NATIONAL UNIVERSITY OF SCIENCE AND TECHNOLOGY
FACULTY OF APPLIED SCIENCES
DEPARTMENT OF INFORMATICS AND ANALYTICS
ADVANCED DATA BASE AND DATA MINING
SCS5109

Examination Paper

2024

This examination paper consists of 4 pages

Time Allowed:	**3 hours**
Total Marks:	**100**
Examiner's Name:	
External Examiner:	

INSTRUCTIONS

1. Answer any four (4) questions

2. Each question carries 25 marks

MARK ALLOCATION

QUESTION	MARKS
1.	25
2.	25
3.	25
4.	25
5.	25
TOTAL FOR ANY FOUR QUESTION	100

QUESTION ONE

1. Design a university database schema using EER concepts of specialisation and generalisation, along with primary and foreign key dependencies. Primary key attributes should be underlined, while foreign key dependencies should appear as arrows from the foreign key attributes of the referencing relation to the primary key of the referenced relation. State any assumptions you will make.

[25]

QUESTION TWO

2. (a) A manufacturer makes three models of vehicle that vary in their engine type and number of seats. One model is available in two paint colours. Engines are either petrol or electric. Engines come from different suppliers, dependent on their fuel and horsepower.

      (i) Draw a suitable E/R diagram. [8]

      (ii) By writing out a suitable number of short tables, give a small relational database example holding vehicles and engines. Make sure every field has a name. Underline table keys. [6]

      (iii) Which fields in your example are foreign keys? Explain whether your example satisfies referential integrity. [5]

  (b) The SQL language contains a GROUP BY construct.

      (i)   Explain why the data returned by GROUP BY has to be passed through a reduction operator before it can be returned in a table? [3]

      (ii)   What mathematical properties should such a reduction operator have? Explain why. [3]

QUESTION THREE

3. Discuss how the emerging technologies are affecting enterprise database management [15]
Identify and discuss any two main challenges database designers face in designing databases that serve the needs of all levels of the enterprise. [10]

QUESTION FOUR

4. Consider a database that consists of the following relations.
    SUPPLIER(Sno, Sname)
    PART(Pno, Pname)
    PROJECT(Jno, Jname)
    SUPPLY(Sno, Pno, Jno)
    The database records information about suppliers, parts, and projects and includes a ternary relationship between suppliers, parts, and projects. This relationship is a many-many-many relationship. Specify and execute the following queries.

    a) Retrieve the part numbers that are supplied to exactly two projects.
    b) Retrieve the names of suppliers who supply more than two parts to project 'J
    c) Retrieve the part numbers that are supplied by every supplier.
    d) Retrieve the project names that are supplied by supplier 'S1' only.
    e) Retrieve the names of suppliers who supply at least two different parts each to at least two different projects.

                                                                                    [25]

QUESTION FIVE

Consider the table below:

Customer ID	Transaction ID	Items Bought
1	0001	{a, d, e}
1	0033	{a, b, c, e}
2	0024	{a, b, d, e}
2	0012	{a, c, d, e}
3	0031	{b, c, e}
3	0015	{b, d, e}
4	0022	{c, d}
4	0029	{a, b, c}
5	0040	{a, d, e}
5	0038	{a, b, e}

a) Compute the support for itemsets {e} , {b, d} and {b, d, e} by treating each transaction ID as a market basket.   [6 marks]
(b) Use the results in part (a) to compute the confidence for the association rules {b, d} → {e} and {e} → {b, d}. Is confidence a symmetric measure?   [7]
(c) Generate elements for k=2; C2=apriori-gen(L1)   [6]
(d) Generate elements for k=3; C3=apriori-gen(L2)   [6]

QUESTION SIX

Use two binary attributes, leg cover (pants or skirts) and facial hair (some or none), to predict the sex of a person. From the data set of 2,000 individuals, half are male and half female. 75% of the males have no facial hair. Skirts are worn by 50% of the females. All females are bare-faced, and no male wears a skirt.

a) What is the initial entropy in the system? [5]
b) Compute the information gain by initially choosing the attribute leg-cover and initially choosing facial hair. [13]
c) Based on your answers, which attribute should be used as the root of a decision tree? [7]

log$_2$ approx.
log2(1/8) = -3.0
log2(1/4) = -2.00
log2(1/3) = -1.58
log2(3/8) = -1.42
log2(3/7) = -1.22
log2(1/2) = -1.00
log2(4/7) = -0.81
log2(5/8) = -0.68
log2(2/3) = -0.58
log2(3/4) = -0.42
log2(7/8) = -0.19

END OF QUESTION PAPER