Question Paper Code : Q 2141

B.E./B.Tech. DEGREE EXAMINATION, NOVEMBER/DECEMBER 2009.

Sixth Semester

Computer Science and Engineering

CS 1004 — DATA WAREHOUSING AND MINING

(Common to Seventh Semester Information Technology)

(Regulation 2004)

Time : Three hours   Maximum : 100 marks

Answer ALL questions.

PART A — (10 × 2 = 20 marks)

1.   List the characteristics of a Data warehouse.

2.   What is bit map indexing?

3.   What are the various data sources for data ware house?

4.   What are the types of concept hierarchies?

5.   Give the difference between operational database and informational database.

6.   Give the difference between Boolean association rule and quantitative association rule.

7.   Compare clustering and classification.

8.   What is meant by hierarchical clustering?

9.   Write down some applications of data mining.

10.   What is web usage mining?

11. (a) (i) With a neat sketch explain the architecture of Online analytical Mining. (10)

      (ii) Explain the types of OLAP server in detail. (6)

Or

    (b) (i) Discuss the various types of data warehouse schema with a suitable example. (8)

      (ii) Explain the data model which is suitable for data warehouse with examples. (8)

12. (a) (i) How can data mining task be specified? Explain. (10)

      (ii) Explain how data mining system can be integrated with database/ data warehouse system. (6)

Or

    (b) (i) Give an account of Data Mining Query Language (DMQL). (6)

      (ii) Suppose that the university course database for ABC-University includes the following attributes describing students name, address, status (under graduate or post graduate), department and CGPA (Cumulative Grade Point Average). (10)

         (1) Propose a concept hierarchy for the attributes address, status, department and CGPA.

         (2) For each concept hierarchy that have proposed above, what type of concept hierarchy is it?

         (3) Define each hierarchy using DMQL syntax.

         (4) Write a DMQL query to find the characteristics of students who have an excellent CGPA.

13. (a) (i) Write down and explain the Apriori algorithm for finding frequent item set. (10)

      (ii) Consider a database having ten transactions. Let min_sup = 20%. (6)

TID	List of items_IDs
T100	I1, I2, I5
T200	I2, I4
T300	I2, I3

TID	List of items_IDs
T400	I1, I2, I4
T500	I1, I3
T600	I2, I3
T700	I1, I3
T800	I1, I2, I3, I5
T900	I1, I2, I3
T1000	I1, I2

Find all frequent item set using Apriori algorithm.

Or

(b) (i) Write down the algorithm for mining frequent itemsets without candidate generation. Also explain the same. (10)

(ii) Draw FP tree for the database given in question (13 (a) (ii)) and find the frequent patterns. (6)

14. (a) (i) How do you estimate classifier accuracy? Discuss. (8)

(ii) With an example explain the use of tree pruning. (8)

Or

(b) (i) Give an account of the requirements of clustering algorithms. (8)

(ii) Describe the working of PAM (Partioning Around Medoids) algorithm. (8)

15. (a) What is spatial database? Explain the methods of mining spatial databases. (16)

Or

(b) Discuss the method of mining text databases in detail. (16)

————————