

Extra Credit Homework – “Language shapes the way we think, and determines what we can think about.” – Benjamin Lee Whorf

DUE: December 3 by 11:59:59 PM

Assigned: November 19

Background

This program will open a file and read its contents, then output: a word count, the longest word, and the average word length. It will be written in python.

Assignment Requirements

- The name of your source code file shall be `word_count_ec.py`
- The program will not take any input from the keyboard
- The file to be read will be specified when the program is run
- Words will be split by any character in the following string: `" \t\n\r.!?,\\"`
 - NOTE: You may need to delete and retype the double quote that is intended to be a character that we break on
- Words will be have its ends stripped of the following characters: `"!?,.-;:'()[]{}@#$$%^&* _+</>`~ \t\n\r"`
 - If one of these characters appear in the middle of a word, like the apostrophe in a word like “don’t”, it must be left alone
 - Numbers will not be tested
 - Empty strings must not be counted
- The word count is a whole number
- The longest word must be enclosed in double-quotation marks
- The average word length must always show one decimal place
- Input files are provided
 - Quote attribution for the file input.LONG: Edgar Allan Poe
 - The input you will be graded against will not be provided for this assignment
- Only submit your source code

Sample Run

This sample run uses the file `input.SAMPLE` (quote attribution: Frank Herbert)

```
> python3 word_count_ec.py input.SAMPLE
There were 61 words read.
The longest word is "obliteration"
The average word length is 3.9 letters.
```

```
> python3 word_count_ec.py input.LONG
There were 299 words read.
The longest word is "darling-my"
The average word length is 3.8 letters.
```

Hints

- Your longest word may be different; that's acceptable. There are multiple words that can be the longest, and which word you end up selecting depends on your method of identification
- Translating code from one language to another is not a task done line by line. It is performed action by action. Because this is a python assignment, you need to be in a python mindset to successfully complete it
- Unlike many other homework assignments, writing our own functions is not really necessary this time around
 - Python provides a lot of the functionality we need out-of-the box for this assignment
 - As a measure, my solution code is ~21 lines
- Python imports that might prove useful:
 - argparse (do the tutorial before reading the documentation)
 - re (string splitting using multiple delimiters requires regular expressions)
 - If I don't cover this in class, remind me
- Python functions that might prove useful:
 - range()
 - len()
 - sorted() or list.sort()
- Python principles that might prove useful:
 - List comprehensions
 - Format strings, output formatting mini-language
 - Indexing

Reminders

- Be sure to include a comment block at the top of every file with the required information
 - Refer to the General Homework Requirements handout on Blackboard
- Provide meaningful comments
 - If you think a comment is redundant, it probably is
 - If you think a comment is helpful, it probably is
 - Remember that you are writing comments for other programmers, not people who know nothing (obligatory Jon Snow) about coding
 - Comments are more helpful when they explain why, not what or how
- There will be no extensions

Preparing and Submitting

- Your code must be able to compile and run on the EECS lab machines
 - You are responsible for testing your code
 - "But it runs fine on my machine!" will **not** earn you any points

- Submit **ONLY** your source code file
- Homework submission will be handled exclusively through Blackboard