GANPAT UNIVERSITY
M. Tech SEMESTER-II Computer Engineering
REGULAR EXAMINATION JUNE 2012
3CE203: Data Mining & Data Warehousing

Time: 3 Hours]   [Total Marks: 70

Instructions:

1. Figures to the right indicate full marks
2. Each section should be written in a separate answer book
3. Be precise and to the point in your answer

SECTION-I

Q-1 Answer the following: [12]

(a) Suppose that a data warehouse consists of the three dimensions time, doctor, and patient and two measure count and charge where charge is the fee that a doctor charges a patient for a visit.

(1). Enumerate three classes of schemas that are popularly used for modeling data warehouses.

(2). Draw a schema diagram for the above data warehouse using one of the schema classes listed in (a).

(3). Starting with the base cuboid[day,doctor,patient] what specific OLAP operations should be performed in order to list the total fee collected by each doctor in 2012.

(4). To obtain the same list, write an SQL query assuming the data are stored in a relational database with the schema fee(day, month, year, doctor, hospital, patient, count, charge).

(b) Define Heterogeneous Databases and Legacy Databases.

Q-2 Answer the following:

(a) Discuss FP growth Tree.   [05]

(b) Discuss the application of Data Mining in Medical Gene patterns.   [06]

OR

Q-2 Answer the following:

(a) The business database is given two attributes : GDP and CPI with other   [06]
attributes. Write Pl/Sql Query to find out the support and confidence value
for GDP =2500 and CPI=25. Write all queries starting from creation of
table.

(b) What is association rule mining? Write Apriori Algorithm.   [05]

Q-3 Answer the following: [12]

(a) What is snowflake schema? Discuss it with an example. Write Sql query
for it.

(b) What is Information gain in decision tree induction? Find out Information
gain for various attributes with one sample customer database.

OR

Q-3 Answer the following: [12]
(a) Discuss classification by decision tree induction
(b) Write an algorithm of Naïve Bayesian Classification.

SECTION-II

Q-4 Answer the following: [06]
(a) Explain major challenges faced in Data Mining   [05]
(b) Explain various schemas in Data warehousing
                         OR

Q-4 Answer the following: [06]
(a) Explain Rollup, drill down and dice operation in cube.   [05]
(b) Write a code for implementation of Association Rule.

Q-5 Answer the following: [12]
(a) Discuss the methods to improve the efficiency of Apriori Algorithm.
(b) Discuss three tier architecture of Data Warehousing

Q-6 Answer the following: [12]
(a) Explain the methods for handling noisy data.
(b) Explain cuboids with one example.
                         OR

Q-6 Answer the following: [12]
(a) Give the difference between OLAP and OLTP.
(b) Explain the methods of implementation of Data warehouse.

End of Paper