

Contents

- 1 Due Time
- 2 Topics
- 3 Overview
- 4 Phase 2: Teacher Lied
- 5 Sample Run
- 6 Rubric

Navigation

[Home](#)

[Information](#)

[Syllabus](#)
[Schedule](#)
[Lecture Archive](#)

[Classwork](#)

[Labs](#)
[Submitting Work](#)

Due Time

This lab is due one week from your lab's start time.

Topics

- Dictionaries
- Sets
- String manipulation

Please refer to my Moving from C++ to Python (<https://docs.google.com/document/d/1BYeChQ5Tt9qxm5uY5CzJ4H5nTY6kafobsGCU7Y984GA/edit?usp=sharing>) document for all reading and example code.

Overview

```
"If his chest were a cannon, he would have fired his heart upon it"
-Captain Pickard, Star Trek: First Contact

"Actually, I never read it."
-Lily Sloane, Star Trek: First Contact
```

In this lab, you will do some processing on a file that contains all, yes all, of the text from Moby Dick. (<https://people.eecs.ku.edu/~jwgibbo/eecs268/2022spring/labs/lab02/mobydick.txt>)

In this program it will be helpful to define the following functions:

- build_count(text)
 - returns a dictionary with every word in the text and the count of how many times that words appeared in the text
 - words should be "cleaned" (see function *clean_word*)
- clean_word(word)
 - returns the word with the following changes:
 - converted to all lower-case
 - all occurrences of '!', '?', ':', ';', ',', '|', '.', and '--' (those are two dashes side-by-side) are removed
 - all leading and trailing whitespace is removed
- unique_words(word_counts)
 - takes in a dictionary of words and their counts and returns a list (not a dict!) of all words that have a count of exactly 1
- main()
 - In charge of printing a welcome message
 - Obtains file name of input file (e.g. input.txt) from user
 - Creates the following output files:
 - word_data.txt** - contains the words and their corresponding counts
 - unique_words.txt** - contains a listing (no counts) of all words that appear exactly once

NOTES:

- Output does NOT need to be sorted (either alphabetically or by count)
- There are some trickier to deal with words that I'm not holding you responsible for handing, such as words that are surrounded with single quotes.

Phase 2: Teacher Lied

You know how I just said you don't have to handle words that begin and end with single quotes?

You have to detect any words that **begin and end** with single quotes and remove the quotes from the words.

NOTE: You **must not** remove single quotes from words that are using them as contractions.

Sample Run

```
=====Welcome to the word counter!=====
Enter a file name: mobydick.txt
The file mobydick.txt has been processed.
Output stored in word_data.txt and unique_words.txt
Exiting...
```

Rubric

268

- [25pts] build_counts
- [15pts] clean_word
- [15pts] unique_words
- [15pts] main
- [15pts] Correctly handles words that begin and end with single quotes
- [10pts] Formatting of output files
- [5pts] Documentation

Only put your .py file and file containing Moby Dick in your submission

Designed by: John Gibbons

Retrieved from "https://wiki.ittc.ku.edu/ittc_wiki/index.php?title=EECS268:Lab2&oldid=23712"