

CS Homework #22: Working with files and text

Deadline: 4/4/2020, 9:00 pm.

*Save your code as `lastname_homework22.py` and submit on **GOOGLE CLASSROOM***

Task 1

Go to Project Gutenberg top 100 books page: <https://www.gutenberg.org/browse/scores/top>
Choose a book (perhaps, a book that you know or already read). Click on the link.
Choose "Plain Text UTF-8" format and save the file to your Python working directory. One way to do that is to right-click on the link and choose "Save link as". (Or use a method that works best for you).

Task 2

Open the file in Python. Make sure to implement proper error handling and use "with open" approach similar to what we did in class.

Task 3

Read the first 1000 characters from the book and print on the screen.

Task 4

Next, read the whole book and display the total number of characters, words, and sentences (assume that all sentences are separated by periods/dots). What is the average number of characters in a word in this book? What is the average number of words in a sentence?

Task 5

Is this a "happy" book or a "sad" book? Let's find out how many times the words "happy" and "sad" appear in the book. Use the string method `count()` to find out. Make sure to count both "sad"/"Sad" and "happy"/"Happy".

(In "Hamlet", for example, the word "sad" appears 10 times while the word "happy" appears only 4 times ("happy" 3 times, "Happy" once).

(Optional/Challenge) Task 6

Compare two approaches:

- (1) counting the number of times a specific word appears in the string, which contains all of the book's text (as in Task 5).
- (2) counting the number of times a specific word appears in the LIST, which contains all words from the book's text (in other words, after you apply `split()` method).

Why is there a difference? Is there a way to fix it?