A BASIC SPELL CHECKER

By:

Vishwas Hasija

002251121

I. Objective:

To get accustomed with the basics of Spell Checking. Spell checker is important tool on the PC and Internet. Right from what we type in search box, or the red squiggles we see as we enter text via browser or the articles we write using an online and offline word processor spell checkers plays an important role.

II. Implementation:

Programming Language: Python 3.7

Integrated development Environment: Spyder

Libraries: Difflib and Collections

Steps followed:

1. Form_dict() is the function used to create the dictionary. Replaced all the next line character ('\n') with space (' ') and deleted simultaneous occurrences of multiple next line character. While creating dictionary, ignored special characters, numeric values and duplicate words using regex. Created dictionary keeping lexicographical order and number of occurrence in mind.
2. User_input() is the function used to get input from the user. Words are stored in list.
3. Find_nearest() function, suggest three words which are nearest to the each word.
4. These words are stored in List in sorted order.
5. Each suggested word is compared to given word on the basis of length of both the words. If the length of both the words are of equal length, than either words are equal or rearrange the given word or replace one character in given word. If given word is greater than suggested word by one character, than delete the once extra character. If given word is smaller than suggested word by one character, than add one character.
6. Delete_alpha() is the function used to delete one extra character from given word which is not in suggested word.
7. Add_alpha() is the function used to add one missing character to given word which is not in the given word and present in suggested word.
8. Replace() is the function used to replace one character which is in given word and not in suggested word.
9. Rearrange() is the function used to re- arrange the characters of the given word.
10. Equal() is the function which checks if both words are equal or not.

11. All the functions rearrage(), replace(), delete_alpha() and add_alpha() perform respective function for one time only to ensure that words are one distance away from the suggested word. All these function returns the edited word and call check().
12. Check() compares the edited word and suggested words if both are equal edited word is printed, if both words are not equal next suggested word is considered.
13. If there are no suggested word than original given word is printed.

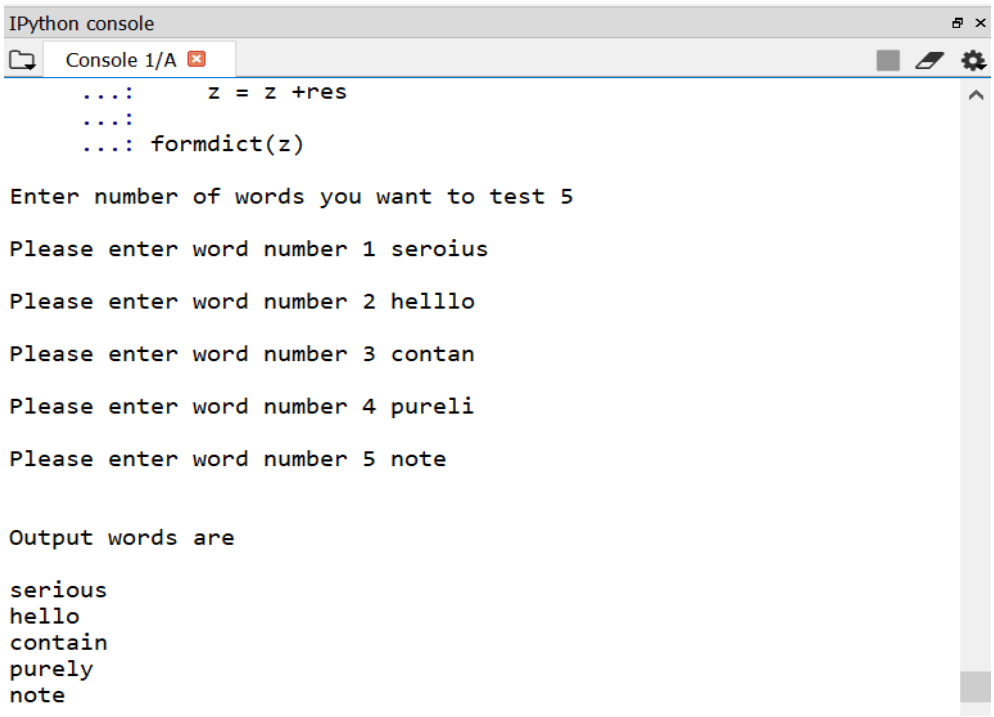
Steps to run:

1. This program can be run on any Integrated Development Environment supporting python.
2. User needs to change the location of corpus and run the program.
3. User will be prompted to enter the number of words and words.
4. After processing, suggested words will be printed for corresponding entered word.

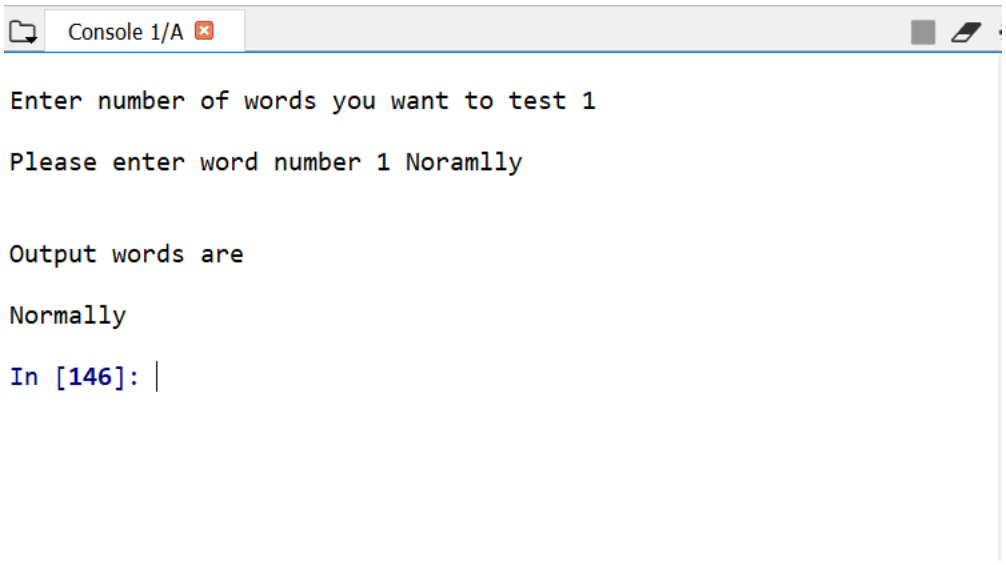
Program run and Outputs:

Program has been tested for positive and negative test cases.

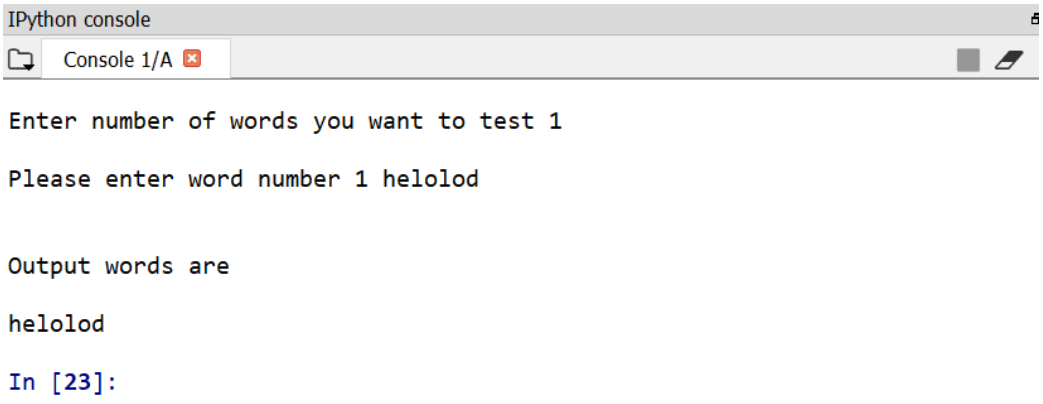
1. Program run with the given corpus.



Output

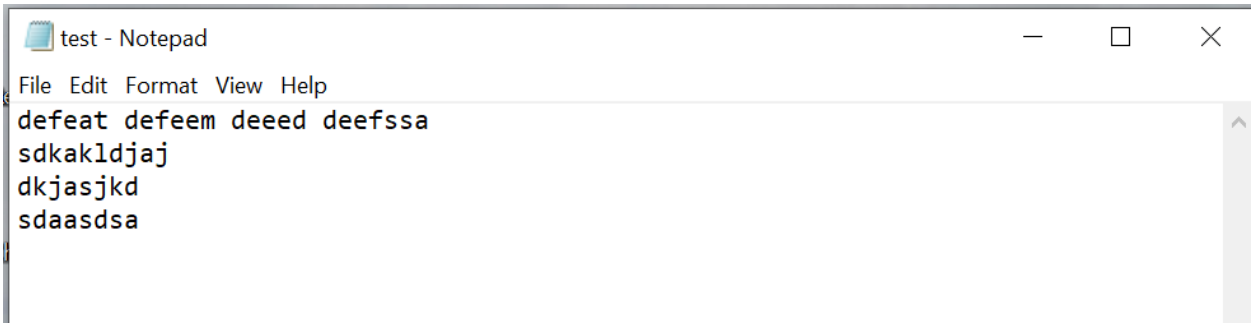


Output

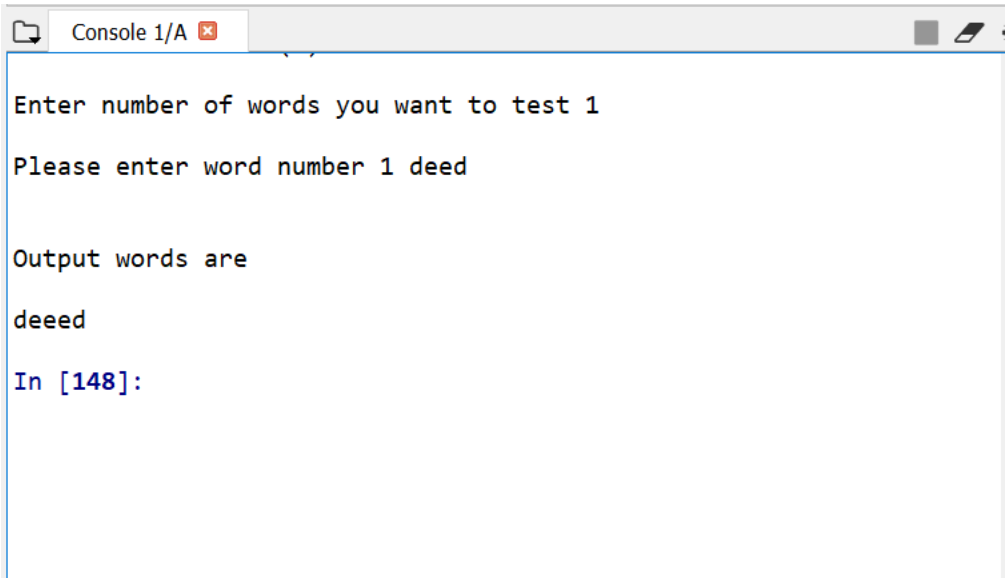


Output

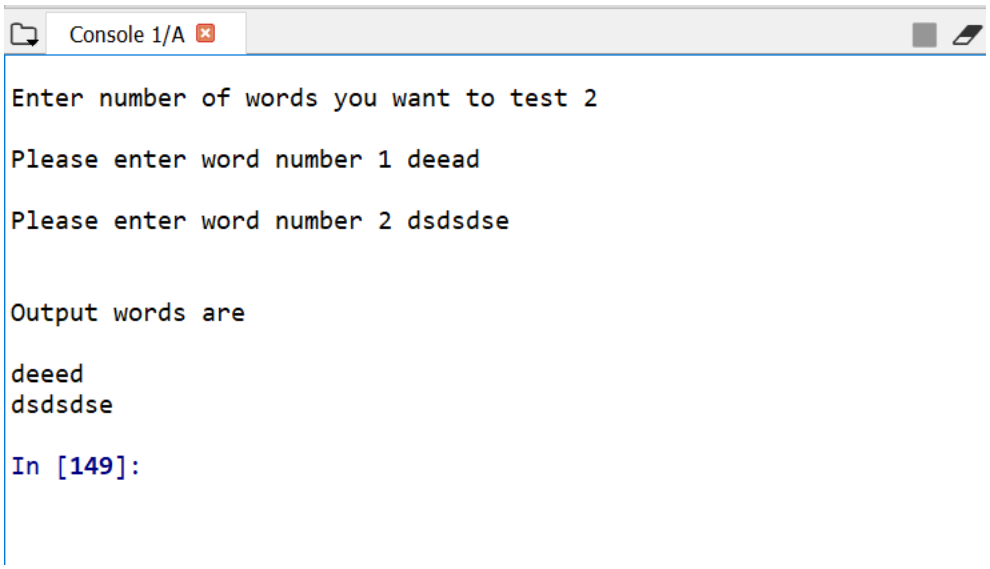
2. Program with dummy corpus and no closest match found for entered word.



Dummy Corpus



Output


Console 1/A
Enter number of words you want to test 2
Please enter word number 1 deead
Please enter word number 2 dsdsdse
Output words are
deeed
dsdsdse
In [149]:

Output