CMPE5101 Take Home Final Exam

Dr. Selim Mimaroglu
Bahçeşehir University

Due Date: January 18, 2010 7:00pm (sharp)

1 Programming Assignment

Using Java programming language, BitSet class and Jacquard-Tanimoto distance metric **you will be implementing a well known Agglomerative Hierarchical Clustering Algorithm, AGNES**. You shall use my class notes and the textbook as a reference. AGNES algorithm is well defined in my notes, therefore it is needles to repeat this definition.

Following are some rules that you **MUST** obey:

1. Use Java
2. Use BitSet class for representing bit vector objects
3. Use Jacquard-Tanimoto distance metric
4. You cannot discuss this project with your peers, this is an individual work
5. You cannot get any programming help from internet or any other source. When evaluating your project I will run *plagiarism detection* software. Any cheating attempt will be punished severely (You will get an **F** and I will report the incident to the university)

1.1 Input

Input is an ascii file named as *input.txt* residing in the same folder. In input.txt each row represents a bit vector object as follows:

0,1,1,1,1,0,0,0,1,1
0,1,1,1,1,0,0,0,0,0
0,1,1,1,1,0,0,1,1,1
1,1,1,1,1,0,0,0,1,1
0,0,0,0,0,0,0,0,0,1

In this database first object corresponds to the first row, which is 0,1,1,1,1,0,0,0,1,1

1.2 Jacquard-Tanimoto distance metric

Jacquard-Tanimoto distance metric between two bit vectors i, j is defined as follows:

$$d(i,j) = \frac{|i \oplus j|}{|i \vee j|}$$

For example, the Jacquard-Tanimoto distance metric between the first and the third objects in the database above can be found as follows:

$$\begin{aligned} i &= 0,1,1,1,1,0,0,0,1,1 \\ j &= 0,1,1,1,1,0,0,1,1,1 \\ d(i,j) &= \frac{1}{7} \\ &= 0.142857142857142857142857142857 \end{aligned}$$

1.3 Sample Run 1

Input Number of Objects:
5
Input Method (single, average, complete):
complete
Following clusters are found:
3 merged with 5
1 merged with 2
1,2 merged with 4
1,2,4 merged with 3,5

2 Submission

Submit your work electronically as an email to **smimarog@cs.umb.edu** by the due date. DO NOT FORGET TO INCLUDE SEVERAL TEST RESULTS. Subject of your email should be **Final Exam**. I will send you a confirmation when I receive your work. **If you don't receive a confirmation, it means I haven't received your submission. It is your responsibility to make sure that I got your submission in good order**.

Do not use any packages. Write a detailed **memo.txt** explaining your design, and each class. memo.txt should also include several test runs.

Make sure your attachment is virus free, otherwise I will not evaluate your submission.