

B.E/B.TECH DEGREE EXAMINATIONS MAY/JUNE-2011

REGULATIONS 2008

SIXTH SEMESTER

CS 701- DATA WAREHOUSING AND DATA MINING

INFORMATION TECHNOLOGY

TIME: Three hours

Maximum:100 marks

ANSWER ALL QUESTIONS

PART-A(10*2=20 marks)

1. State what data cleaning routines attempt to fill?
2. What is multidimensional database?
3. Classify OLAP tools?
4. What is fact constellation?
5. What is classification?
6. State why data preprocessing is an important issue for data warehousing and data mining?
7. List the two interesting measures of an association rule
8. State the need for pruning phase in decision tree construction
9. Define clustering
10. What is outlier analysis?

PART-B(5*16=80 marks)

11(a) List and discuss the characteristics and main functions performed by the components of a data warehouse. Give diagrammatic illustration? (16) **(or)**

(b)(i) Explain why a data warehouse is well equipped for providing the data for data mining?(8)

(ii) List and discuss the three important issues that have to be addressed during data integration?(8)

12(a) In data warehouse technology, a multiple dimensional view can be implemented by a relational database technique(ROLAP), or by a multidimensional database technique(MOLAP) or by a hybrid database technique(HOLAP). Describe each implementation technique(16)

or

(b) With relevant examples discuss the different OLAP operations(16)

13(a) Define each of the following data mining functionalities, characterization, discrimination, association and correlation analysis, classification, prediction, clustering and evolution analysis. Give examples of each data mining functionality, using a real-life Database that you are familiar with.(16)

or

(b) With an example how a data mining system can be integrated with a data warehouse(16)

14(a) Develop an algorithm for classification using decision trees. Illustrate the algorithm with a relevant example.(16)

or

(b) Discuss the apriori algorithm for generating association rules. Illustrate the algorithm with a relevant example.(16)

15(a) Consider five points $\{X_1, X_2, X_3, X_4, X_5\}$ with the following coordinates as a two dimensional sample for clustering

$X_1=(0,2)$

$X_2=(0,);$

$X_3=1,5,0);$

$X_4=(5,0);$

$X_5=(5,2)$

illustrate the k-Means partitioning algorithm using the above set of data set (16)

or

(b) Discuss any two applications of data mining(16)