

Note

You need to use Python dictionaries to accomplish the following tasks. You have been hired to help analyze movie and sitcom scripts. Before any episode is filmed, the production crew needs you to provide the number of occurrences of each word in the script. You have been given two text files of a sitcom script to help you test out your program logic. They are `script01.txt` and `script02.txt`.

Part A

Character count: Read both files - `script01.txt` and `script02.txt`. Output the character frequency pair based on both. In doing so you need to account for letters of the alphabet only (i.e. a thru z) and by ignoring case (i.e. treat A and a, B as b, etc.). The output should be formatted as follows. Note that counts are right justified:

```
a  26
b   2
c 101
etc.
```

Name your main file as `A08_Gwida.py`. So if your GWID is G19860011 then you should name your submission file as `A08_Gwida.py`. You should also have a module file called `moduleA.py`. This module file should contain three functions `get_text(file_name)`, `process_data(text_data)`, and `print_output(data_dictionary)`. Please make sure that you comment your Python code.

Part B

Word count: Read both files - `script01.txt` and `script02.txt`. Output the word frequency pair based on both. The output should be formatted as follows (note there is a single space between the largest word and the largest count):

```
Are           26
Hearsay       2
Stanley      123
Interesting    1
etc.
```

Name your file as `A08_Gwidb.py`. So if your GWID is G19860011 then you should name your submission file as `A08_Gwidb.py`. You should also have a module file called `moduleB.py`. This module file should contain three functions `get_text(file_name)`,

`process_data(text_data)`, and `print_output(data_dictionary)`. Please make sure that you comment your Python code. Please make sure that you comment your Python code.

Part C

Word count: Find the ten most frequently occurring word frequency pair for the file named `script01.txt`. Find the counts for those 10 words for the second input file (`script02.txt`). This time use the wordlist in `stopwords.csv` to eliminate from your counts any word that belongs to that list. Your output should be formatted as follows:

WORD	Count 1	Count 2
Are	26	34
Hearsay	2	0
Stanley	123	121
Interesting	1	2
etc.		

Name your file as `A09_Gwidc.py`. So if your GWID is G19860011 then you should name your submission file as `A09_Gwidc.py`. You should also have a module file called `moduleC.py`. This module file should contain three functions `get_text(file_name)`, `process_data(text_data)`, and `print_output(data_dictionary)`. Please make sure that you comment your Python code. Please make sure that you comment your Python code.

For parts A, B and C, I should be able to run your program from the command line using the following command line syntax:

```
python A09_Gwida.py
python A09_Gwidb.py
python A09_Gwidc.py
```