

CARDIFF UNIVERSITY

EXAMINATION PAPER

Academic Year: 2015/2016

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 – Security, transactions and concurrency

- a. Describe the information that is held in a transaction log. Show how this information can be used to recover from system failures. [5]
- b. Describe the two-phase locking protocol. [2]
- c. What is a deadlock? Give an example of a deadlock involving two transactions which use the two-phase locking protocol. [5]
- d. What is a wait-for graph and how can it be used for deadlock detection? Illustrate this on the example from your answer to Part (c). [4]
- e. Describe two ways of deadlock prevention. [4]

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 five transactions were recorded in a supermarket:

#	Transaction
1	beer, nappy, baby powder, bread, umbrella
2	nappy, baby powder
3	beer, nappy, milk
4	nappy, beer, detergent
5	beer, milk, cola

- i. Given a support threshold $s = 2$, apply the a priori algorithm to find all frequent itemsets and explain each step.

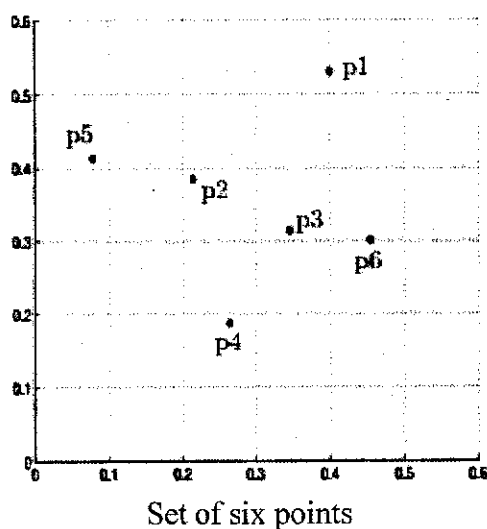
[3]

- ii. Having identified all frequent itemsets, use them to find all association rules with high confidence if the confidence threshold is 0.7.

[6]

- d. A set of six points $\{p_1, \dots, p_6\}$ is given in the graph below. Distances between these points are given in a distance matrix next to it. Apply hierarchical clustering to this set to produce a dendrogram using the complete link (MAX) method.

[3]



	p1	p2	p3	p4	p5	p6
p1	0.0000	0.2357	0.2218	0.3688	0.3421	0.2347
p2	0.2357	0.0000	0.1483	0.2042	0.1388	0.2540
p3	0.2218	0.1483	0.0000	0.1513	0.2843	0.1100
p4	0.3688	0.2042	0.1513	0.0000	0.2932	0.2216
p5	0.3421	0.1388	0.2843	0.2932	0.0000	0.3921
p6	0.2347	0.2540	0.1100	0.2216	0.3921	0.0000

Distance matrix

Please turn over

Question 3 – Non-relational data models

- a. Explain the idea of object-oriented database management systems. Briefly discuss 2 advantages and 2 disadvantages of such systems. [6]
- b. For an XML document found at <http://store1.example.com/bib.xml>, whose content is given below, you are asked to do the following.
- i. Formulate an XQuery that lists the titles and years of all books published by Addison-Wesley after 1991, in alphabetic order. [3]
- ii. Translate the given XML document into a JSON object. [3]

```

<bib>
  <book year="1994">
    <title>TCP/IP Illustrated</title>
    <author><last>Stevens</last><first>W.</first></author>
    <publisher>Addison-Wesley</publisher>
    <price>65.95</price>
  </book>
  <book year="1992">
    <title>Advanced Programming in the Unix environment</title>
    <author><last>Stevens</last><first>W.</first></author>
    <publisher>Addison-Wesley</publisher>
    <price>65.95</price>
  </book>
  <book year="2000">
    <title>Data on the Web</title>
    <author><last>Abiteboul</last><first>Serge</first></author>
    <author><last>Buneman</last><first>Peter</first></author>
    <author><last>Suciu</last><first>Dan</first></author>
    <publisher>Morgan Kaufmann Publishers</publisher>
    <price>39.95</price>
  </book>
  <book year="1999">
    <title>The Economics of Technology and Content for Digital TV</title>
    <editor>
      <last>Gerbang</last><first>Darcy</first>
      <affiliation>CITI</affiliation>
    </editor>
    <publisher>Kluwer Academic Publishers</publisher>
    <price>129.95</price>
  </book>
</bib>

```

- c. Explain 4 properties that are commonly used to define Big Data. [4]
- d. There are 4 types of NoSQL databases. Choose 2 and explain their basic principles. [4]

Question 4 – Semantic web

- a. What is meant by resource in the context of the Semantic Web? What is a URI?
A large subset of URIs share a common syntax with four main components:

<scheme>://<authority><path>?<query>

Briefly explain the role of each of these URI components.

[6]

- b. The following code represents RDF statements written in RDF/XML. Represent these statements as an RDF graph.

[4]

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns="http://www.example.org/~joe/contact.rdf#">
  <foaf:Person rdf:about="http://www.example.org/~joe/contact.rdf#joesmith">
    <foaf:mbox rdf:resource="mailto:joe.smith@example.org"/>
    <foaf:homepage rdf:resource="http://www.example.org/~joe/" />
    <foaf:family_name>Smith</foaf:family_name>
    <foaf:givenname>Joe</foaf:givenname>
  </foaf:Person>
</rdf:RDF>
```

- c. RSS 1.0 is a most widely deployed Resource Description Framework (RDF) application on the Web. An RSS summary is a document describing a channel consisting of URL-retrievable items. The example given below illustrates how channel and items can be specified in RSS 1.0. Describe an algorithm that parses an RSS 1.0 summary in order to display an RSS feed on a web site. You can do this using pseudocode, a flowchart or a textual description of the main steps.

[4]

```
<?xml version="1.0"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns="http://purl.org/rss/1.0/">
  <channel rdf:about="http://example.com/news.rss">
    <title>Example Channel</title>
    <link>http://example.com/</link>
    <description>My example channel</description>
    <items>
      <rdf:Seq>
        <rdf:li resource="http://example.com/2002/09/01/" />
        <rdf:li resource="http://example.com/2002/09/02/" />
      </rdf:Seq>
    </items>
  </channel>
  <item rdf:about="http://example.com/2002/09/01/">
    <title>News for September the First</title>
    <link>http://example.com/2002/09/01/</link>
    <description>other things happened today</description>
  </item>
  <item rdf:about="http://example.com/2002/09/02/">
    <title>News for September the Second</title>
    <link>http://example.com/2002/09/02/</link>
    <description>other things happened today</description>
  </item>
</rdf:RDF>
```

- d. Briefly define ontologies and their basic elements. Explain 2 examples of potential programmatic applications of ontologies.

[6]