

Computer Science I

Find Text in File

CSCI-141

Homework

09/11/2016

1 Problem

We have text documents that we wish to analyze. We want to look at each line in a file and print out the lines which contain a given text string, while at the same time, counting the number of times that text string is displayed.

Task and Examples

You will write Python code to count the number of times a text string appears in a file, and display lines which contain that text string, as well as the count of the string's appearances.

Suppose the file `some.txt` contains the following:

```
words make up other words
This is a line
words make sentences
I have MS word
```

Then we have the following interaction.

```
$ python3 text_count.py
Enter text: word
Enter filename: some.txt
words make up other words
words make sentences
I have MS word
Total count of word is 4
```

We could also have another interaction.

```
$ python3 text_count.py
Enter text: words
Enter filename: some.txt
words make up other words
words make sentences
Total count of words is 3
```

There are some special cases to consider:

Suppose the file `mom.txt` contains the following.

```
testing word in word
mom should be counted twice
in momom
```

Then we have the following interaction.

```
Enter text: mom
Enter filename: mom.txt
mom should be counted twice
in momom
Total count of mom is 3
```

Notice 'mom' is counted three times. It is counted once on the second line and twice on the third line. This is due to the text 'mom' actually appearing there twice. Your solution should handle this situation.

Another special case to look at is when you start to match a word and it ends up not being a complete match. For example, looking for 'vavoom' in 'vavavoom'. A partial match, 'vav', is found. But the matching process fails with the next character, 'a'. Make sure you aren't discarding all of the characters considered at this point, or you will fail to find the match that does occur. Your solution should count the match later on in the word.

Your solution output must match the format above.

Details

You may not use any Python functionality that we haven't discussed in class without permission.

The following limitations exist on this homework:

1. No usage of the Python keyword *in*.
2. Slicing can only be used to get the tail of the string. *i.e.* `myString[1:]`
3. Indexing of strings can only be used to get the first character of the string. *i.e.* `myString[0]`
4. Using Python lists and `split()` is not allowed, as we have not discussed them in class.
5. No usage of the Python `str` class. *i.e.* `myString.startsWith()`

You are required to write the following functions. You can add more functions, i.e helper functions, but at minimum these functions must be present.

- `def count_text_string(search_for, search_in)` - a function that takes a text to be searched for and a text to be searched in. It must use **recursion** to return the number of times that the searched-for string appears in the searched-in string.
- `def count_text_file(search_for, text_file_name)` - a function that takes in the text to be searched for and the name of the file to search for the text in. It will use a **for loop** to print the lines containing the searched for text and the count of the given letter in the file.
- `def main()` - a function that reads in a text input and a file name and uses `count_text_file` to do the work.

Grading

The submission will be graded in the following format:

1. 20 points for getting text and file name
2. 20 points for opening file and iterating over it properly
3. 30 points for printing correct lines
4. 25 points for correct count
5. 5 points for correct style

Submission

Submit the file `text_count.py`, containing the code detailed above that implements the functionality described above to the appropriate *MyCourses* dropbox.