

CS 115 - Introduction to Programming in Python

Lab 04

Lab Objectives: Files, Modules, Functions

1. Write a module, Lab04_Module.py, that contains the following functions:
 - `clean_data()` – takes a string as a parameter and does the following:
 - Replace all newline characters with a single space.
 - Replace punctuation (. , ?) with an empty string.
 - Converts the string to lowercase.
 - Returns the updated version of the string.
 - `find_unique()` – takes a string as a parameter, and returns a string containing each unique word in the string (separated by spaces) and the count of unique words in the string.
 - `count_unique()` – takes a string file name as a parameter. Each line in the file contains a set of unique words, and the count of the unique words as the last value on each line. The functions should sum the values at the end of each line. You may assume that the value is preceded by a comma.
2. Write a program that uses the functions in the module defined above and does the following:
 - a. Using a loop reads the data from 3 input files (words1.txt, words2.txt, words3.txt).
 - b. Cleans the data in each file.
 - c. Finds the unique words and their count for each file, and writes to an output file.
 - d. Using the output file from step c, displays the sum of all unique words from the files.

Sample Input Files:

Sample Run:

Number of unique words in files: 10