HOMEWORK 3

DUE SEPT 27th 5.00pm 2018

(This HW covers Part 1 – Chapter 2 through 8)

(A single .PY file maybe submitted if that makes it easier for you to write. Ensure that each HW problem is solved within a contained function. If you have multiple .py files, zip them up before uploading to wolfware. Happy Coding!)

1. Write a program that removes all empty lines from the attached text file "1-SampleText.txt' and copy the contents to a new file without the empty lines.

2. Write a function CodeWord(filename) that reads a file '2_SampleFile.txt' containing a large amount of text. For every instance of the word – CNC contained in the file, replace this word with the word – 'XYZ'. Output this new text to another file – "CodedText.txt". Also output the number of times the word – CNC appeared in the file to the console screen.

3. Open the .zip file- "3_Mfg_ZipFiles.zip", containing a bunch of text files. Write a program to read each text files contained in the .zip file and count the occurrences of all words that appear in every text file. Words are not case sensitive. For example if the word 'name' appears in any file, count how many times the word 'name' or 'NAME' or 'Name' appears in all files.

   As the output of the program, create a file that contains the word list obtained from the files and the corresponding number of times the word appears across all files. Ensure that the output file lists the words alphabetically sorted in ascending order. No words must be repeat twice. The list must contain a unique set of word list. In addition, do not count any word that appears in the "3-stopwords.txt" file.

   A sample console screen output is shown:
   Input:
   Enter name of .zip file
   Reading 3-stopwords.txt file…

   Output:
   6 files detected.
   Reading 6 files…
   Counting words….
   X unique words detected.
   Output file written to C:\desktop\HW3\Starly_HW3_outputfreq.txt

4. Open and read the contents of the file – '4_NoofParts_assem.txt'. Perform the following:
   a) Calculate how many entries are available in that file excluding the header.
   b) Calculate the sum of all parts from each file. Essentially, finding the sum of all values contained in the 2nd column of the file.
   c) Extract the part ID that has the largest associated no. of parts from the entire list.

5. Open the files contained in the .zip file – "5_Jobs_Completed_log.zip". Scan the files for the line that starts with the word string – "Jobs Completed.. ". Extract the number associated with this line and for all instances that this word string appears across all log files, count the total sum across all files within the .zip file. For example:

    Jobs Completed.. 10 2018-09-04 08:21:28.503153

    Extract the number 10 from this sentence which signifies the total number of jobs completed at the point in time. Find for all instances in which the string - 'Jobs Completed' appears, find the total number of jobs completed from across the provided log files.

6. For the same files above, calculate the following:
    a) How many jobs were completed in the period between 15th Aug 2018 to 15th Sept 2018?
    b) The day of the year in which the maximum number of jobs was completed.
    c) Calculate the total number of days elapsed between the time at which the first job batch was completed to the last batch completed.

7. Read the following (7_Part1.stl) triangular mesh file available in .ASCII format. For those of you aware of 3D printing, you will recognize this is as a .STL file. For those who aren't aware of what a .STL file is, please read up online on the format of a .STL ascii file. An ASCII .STL file can be opened in any text editor. Every vertex of the triangle is preceded by the word 'vertex' followed by the x, y and z coordinates in millimeters. Write functions that calculate the following:
    a. Find the total number of triangles listed in the file.
    b. Store the coordinates of each triangle in a 3-tuple list containing N indices, where N is the total number of triangles.
    c. List the area of each triangle in the file. Compute also the total area of all triangles listed in the file.

8. Open and read the contents of the file "8_DeptReportstatus.txt". In it contains recent awards made to a particular department at NCSU with various sponsor agencies, amount awarded, start date, end date and the faculty involved in every project. For every project, the faculty code that appears first is the main project leader - PL. The remaining faculty after the PL on that list are co-investigators – CoPL. From the file, find the following:
    a. Which project received the maximum award and list the faculty associated with that award?
    b. Which faculty appears the most number of times over the last 10yrs in any status – PL or CoPL?
    c. List the Faculty ID/s that appears the most number of times over the last 5yrs in the status of PL?
    d. List the Faculty ID/s that appears the least number of times over the last 5yrs in the status of PL?