

CS 101A Homework 24

Due 4/17/2020, 9:00 pm.

Save your code as lastname-homework24.py and submit on Google Classroom

Task 1

Similar to the previous homework, download a book of your choice (as a text file) and open in Python (if you have already done that, feel free to use the same book as in the previous homework).

Makes sure to implement proper error handling and use “with open” approach when loading the text.

Task 2

Remove all punctuation from your text. Make all words lower case. Generate a list of all words from the book’s text similar to how we did it in class using .split() method.

Task 3

Count the number of times the word “sad” appears in htext (the whole book as a single string variable), using .count() method.

Count the number of times the word “sad” appears in the list from task 2.

Is there a difference? Identify the words that you incorrectly included using .count() – see the classwork code if you need a reminder how to do that.

Task 4

Using collections.Counter, find and print 20 most frequent words in the text. Can you print each word on a new line with the number of times it is used next to it?

Task 5

Create a list (or a dictionary) of 10 most frequent words that are 5 or more characters long. Print your top 10 (one word per line with the number of times it is used next to it).