JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD
B. Tech IV Year I Semester Examinations, November - 2013
Data Warehousing and Data Mining
(Computer Science and Engineering)

Time: 3 Hours   Max. Marks: 75

Answer any Five Questions
All Questions Carry Equal Marks
- - -

1.a) Present an example where data mining is crucial to the success of a business. What *data mining functions* does this business need? Can they be performed alternatively by data query processing or simple statistical analysis?

b) *Data quality* can be assessed in terms of accuracy, completeness, and consistency. Propose two other dimensions of data quality?   [8+7]

2. a) What are the characteristics of the OLAP and the basic data warehouse environments as they relate to information delivery needs?

b) Discuss why *relevance analysis* is beneficial and how it can be performed and integrated into the characterization process. Compare the result of two induction methods: (1) with relevance analysis and (2) without relevance analysis.   [8+7]

3. How can efficiency of Apriori-based be improved? Describe briefly any of five variations of the Apriori algorithm?   [15]

4.a) Briefly outline the major steps of *decision tree classification*?

b) What is *association-based classification*? Why is association-based classification able to achieve higher classification accuracy than a classical decision-tree method? Explain how association-based classification can be used for text document classification?   [6+9]

5.a) Briefly outline how to compute the *dissimilarity* between objects described by Asymmetric binary variables?

b) Give an example of how specific clustering methods may be *integrated*, for example, where one clustering algorithm is used as a preprocessing step for another. In addition, provide reasoning on why the integration of two methods may sometimes lead to improved clustering quality and efficiency?
[6+9]

6. The concept of microclustering has been popular for on-line maintenance of clustering information for data streams. By exploring the power of microclustering, design an effective *density-based* clustering method for clustering evolving data streams.   [15]

7. Outline an implementation technique that applies a similarity-based search method to enhance the quality of clustering in multimedia data?   [15]

8. Why is the establishment of *theoretical foundations* important for data mining? Name and describe the main theoretical foundations that have been proposed for data mining. Comment on how they each satisfy (or fail to satisfy) the requirements of an ideal theoretical framework for data mining?   [15]

--ooOoo--