CARDIFF UNIVERSITY

EXAMINATION PAPER

Academic Year:	2016/2017
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.

Question 1 – Information, data and knowledge

    a. Briefly define *data* and *information*. Explain their roles in *knowledge* creation. Justify your answers using a value chain model as an example.

                          [4]

    b. Knowledge may be categorised according to three criteria: *abstraction, accessibility,* and *purpose.* Identify and briefly discuss an example of knowledge relevant to each of these criteria.

                          [6]

    c. What is *metadata*? Give an example of how it is used.

                          [2]

    d. Identify and describe TWO types of metadata.

                          [4]

    e. Identify TWO methods of storing metadata. For each method, state one situation where it would be particularly appropriate to use.

                          [4]

Question 2 – Security, transactions and concurrency

a. What is a *view*? Provide an example of where such database privilege would be appropriate to use.

[2]

b. Briefly define database *security* and database *integrity*.

[2]

c. A database transaction must satisfy the ACID properties. Identify, and briefly explain the importance of, TWO of these principles.

[4]

d. Consider the following five transactions and schedule. With the aid of a precedence graph, explain why this schedule is, or is not, serializable.

T1	T2	T3	T4	T5
w(Y)				
	r(X)			
		r(Y)		
			w(X)	
				w(Q)
	r(Z)			
			r(Y)	
			r(Q)	
r(Z)				
w(Z)				
	r(X)			
		r(Q)		

[5]

e. Briefly describe *timestamping* and why it is used.

[3]

f. Identify TWO advantages and TWO disadvantages of timestamping.

[4]

Please turn over

Question 3 – Non–relational data models

a. Translate the given JSON object into XML.

```
{
  "Simpsons": [
    {
        "firstName": "Homer",
        "lastName" : "Simpson",
        "age" : "40",
        "sex" : "Male"
    },
    {
        "firstName": "Marge",
        "lastName" : "Simpson",
        "age" : "30",
        "sex" : "Female",
        "relatives" : [ {"first" : "Homer", "last" : "Simpson"}]
    },
    {
        "firstName": "Lisa",
        "lastName" : "Simpson",
        "age" : "15",
        "sex" : "Female",
        "pet" : [ {"name" : "Snowball"},
                  {"name" : "Santa's little helper"}]
    }
  ]
}
```

[4]

b. Formulate an XQuery which returns the full names of all members of the family that are under the age of 35 from the Simpsons.xml document. Alphabetically order your answer by their first name.

[3]

c. Discuss the similarities and differences between JSON and XML.

[3]

d. A graph database is a type of NoSQL database. With the aid of a diagram, explain its general concepts.

[4]

e. Explain the idea of *object-orientated database management systems*. Identify four main features of such systems.

[6]

Question 4 – Data mining

a. Briefly explain *inter-annotator agreement* and why it is measured.

[3]

b.

Instance	Annotator A	Annotator B
1	Red	Red
2	Blue	Blue
3	Red	Blue
4	Red	Blue
5	Red	Red
6	Red	Blue
7	Blue	Red
8	Blue	Red
9	Blue	Blue
10	Red	Blue
11	Red	Red
12	Blue	Blue

Calculate inter-annotator agreement for the given dataset using Cohen's Kappa $\left(k = \frac{A_o - A_e}{1 - A_e}\right)$.

[4]

c. Briefly define *supervised machine learning* and *rule based classification*.

[3]

d.

Instance	Number of sides	Number of corners	Colour	Shape
1	4	4	Red	Square
2	3	3	Blue	Triangle
3	1	0	Green	Circle
4	1	0	Red	Circle
5	3	3	Blue	Triangle
6	4	4	Green	Triangle
7	4	4	Yellow	Square
8	3	3	Red	Triangle
9	3	3	Blue	Triangle
10	4	4	Green	Square
11	1	0	Red	Circle
12	4	4	Green	Circle
Test	4	4	Green	?

Using Bayes' theorem $\left(P(C|A) = \frac{P(A|C) \cdot P(C)}{P(A)}\right)$, suggest the shape for the given test instance. Justify your answer by showing which class has the highest probability.

[10]