Homework - 2

Q1) You are to design and implement (in any object oriented programming language, preferably Java) an email address book. (**30 points**)

Requirements:
a) Every email address must be stored along with the name of its owner.
b) An email address can be personal, such as alivelioglu@xyz.com or it can belong to a group, such as cengstudents4th@gtu.edu.tr which contains an arbitrary number of personal or group addresses.
c) Your application must be able to print on screen all the addresses in the address book. Personal addresses must be printed as : "alivelioglu@xyz.com Ali Velioglu" while group addresses must be printed by printing on screen every email (personal or group) address that they contain.
d) Implement email addresses through the **Composite** design pattern.
e) Draw a detailed class diagram of your solution.

Q2) A data mining teacher is trying to teach clustering to her students. Given a set of points, clustering is the process of grouping these points, so that similar points belong to the same groups, i.e. clustering = partitioning a set of points.

In order to achieve this, she has asked you to develop an application that demonstrates various clustering algorithms (k-means and meanshift to be exact) and for some unknown reason she has also begged you to use the **Template Method** design pattern. After careful deliberation you have decided to implement each clustering demo as follows:

K-means or Meanshift
1. Load data from file 2. Initialize the clustering parameters 3. Implement clustering 4. Show the clustering output

Your data is a text file that contains a few hundred thousand integer 3d vectors in $[0,255]^3 \cap \mathbb{Z}^3$. As output, your application should print on screen the values of the cluster centroids. e.g:
"For kmeans the clusters are : [3, 155, 78], [199, 77, 48]"
"For meanshift we have found 3 clusters: [4, 200, 21], [1, 81, 181], [44, 5, 4]"

Kmeans is a parametric clustering algorithm (https://en.wikipedia.org/wiki/K-means_clustering). It has only 2 parameters: the value of k and the distance function to use.

Meanshift is a non-parametric clustering algorithm (https://en.wikipedia.org/wiki/Mean_shift). It has a single parameter: the kernel. Use a flat kernel with a user defined radius (r=5, 10, etc) and the Euclidean distance.

Think carefully about which steps of the algorithms are common, which are specific to each algorithm and implement this application using the Template method pattern and draw its class diagram. (**70 points**)

Note: send your report (in pdf) and *.java source files (not the *.class files).