Jagan Institute of Management Studies

3, Institutional Area, Sector-5, Rohini, Delhi-110085

Paper 205: Object Oriented Programming using C++
Assignment #2

Objective: To assess the performance of student candidate on CO4

Marks: 30 (team work 10, concept and solution 10, implementation 10)
Time Duration: 1 weeks

Project Teams: 2-3 members per team. Roles should be properly assigned and contribution of each member justified.

Mode of submission:
1. Written submission should include case description, objectives, data descriptions, file descriptions, source code, screenshots and conclusions
2. Soft copy submission (code and description) should be uploaded on link shared in class.

Description:

You are given a dataset of a specific datatype in a text file. Design a program in C++ which reads all data and performs a lexical analysis of each individual data unit (integers, characters, words etc). Now create separate text files for each lexical unit. Name all the files as the name of each unit respectively. Every file must have first value as the total count of that unit stored in that file. Make sure that dataset contains at least 1000 units of data and should be of same type. The program should work even if another dataset is taken as input of some other datatype.

Example: a given file "sample.txt" contains a research article of appx 5000 words
Output :
1. Words.txt: contains all the words from source file
2. Numbers.txt: contains all numeric data from source file
3. Mixed.txt: contains all words that contain numeric digits e.g. 5A, 64E, word2vec etc
4. Special.txt: contains all special symbols used in the source file
5. capitalWords.txt: contains words that were all capitalized like INDIA, IIT