

61FIT3SS1 Python Programming - Fall 2021 Assignment 2

For this assignment, you will write a Python program work with an input text file. Your program will read text paragraphs from the file, process the text and write the result into a JSON output file.

Question 1 (5 points):

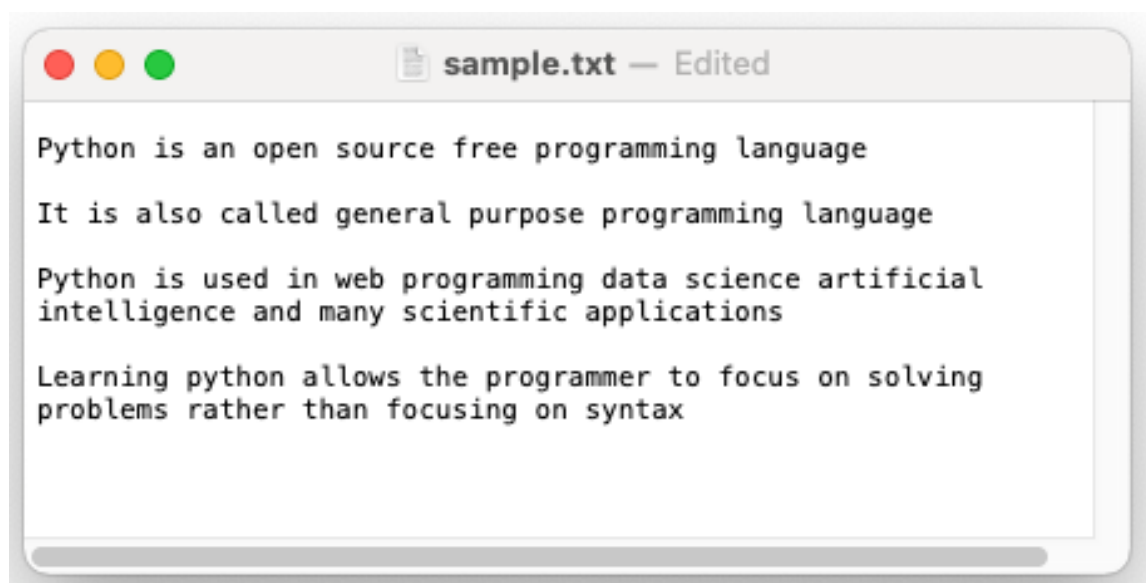
Create a function named `countWords_1()` that takes two file names as the inputs. The first one is the input file's name (a text file) and the second one is the output file's name (a JSON file). The function will count the number of times the words in the input text file occur then write the result (in JSON format) into the output file.

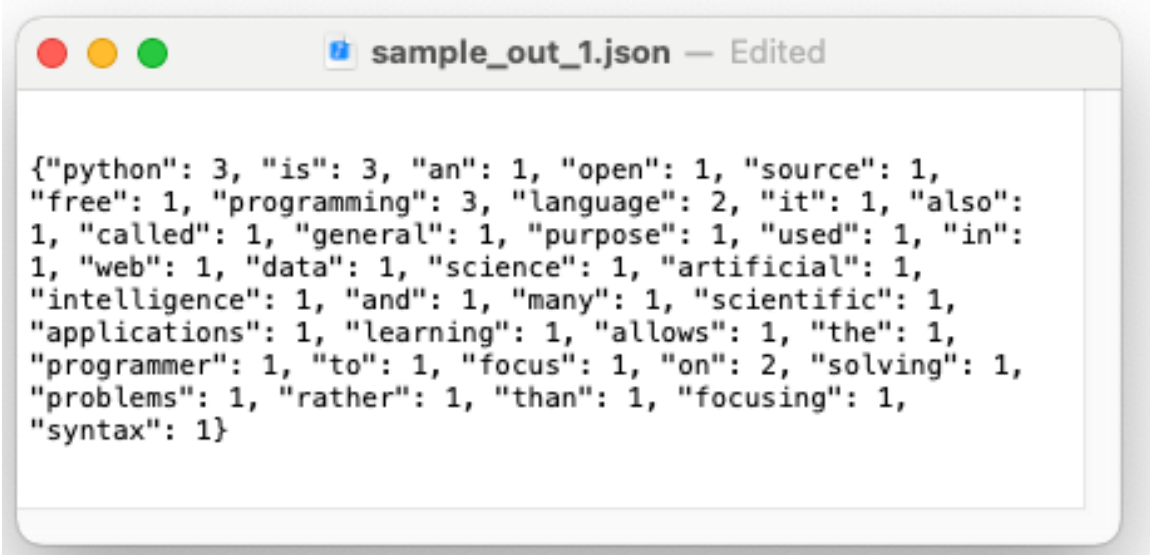
Three input text files (`sample.txt`, `sample1.txt` and `sample2.txt`) are provided for testing your function:

https://drive.google.com/drive/folders/1TAyRzGqXv-jE9RFmvSJNJdruwD21_C8i?usp=sharing

Following is an example result when executing your function with two arguments:

```
countWords_1('sample.txt', 'sample_out_1.json') :
```





```
{
  "python": 3, "is": 3, "an": 1, "open": 1, "source": 1,
  "free": 1, "programming": 3, "language": 2, "it": 1, "also":
  1, "called": 1, "general": 1, "purpose": 1, "used": 1, "in":
  1, "web": 1, "data": 1, "science": 1, "artificial": 1,
  "intelligence": 1, "and": 1, "many": 1, "scientific": 1,
  "applications": 1, "learning": 1, "allows": 1, "the": 1,
  "programmer": 1, "to": 1, "focus": 1, "on": 2, "solving": 1,
  "problems": 1, "rather": 1, "than": 1, "focusing": 1,
  "syntax": 1}
}
```

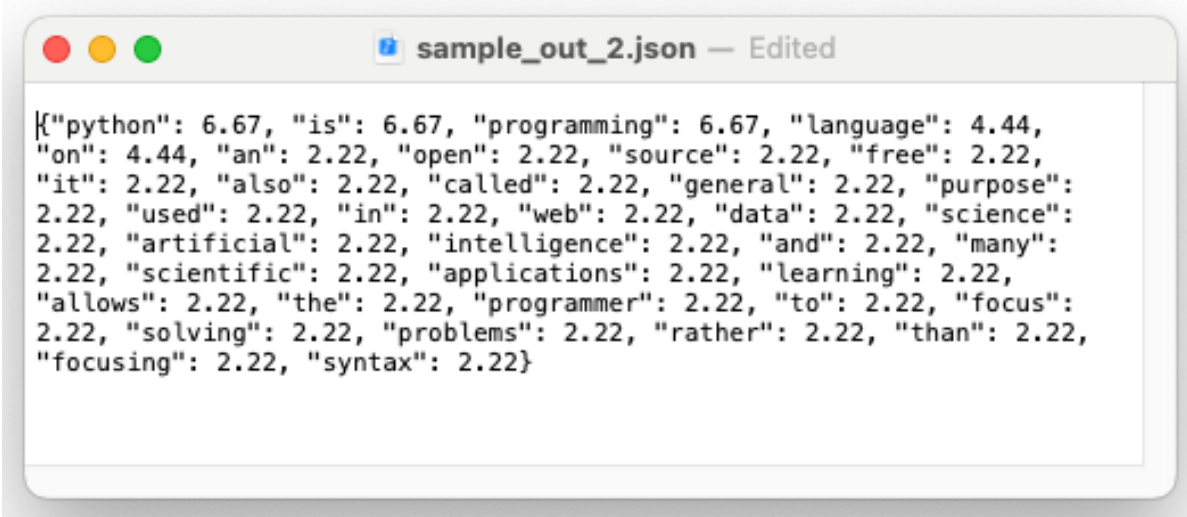
Question 2 (2 points):

Upgrade your `countWords_1()` function into `countWords_2()` function that takes the same input parameters and do the following:

- Calculate the words' frequency in percentage (with two digits in the decimal places)
- Sort the list of words based on the frequency value, from highest to lowest
- Save the result to the output file in JSON format.

Following is an example result when executing your function with two arguments:

```
countWords_2('sample.txt', 'sample_out_2.json') :
```



```
{
  "python": 6.67, "is": 6.67, "programming": 6.67, "language": 4.44,
  "on": 4.44, "an": 2.22, "open": 2.22, "source": 2.22, "free": 2.22,
  "it": 2.22, "also": 2.22, "called": 2.22, "general": 2.22, "purpose":
  2.22, "used": 2.22, "in": 2.22, "web": 2.22, "data": 2.22, "science":
  2.22, "artificial": 2.22, "intelligence": 2.22, "and": 2.22, "many":
  2.22, "scientific": 2.22, "applications": 2.22, "learning": 2.22,
  "allows": 2.22, "the": 2.22, "programmer": 2.22, "to": 2.22, "focus":
  2.22, "solving": 2.22, "problems": 2.22, "rather": 2.22, "than": 2.22,
  "focusing": 2.22, "syntax": 2.22}
}
```

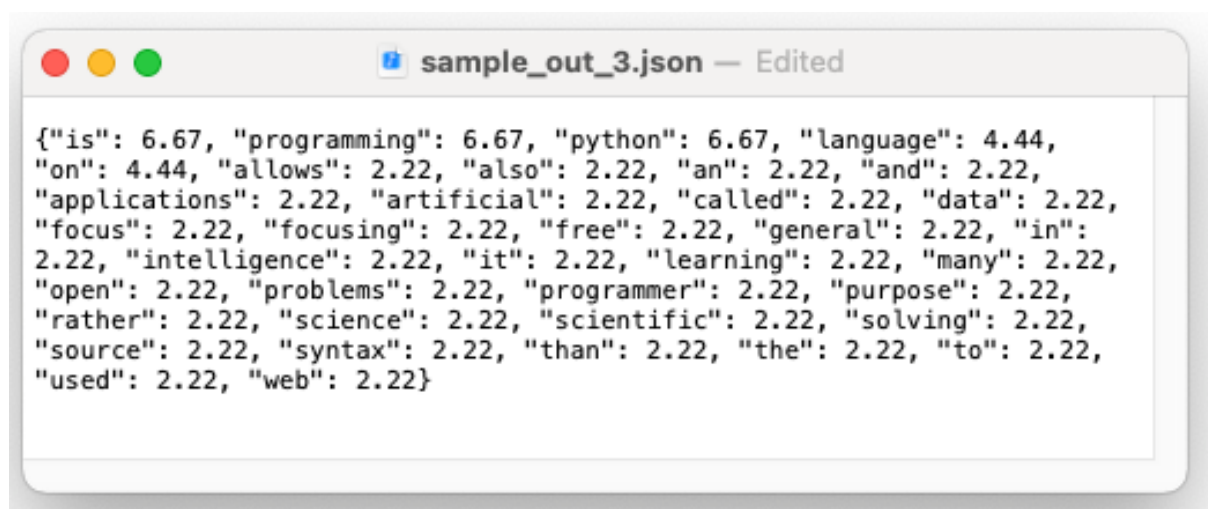
Question 3 (3 points):

Upgrade your `countWords_2()` function into `countWords_3()` function that takes the same input parameters and do the following:

- Calculate the words' frequency in percentage (with two digits in the decimal places)
- Sort the list of words based on the frequency value, from highest to lowest. Any words that have equal frequency values will be arranged in alphabetical order.
- Save the result to the output file in JSON format.

Following is an example result when executing your function with two arguments:

```
countWords_3('sample.txt', 'sample_out_3.json') :
```



Submission Instruction

Please create a new folder named `61fit3ss1_assignment2_<your student ID>`. In the folder, create the following files:

- `question1.py` : contains your program code for Question 1
- `question2.py`: contains your program code Question 2
- `question3.py`: contains your program code for Question 3

Please compress your folder into a `.zip` file and submit this file via MS Teams.