Total No. of Questions : 5]

SEAT No. : [ ]

P3349

[Total No. of Pages : 5

[6027]-32

Second Year M.C.A. (Management)

IT-32 : DATA WAREHOUSING AND DATA MINING

(2020 Pattern) (Semester - III)

Time : 2½ Hours] ***[Max. Marks : 50***

Instructions to the candidates:

1) ***All questions are compulsory.***
2) ***Draw neat labelled diagrams wherever necessary.***

Q1) Answer the following multiple-choice questions. **[20×0.5]**

i) Generation of concept hierarchy who's depends on?

a) Static or dynamic data set b) Only several relations

c) System call mechanism d) Basic terminology

ii) Which cache is use in connected and unconnected transformations?

a) Dynamic cache and static cache

b) Content Delivery Network cache

c) Database cache

d) Import cache & File system cache

iii) Which is the Function used for K-means clustering?

a) K-mean b) Heatmap

c) K-means d) K-medians

iv) What to use in Hierarchical clustering for finding the right number of clusters?

a) The Elbow method b) Decision Trees

c) Dendrograms d) Histograms

v) _____collects all the information and the subjects about an entire organization.

a) Data mart

b) Virtual warehouse

c) Data warehouse view

d) Enterprise warehouse

P.T.O.

vi) Which is not a valid layer in Three-layer data warehouse architecture in conceptual view?

a) Processed data layer
b) Real-time data layer
c) Derived data layer
d) Reconciled data

vii) Which is true about star schema?

a) Suited for operational data processing
b) Used in operational systems
c) Used in operational data stores
d) Used to develop data warehouses & dimensional data marts

viii) Dimensionality refers to

a) Cardinality of key values in a star schema
b) The data that describes the transactions in the fact table
c) The level of detail of data that is held in the fact table
d) The level of detail of data that is held in the dimension table

ix) Data transformation includes _____

a) A process to change data from a detailed level to a summary level
b) A process to change data from a summary level to detailed level
c) Joining data from one source into various sources of data
d) Separating data from one source into various sources of data

x) A cross tab section with one attribute is obtaining by using

a) Slice
b) Dice
c) Pivot
d) Both slice and dice

xi) What is true about data mining?

a) Data mining is defined as the procedure of extracting information from huge sets of data
b) Data mining also involves other processes such as data cleaning, data integration, data transformation
c) Data mining is the procedure of mining knowledge from data
d) All of the mentioned

xii) In a hypercube, each dimension belongs to _____only.

a) One cube
b) Multi cube
c) None of the mentioned
d) Both a and b

xiii) Parametric Data Reduction involves

a) Regression b) Classification
c) Clustering d) Association

xiv) _____is the application of data mining techniques to discover patterns from the web.

a) Text mining b) Multimedia mining
c) Web mining d) Link mining

xv) Google PR checker is a tool designed for _____

a) Page Rank b) Hits
c) Plagiarism d) Summarization

xvi) For KNN, which of the following distance measure is NOT valid for continuous variables?

a) Manhattan distance b) Hamming distance
c) Euclidean distance d) Minkowski distance

xvii) A collection of databases that offer a unified approach for organizing data and classifying data according to subject.

a) Data mart b) Enterprise data warehouse
c) Operational data store d) Data mining

xviii) _____is a base document for data extraction.

a) Logical data map b) FP Tree
c) Granularity d) Association

xix) Microsoft SQL server is an example of

a) HOLAP b) MOLAP
c) ROLAP d) SOLAP

xx) Propagate the updates from the data sources to the warehouse.

a) Data Load b) Refresh
c) Data Transformation d) Data cleaning

Q2) a) Define Data warehouse & explain architecture of data warehouse with neat diagram. **[5]**

b) Explain different data warehouse schemas. **[5]**

OR

Q2) a) Compare OLAP and OLTP. **[5]**

b) Explain kimball life cycle diagram in detail. **[5]**

Q3) a) Explain different steps involed in cleaning & transformation of data. **[5]**

b) What is text-mining? Explain applications of text-mining. **[5]**

OR

Q3) a) Brief OLAP operations with proper example. **[5]**

b) Explain data mining architecture. **[5]**

Q4) a) Apply FP-Growth tree algorithm to construct FP-Tree & find the frequent item sets. Consider following data set with minimum support 30%. **[5]**

TID	List of Items
1	Apple, Mango, Cocount, Banana
2	Banana, Berries, Coconut, Mango, Chikoo
3	Berries, Coconut, Dates, Grapes
4	Watermelon, Coconut, Mango
5	Apple, Banana, Chikoo, Dates
6	Apple, Berries
7	Apple, Coconut, Grapes, Banana
8	Mango, Watermelon, Coconut
9	Banana, Mango, Coconut

b) Consider the data set from ***Q4)*** a) and calculate support & confidence of following item sets. **[5]**

i) { Apple, Coconut }

ii) { Apple, Coconut, Banana }

iii) { Mango, Watermelon }

iv) { Coconut, Berries }

v) {Banana, Mango, Coconut }

OR

Q4) a) Consider following data set and find the frequent item sets with minimum support count 2 using Apriori Algorithm. **[5]**

TID	Items
T1	Hot Dogs, Buns, Ketchup
T2	Hot Dogs, Buns
T3	Hot Dogs, Coke, Chips
T4	Chips, Coke
T5	Chips, Ketchup
T6	Hot Dogs, Coke, Chips

b) Consider same data set in ***Q4)*** a) & generate association rules [minimum 3] **[5]**

Q5) a) Write K-mean clustering algorithm and apply on the following data set to group it into two clusters. **[5]**

D = (0.5, 0.7, 0.9, 0.6, 0.4, 0.3, 0.10, 0.12, 0.14, 0.16, 0.35, 0.36, 0.11, 0.45, 0.58)

b) How KNN algorithm work? **[5]**

OR

Q5) a) D = { 14, 13, 15, 16, 16, 19, 20, 35, 37, 39, 40, 41, 50, 56, 72, 17, 18, 12, 24, 22, 21 }

K = 3 C_1 = 13, C_2 = 20, C_3 = 50

Apply K-mean clustering. **[5]**

b) Explain Agglomeratire and divisive clustering mechanism. **[5]**