J51

USN | | | | | | | | | | |

CS845

Eighth Semester B.E. Degree Examination, May/June 08
Data Mining and Warehousing

Time: 3 hrs. Max. Marks:100

Note : ***Answer any FIVE full questions.***

1 a. With a neat figure explain three-tier data warehouse architecture. **(08 Marks)**

b. Explain with example, different OLAP operations in multidimensional data model. **(08 Marks)**

c. Find number of cuboids for a data warehouse, which consists of 20 dimensions each with about five levels of granularity. **(04 Marks)**

2 a. Write a note on indexing OLAP data. Give examples. **(06 Marks)**

b. A data warehouse consists of the three dimensions time, doctor and patient and two measures count and charge, where charge is the fee that a doctor charges a patient for a visit. Perform the following :

i) Draw a star schema for above data warehouse

ii) Starting with the base cuboid [day, doctor, patient], what specific OLAP operation should be performed to list the total fee collected by each doctor in 2007? **(06 Marks)**

c. List different types of data transformation. Explain with examples variants of normalization. **(08 Marks)**

3 a. List and describe the five primitives for specifying a data mining task **(08 Marks)**

b. Provide the definition of a following hierarchy of the schema
date (day, month, quarter, year) and an
item_hierarchy (item_ID, brand, type, place_made, supplier), which is having two relations item and supplier using DMQL. **(08 Marks)**

c. Briefly define four major types of concept hierarchy. **(04 Marks)**

4 a. A database has four transactions. Let min_sup = 40% and min_conf = 60%. Find all frequent item sets using Apriori algorithm. **(10 Marks)**

TID	Date	Items.bought
T_{100}	10/15/07	{K, A, D, B}
T_{200}	10/15/07	{D, A, C, E, B}
T_{300}	10/19/07	{C, A, B, E}
T_{400}	10/22/07	{B, A, D}

b. Give and explain different approaches to mining multilevel association rules with example. **(10 Marks)**

5 a. Give and explain the algorithm for decision free induction. Write a decision tree for the concept buys_computer. **(10 Marks)**

b. What are Baysian belief networks? Explain the concept by taking example of probability of getting lung cancer by considering the family history and smoker as parent nodes. **(10 Marks)**

6 a. Enumerate and briefly explain typical requirements of clustering in data mining. **(10 Marks)**

b. Explain in detail balanced iterative reducing and clustering using hierarchies. Also draw Cluster Feature (CF) tree. **(10 Marks)**

7 a. What is conceptual clustering? Explain the concept of classification tree and draw the classification tree for animal data. **(10 Marks)**

b. What is an outlier? Explain in brief different types of outlier detection. **(10 Marks)**

8 a. What are the parameters to be considered while choosing a data mining system? Explain. **(10 Marks)**

b. What is the difference between direct query answering and Intelligent query answering? Suppose that a user requests the price, address and rating of hotels at a particular holiday location. Give examples of how these queries could be answered using both the methods. **(10 Marks)**

*****