

Minor changes for MCA

MCA 4.1	DATA MINING TECHNIQUES	
Instruction: 4 periods / week		Credits: 5
Internal : 30 marks	University Exam: 70 marks	Total : 100 Marks

Objectives:

The course is designed to meet the objectives of:

1. understand classical models and algorithms in data warehousing and data mining,
2. enable students to analyse the data, identify the problems, and choose the relevant models and algorithms to apply,
3. assess the strengths and weaknesses of various methods and algorithms and to analyse their behaviour.

Outcomes:

Students upon completion of this course will be able to:

1. learn implementation of classical algorithms in data mining and data warehousing
2. learn to identify the application area of algorithms, and apply them
3. learn clustering application and recent works in data mining

SYLLABUS

UNIT – I

Introducing to Data Mining : Introduction, What is Data Mining, Definition, KDD, Challenges, Data Mining Tasks, Data Preprocessing, Data Cleaning, Missing data, Dimensionality Reduction, Feature Subset Selection, Discretization and Binaryzation, Data Transformation; Measures of Similarity and Dissimilarity – Basics.

UNIT – II

Association Rules : Problems Definition, Frequent Item Set Generation, The APRIORI Principle, Support and Confidence Measures, Association Rule Generation; APRIORI Algorithm, The Partition Algorithms, FP- Growth Algorithms, Compact Representation of Frequent Item set- Maximal Frequent Item Set, Closed Frequent Item Sets.

UNIT – III

Classification : Problem Definition, General Approaches to solving a classification problem, Evaluation of classifiers, Classification Techniques, Decision Tree – Decision tree Construction, Methods for Expressing attribute test conditions, Measures for Selecting the Best Split, Algorithm for Decision tree Induction; Naive Bayes Classifier, Bayesaian Belief Networks; K – N earnest neighbour classification – Algorithm and Characteristics.

UNIT – IV

Clustering : Problem Definition, Clustering Overview, Evaluation of Clustering Algorithms, Partitioning Clustering -K-Means Algorithm, K-Means Additional issues, PAM Algorithm; Hierarchical Clustering – Agglomerative Methods and divisive methods, Basic Agglomerative Hierarchical Clustering, Strengths ans Weakness; Outlier Detection.

Prescribed Textbooks:

- Data Mining – Concepts and Techniques – Jiawei Han, Micheline Kamber, Morgan Kaufmann Publishers, Elsevier, 2 Edition, 2006.
- Introduction to Data Mining, Pang – Ning Tan, Vipin Kumar, Michael Steinbach, Pearson Education.

Reference Books:

- Data Mining Techniques, Arun K Pujari, 3rd Edition, Universities Press.
- Data Warehouse Fundamentals, Pualraj Ponnaiah, Wiley Student Edition.
- Data Mining, Vikram Pudi, P Radha Krishna, Oxford University Press

Model Paper

MCA 4.1: Data Mining Techniques

Time: 3 Hrs

Max. Marks: 70

Answer the following questions. Each Question Carries 14 Marks.

- 1.(a) What is OLAP? Discuss about OLAP operations.
- (b) What is Multidimensional Model?
- (c) What is Generalized Rule?
- (d) Explain about different cluster categories.
- (e) How can the root of the decision tree was initialised?
- (f) What kinds of data used in Data Mining?
- (g) What is Meta data? Describe the types of Metadata.

Unit-I

- 2.(a) What is Data Mining? Discuss the difference between Warehouse and Database.
- (b) What is Warehouse Schema? Discuss about Star Schema.
(or)
- (c) Discuss about the phases of data mining Architecture.
- (d) Discuss about Indexing Techniques used to organize the data in the Data mining.

Unit-II

- 3.(a) Discuss and Compare
 - i. Priori Algorithm.
 - ii. PAM Algorithm
- (b) What is Association Rule? Is every subset of any itemset must contain either a frequent set or a border set. Justify.
(OR)
- (b) Discuss about FP-Tree Growth Algorithm.
- (c) Discuss about various categories of Association Rules

Unit-III

- 4.(a) What is Decision Tree? Discuss about Decision Tree Construction Algorithms.
- (b) Discuss about Bayesian Classification.
(OR)
- (c) What is the purpose of Neural Network? Discuss about types of Neural Networks?
- (d) Discuss about salient features of the generic algorithm? How can a data mining problem can be an optimization problem? How can do you use generic algorithms for such a cases?

Unit-IV

5.(a) What is clustering? Discuss about various categories of clustering.

(b) Discuss about DBSCAN Algorithm.

(or)

(c) Discuss about BIRCH clustering

(d) Discuss about STING Algorithm with example.

Minor Changes for MSc Computer Science

MCS 202	DATABASE MANAGEMENT SYSTEMS	
Instruction: 4 periods / week		Credits: 5
Internal : 30 marks	University Exam: 70 marks	Total : 100 Marks

Objectives:

The course is designed to meet the objectives of:

1. the need of a database management system (DBMS)
2. the concept of data normalization
3. the concept of entity relationships
4. the concept of a client/server database, and
5. the concepts of MongoDB

Outcomes:

Students successfully completing this module will be able to:

1. design database, different operations, queries performed for a management system problems,
2. understand and design of ER-diagram in DBMS,
3. Implementation of different normalizations for database size reduction and removal of redundancy, and
4. Understand and design application development in MongoDB

SYLLABUS

Unit-I

Databases and Database Users: Introduction, Characteristics of the Database Approach, Actors on the Scene, Workers behind the scene, Advantages of the using the DBMS Approach.

Database System Concepts and Architecture: Data Models, Schemas and Instances, Three Schema architecture and Data Independence, Database Languages and Interfaces, Centralized and Client/Server Architecture for DBMS, Classification of Database Management Systems.

Data Modeling Using the ER Model: Conceptual Data models, Entity Types, Entity Sets, Attributes and Keys, Relationship types, Relationship sets, roles and structural Constraints, Weak Entity types, Relationship Types of Degree Higher than Two, Refining the ER Design for the COMPANY Database.

The Enhanced Entity-Relationship Model: Sub classes, Super classes and Inheritance, Specialization and Generalization, Constraints and Characteristics of Specialization and

Generalization Hierarchies, Modeling of Union Types using Categories, An Example University ERR Schema, Design Choices and Formal Definitions.

Unit-II

The Relational Data Model and Relational Database Constraints: Relational Model Concepts, Relational Model Constraints and Relational Database Schemas, Update Operations, Transactions and Dealing with Constraint Violations.

The Relational Algebra and Relational Calculus: Unary Relational Operations: SELECT and PROJECT, Relational Algebra Operations from set Theory, Binary Relational Operations: JOIN and DIVISION, Additional Relational Operations, Examples, The Tuple Calculus and Domain Calculus.

SQL-99: Schema Definition, Constraints, Queries and Views: SQL Data Definitions and Data Types, Specifying Constraints in SQL, Schema Change Statements on SQL, Basic Queries in SQL, More Complex SQL Queries, INSERT, DELETE and UPDATE statements in SQL, Triggers and Views.

Unit-III

Functional Dependencies and Normalization for Relational Databases: Informal Design Guidelines for Relation Schemas, Functional dependencies, Normal Forms Based in Primary Keys, General Definitions of Second and Third Normal Forms, Boyce-Codd Normal Form.

Relational Database Design Algorithms and Further Dependencies: Properties of Relational Decompositions, Algorithms fro Relational Database Schema Design, Multivalued Dependencies and Fourth Normal Form, Join Dependencies and Fifth Normal Form, Inclusion Dependencies, Other Dependencies and Normal Forms.

Unit-IV

Document oriented data: principles of schema design, designing an e-commerce data model, Nuts and bolts on databasem collections and documents.

Queries and Aggregation: e-commerce's queries, MongoDB's query language, aggregating orders, aggregating in detail.

Updates atomic operations and deletes: Document updates, e-commerce updates, atomic document processing, nuts and bolts: MongoDB updates and deletes.

Prescribed Text :

RamezElmasri, Shamkant B. Navathe, "Fundamentals of Database Systems", Fifth Edition, Pearson Education (2007)

Chapters: 1.1 to 1.6, 2, 3.1 to 3.6, 4.1 to 4.5, 5, 6, 8, 10, 11

MongoDB in Action, Kyle Banker, Manning Publication and Co.

Chapters: 4,5 and 6

Reference Books :

1 C.J. Date, A.Kannan, S. Swamynathan, "An Introduction to Database Systems", VII Edition Pearson Education (2006).

2. Database system concepts, Silberschatz, Korth, Sudarshan, Mc-graw-hill,5thEd

3. MongoDB:LearnMongoDB in a simple Way, Dan Warnock

Model Paper

MCS 202: Database Management Systems

Time: 3 Hrs

Max.

Marks: 70

Answer Question No.1 Compulsory:

7 x 2 = 14 M

Answer ONE Question from each unit:

4 x 14 = 56 M

1. a) Aggregate Functions
b) failure of a transaction
c) Entity types.
d) functional dependencies
e) Null Values
f) Centralized databases
g) MongoDB Schema design
h) First normal Form

Unit-I

2. a) Discuss about three level architecture with representation of data in each level.
b) Discuss about levels of RAID.

(or)

3. a) Compare and Construct the indexing of data by using B and B⁺ Trees.
b) Discuss about Data Independency with an example.

Unit-II

4. a) Discuss about Arithmetic functions in SQL with example?
b) Express the following statements in terms of Relational Algebra
 - i)Fetch the Department Numbers consisting of more than three employees.
 - ii)Fetch the Employee aggregated salary for a department.

(or)

5. a) Discuss about Views and its Limitations?
b) What is Index? Create an index for the employees belongs to the Accounts and Sales departments.

Unit-III

6. a) What is Functional Dependency? Explain the role of FD's in construction of Relational Schema.
b) Can I say that BCNF is equivalent Normal Form for III NF, Justify?

(or)

7. a) What is Non-Loss Dependency? Explain with an example.
b) Discuss the following
 - i . II NF
 - ii . Multi-valued Dependency.

Unit-IV

8. Explain MongoDB query selectors.

(or)

9. Explain how documents are processed in MongoDB.