

L435/L555
Fall 2016

In-class Exercise #5

- 1) Download a book from Project Gutenberg. Write a program to read in that file and count: a. the number of lines, b. the number of whitespace-delimited words, and c. the number of total characters (i.e., the accumulation of line lengths), and. For word counting (b), use the command `.split()` to convert a string to a list. Output your number in the same style as the output of the unix command `wc`, and compare the outputs.
- 2) Instead of hard-coding the name of your book file, extend your program so that it accepts the name of a file from the command line, opens that file, and calculates the appropriate numbers (as in #1). Can you also run this program in a pipeline?
- 3) Take the book from Project Gutenberg and make a few changes by hand, so that you have two different versions. Now, write a python program that operates somewhat like the unix command `diff`, reading in two files (given as arguments) and outputs (at minimum) the line numbers where the files differ. Assume that the files are of the same length and the only differences are restricted to the content of individual lines (i.e., you don't have to worry about identical lines that are offset from each other.) Try to get the output to match `diff`.

Note that, at some point, you may find #3 useful, when you start to have a specific notion of what makes files different, and you want to ignore some types of differences.