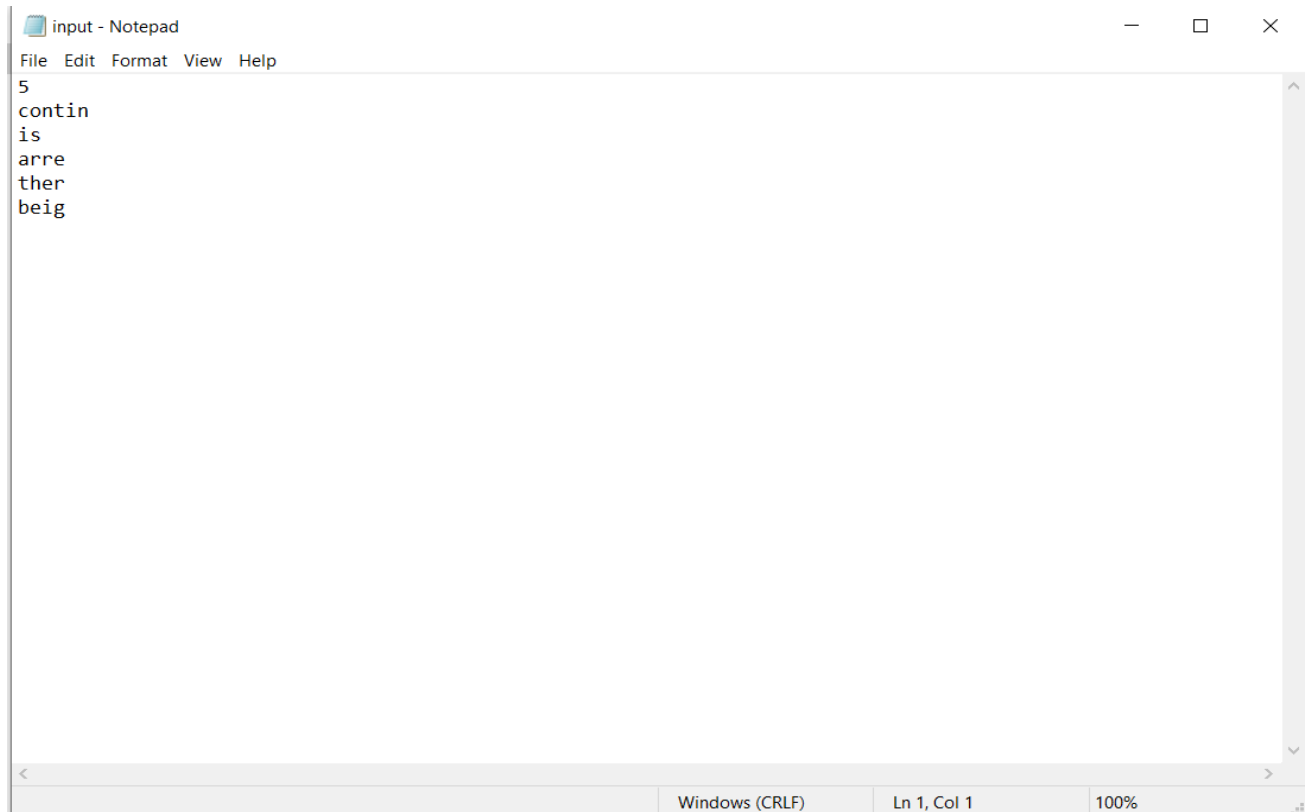


A screenshot of a Notepad window titled "frequency - Notepad". The window contains a list of words and their frequencies, each on a new line. The text is as follows:

```
frequency of the is 160554
frequency of projet is 115
frequency of gutenber is 438
frequency of ebook is 179
frequency of of is 80292
frequency of golden is 34
frequency of bough is 7
frequency of third is 343
frequency of edition is 29
frequency of vol is 167
frequency of by is 12560
frequency of james is 127
frequency of george is 303
frequency of frazer is 3
frequency of this is 8142
frequency of is is 19633
frequency of for is 13908
frequency of use is 732
frequency of nyone is 93
frequency of anywhere is 67
frequency of at is 12355
frequency of no is 4704
frequency of cost is 109
frequency of and is 67067
frequency of with is 19531
frequency of almost is 563
frequency of restritions is 11
frequency of whatsoever is 15
frequency of you is 11245
```

The status bar at the bottom indicates "Windows (CRLF)", "Ln 1, Col 1", and "100%".

## Input.txt file

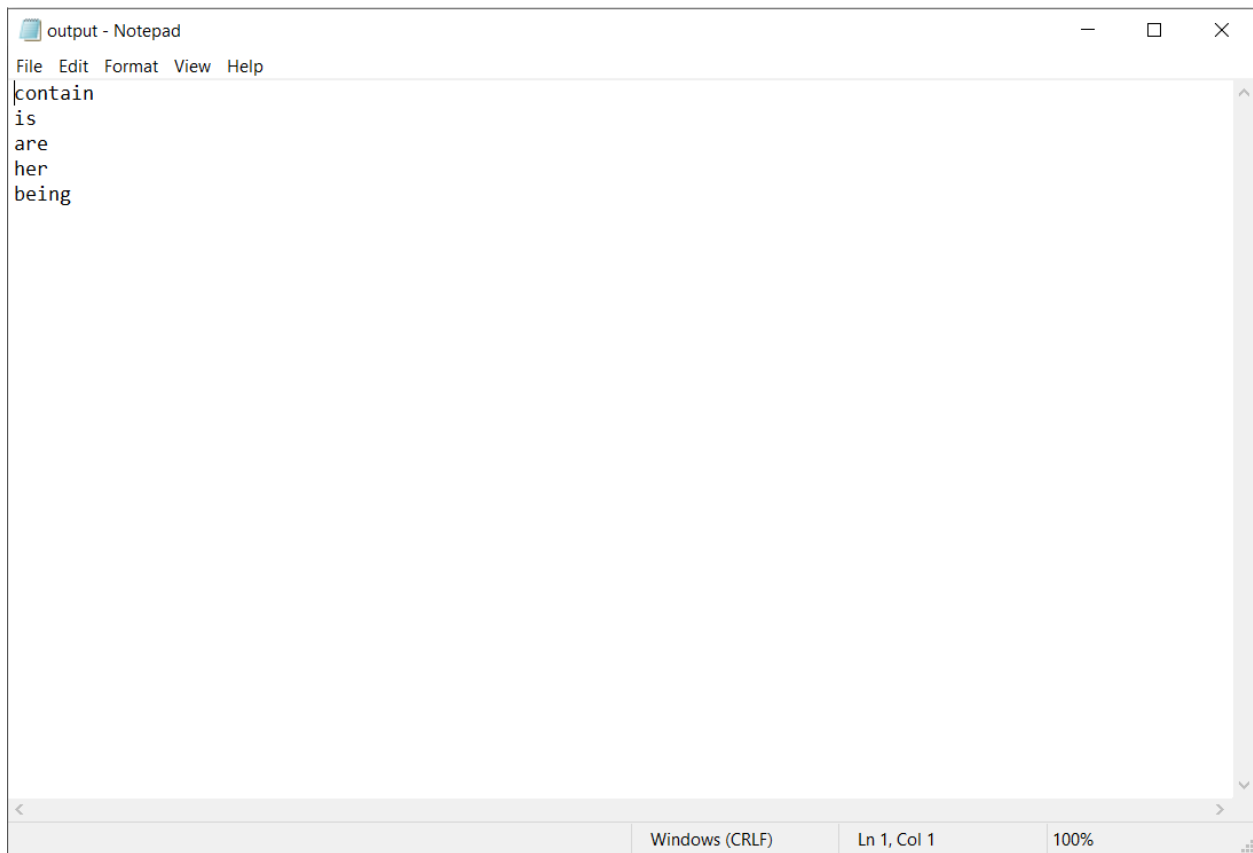


A screenshot of a Notepad window titled "input - Notepad". The window contains a list of words, each on a new line. The text is as follows:

```
5
contin
is
arre
ther
beig
```

The status bar at the bottom indicates "Windows (CRLF)", "Ln 1, Col 1", and "100%".

## Output.txt file



```
output - Notepad
File Edit Format View Help
contain
is
are
her
being
Windows (CRLF) Ln 1, Col 1 100%
```

## REFERENCE

1. Spell Checking Techniques in NLP: A Survey: by Neha Gupta and Pratistha Mathur  
<https://pdfs.semanticscholar.org/c64f/1b3a1bd7f7fe4cad469b4b94c45ad12b5d.pdf> also  
available in [www.ijarcsse.com](http://www.ijarcsse.com)
2. Daniel Jurafsky, James H. Martin, Speech and Language Processing, PEARSON, 3rd  
edition.