CARDIFF UNIVERSITY

EXAMINATION PAPER

Academic Year: 2017/2018

Examination Period: Spring

Examination Paper Number: CMT207

Examination Paper Title: Information Modelling and Database Systems

Duration: 2 hours

Do not turn this page over until instructed to do so by the Senior Invigilator.

Structure of Examination Paper:

There are 5 pages.
There are 4 questions in total.
There are no appendices.
The maximum mark for the examination paper is 60 and the mark obtainable for a question or part of a question is shown in brackets alongside the question.

Students to be provided with:

The following items of stationery are to be provided:
1 answer book.

Instructions to Students:

Answer 3 questions.

Important note: if you answer more than the number of questions instructed, then answers will be marked in the order they appear only until the above instruction is met. Extra answers will be ignored. Clearly cancel any answers not intended for marking. Write clearly on the front of the answer book the numbers of the answers to be marked.

The use of a translation dictionary between English or Welsh and another language, provided that it bears an appropriate departmental stamp, is permitted in this examination.

Please turn over.

Question 1 – Security, transactions and concurrency

a. In the context of database transaction management, explain the following briefly: [8]

    i.   ACID properties
    ii.   S-lock and X-lock
    iii.   serialisable transactions
    iv.   transaction deadlock

b. What is a B-tree of order m? How does it improve the performance of a database?   [4]

c. A B-tree index does not always help improve the performance of a database system. Describe two cases in which the use of a B-tree index may not be effective.   [2]

d. Suppose that we have a file of records whose key values are integers and that a B-tree of order 2 is to be maintained for this file. Records with the following key values are to be inserted into an initially empty file in the given order:

23, 5, 39, 40, 56, 34, 89, 49, 20, 27, 4, 12, 10, 43, 51, 19, 98, 31, 72

    i.   Draw the B-tree as it appears after the insertion of the key value 27.   [3]
    ii.   Draw the B-tree as it appears after the final insertion.   [3]

Question 2 - Data mining

a. Briefly explain the market–basket data model. In relation to this model, define frequent itemsets. [4]

b. Describe the basic idea behind the a priori algorithm. Provide a pseudocode for this algorithm. [4]

c. The following six transactions were recorded in a supermarket:

#	Transaction
1	onion, potato, burger
2	potato, burger, milk
3	milk, beer
4	onion, potato, milk
5	onion, potato, burger, beer
6	onion, potato, burger, milk, beer

i. Given a support threshold s = 4, apply the a priori algorithm to find all frequent itemsets and explain each step. [6]

ii. Having identified all frequent itemsets, use them to find all association rules with high confidence if the confidence threshold is 0.85. [6]

Question 3 – Non–relational data models

a. For an XML document found at `http://store1.example.com/prices.xml`, whose content is given below, you are asked to do the following.

   i. Formulate an XQuery that lists the book titles with the minimum price for each book in the form of a `minprice` <u>element</u> with the book title as its `title` <u>attribute</u>.

                                                          [6]

   ii. Provide the results returned by the query. **[2]**

```xml
<prices>
  <book>
    <title>Crime and Punishment</title>
    <source>Amazon.com</source>
    <price>10.99</price>
  </book>
  <book>
    <title>Crime and Punishment</title>
    <source>ebay.com</source>
    <price>8.50</price>
  </book>
  <book>
    <title>War and Peace</title>
    <source>Amazon.com</source>
    <price>24.49</price>
  </book>
  <book>
    <title>War and Peace</title>
    <source>ebay.com</source>
    <price>29.00</price>
  </book>
  <book>
    <title>Sense and Sensibility</title>
    <source>Amazon.com</source>
    <price>14.99</price>
  </book>
  <book>
    <title>Sense and Sensibility</title>
    <source>ebay.com</source>
    <price>12.50</price>
  </book>
</prices>
```

b. Describe 4 properties that are commonly used to define Big Data. **[4]**

c. Compare SQL and NoSQL databases with respect to: **[8]**

    i.   schemas
    ii.   data storage models
    iii.   scaling
    iv.   data manipulation

Question 4 – Semantic web

a. Explain the idea of the Semantic Web.

[4]

b. What is a resource in the context of the Semantic Web?

[4]

c. Clarify the difference between names and identifiers. How are identifiers defined on the Semantic Web?

[6]

d. What is RDF? What are its basic properties?

[6]