MCIT/MCSE/MCTA-301

M. Tech. (Software System) (Third Semester)
EXAMINATION, Feb., 2010

DATA MINING

(Elective – IV)

Time : Three Hours

Maximum Marks : 100

Minimum Pass Marks : 40

Note : Attempt any *five* questions. All questions carry equal marks.

1. (a) Explain all the seven steps of KDD. 10
   (b) Explain all major issues in data mining. 10
2. (a) How is a data warehouse different from a database ? How are they similar ? 10
   (b) Explain OLAP operations in multidimensional data model. Differentiate ROLAP, MOLAP and HOLAP. 10
3. (a) How we can mine frequent itemsets without candidate generations ? 10
   (b) Why a priori algorithm is inefficient ? Explain its any *two* variations that increase the efficiency. 10
4. (a) Briefly outline the major steps of decision tree classification 10

   (b) Why is Naive Bayesian classification called "Naive" ? Briefly outline the major ideas of Naive Bayesian classification. 10
5. (a) How many types of clustering methods are there ? Explain any *one* partitioning-clustering algorithm. 10
   (b) Explain DBSCAN, density based clustering algorithm to discover cluster with arbitrary shape. 10
6. (a) Why current search engine are not sufficient for web resource discovery ? Differentiate web content mining, web structure and web usage mining. 10
   (b) What kind of analysis we can perform on web log files ? Explain with suitable examples. 10
7. (a) What is spatial data mining ? Can we create data cube on spatial data ? If so then explain spatial data cube with suitable example. 10
   (b) Explain the following with suitable example : 10
      (i) Spatial association analysis
      (ii) Spatial clustering
8. Write short notes on any *four* of the following : 20
   (i) Image and Video mining
   (ii) Text mining
   (iii) Temporal mining
   (iv) Multilevel association rule mining
   (v) Concept hierarchy