CM3010

BSc EXAMINATION

COMPUTER SCIENCE

Databases and Advanced Data Techniques

Release date: Monday 19 September 2022 at 12:00 midday British Summer Time

Submission date: Tuesday 20 September 2022 by 12:00 midday British Summer Time

Time allowed: 24 hours to submit

INSTRUCTIONS TO CANDIDATES:

Section A of this assessment paper consists of a set of **TEN** Multiple Choice Questions (MCQs) which you will take separately from this paper. You should attempt to answer **ALL** the questions in Section A. The maximum mark for Section A is **40**.

Section A will be completed online on the VLE. You may choose to access the MCQs at any time following the release of the paper, but once you have accessed the MCQs you must submit your answers before the deadline or within **4 hours** of starting whichever occurs first.

Section B of this assessment paper is an online assessment to be completed within the same 24-hour window as Section A. We anticipate that approximately **1 hour** is sufficient for you to answer Section B. Candidates must answer **TWO** out of the THREE questions in Section B. The maximum mark for Section B is **60**.

Calculators are not permitted in this examination. Credit will only be given if all workings are shown.

You should complete **Section B** of this paper and submit your answers as **one document**, if possible, in Microsoft Word or PDF to the appropriate area on the VLE. Each file uploaded must be accompanied by a coversheet containing your **candidate number** written clearly at the top of the page before you upload your work. Do not write your name anywhere in your answers.

SECTION A

Candidates should answer the **TEN** Multiple Choice Questions (MCQs) quiz, **Question 1** in Section A on the VLE.

SECTION

Candidates should answer any **TW** questions from Section B

Question 2

An organisation monitoring performance on non-verbal reasoning tests of school and college pupils maintains a database of results.

To inspect the table for research purposes, a sociologist runs the following MySQL command:

```
SELECT AVG(Score) AS Average,
       Year(TestDate) AS TestYear,
       Gender,
       TIMESTAMPDIFF(YEAR, BirthDate, TestDate) AS Age,
       Student.City as City
  FROM Test INNER JOIN Student ON Test.Student=Student.ID
  GROUP BY TestYear, Gender, Age, City
```

An extract of the result is given here:

Average	TestYear	Gender	Age	City
50.5	2019	M	12	Birmingham, Al
80.3	2019	M	12	Berlin, DE
80.9	2019	F	12	Berlin, DE
63	2019	F	12	Birmingham
91	2019	M	12	Seoul
89.3	2018	F	12	Seoul
81	2019	M	11	Birmingham, USA

(a) Which aggregate function is used here? [1]

(b) There is a problem with the database design that risks making the aggregation incorrect. What is it, and how could it be resolved? [6]

(c) For security reasons, the researcher should be given minimal, read-only access to the database. Give a suitable command that the database administrator should run to provide this. [4]

d) From the point of view of handling confidential information about minors, it would be better to give access only to aggregated data. How would you achieve that? You are not required to give the detailed commands, just describe them.) [4]

(e) What limitation would that create for the researcher? [1]

(f) Database users have found several issues that come from the structure of the data model. These focus on the Student table. This table is defined by the following MySQL command:

```
CREATE TABLE Student(
  ID VARCHAR(25) PRIMARY KEY,
  GivenName VARCHAR(80) NOT NULL,
  FamilyName VARCHAR(80) NOT NULL,
  Gender ENUM('M', 'F') NOT NULL,
  BirthDate DATE NOT NULL,
  School VARCHAR(130),
  City VARCHAR(130));
```

What problems can you see with this table, and how would you resolve them? Be as specific as you can be. [8]

(g) How well would this data work in an object database like MongoDB? What would be the advantages or disadvantages? [6]

Question 3

An entry in the Oxford Medieval Manuscript catalogue begins as follows:

```
<?xml version= 1.0  encoding= UTF-8 ?>
<?xml-model href= https://raw.githubusercontent.com/bodleian/
  consolidated-tei-schema/master/msdesc.rng
  type= application/xml
  schematypens= http://relaxng.org/ns/structure/1.0 ?>
<?xml-model href= https://raw.githubusercontent.com/bodleian/
  consolidated-tei-schema/master/msdesc.rng  type= application/xml
  schematypens= http://purl.oclc.org/dsdl/schematron ?>
<TEI xml:id= manuscript_3945  xmlns= http://www.tei-c.org/ns/1.0 >
   <teiHeader xmlns:tei= http://www.tei-c.org/ns/1.0 >
     <fileDesc>
       <titleStmt>
         <title>Christ Church MS. 341</title>
         <title type= collection >Christ Church MSS.</title>
         <respStmt>
            <resp>Cataloguer</resp>
            <persName>Ralph Hanna</persName>
            <persName>David Rundle</persName>
         </respStmt>
       </titleStmt>
```

(a) What markup language is being used? And what is the root node?   [2]

(b) Is this fragment well-formed? Justify your answer.   [3]

(c) What would be selected by evaluating the XPath expression `//fileDesc//`
`title/@type`?   [2]

(d) What would be selected by evaluating `//resp[text()= Cataloguer ]/`
`../persName`?   [2]

e) Why might you choose the expression given in part c) rather
than the simpler   `persName`? Give two situations where it would be   [4]
preferable. Another extract of the same file is reproduced below.

```
<msItem n= 2 >
  <locus> Fols 9{10 <hi rend= superscript >v</hi></locus>
  <incipit>Seynt austyn sei in e secounde boke of cristen
```

```
doctrine in e ende What euer ing a man fyndi</incipit>
<explicit>goddis spechis in expownynge bi at we heringe
as drinking ben more holpen</explicit>
<note> A prologue to <title>Oon of Foure</title>, ed. FM,
1:44{49 (the page foot) from BL, MS Harl. 6333. The text
appears in a minority of <title>Oon of Foure</title> MSS,
and separately in a transcript at Lambeth Palace Library,
MS 594, pp. 47{48 (s. xvii).</note>
</msItem>
```

(f) This element refers to the second textual item (such as a story or sermon) that the manuscript contains – hence n= 2 . How well would this way of listing contents work in a relational model? How would you approach the problem? [8]

(g) Here is an extract of the file msdesc.rng:

```
<define name= titleStmt >
  <element name= titleStmt >
    <a:documentation
      xmlns:a= http://relaxng.org/ns/compatibility/annotations/1.0 >
    (title statement) groups information about the title of a work
    and those responsible for its content. [2.2.1. The Title Statement
    2.2. The File Description]
    </a:documentation>
    <group>
      <oneOrMore>
        <ref name= title />
      </oneOrMore>
      <zeroOrMore><ref name= model.respLike /></zeroOrMore>
    </group>
    <ref name= att.global.attributes />
```

What is this file, and why is it referenced in the catalogue entry? [3]

(h) What is the difference between valid and well-formed XML? [2]

i) If the first extract in this question had omitted the resp tmt element, would the XML have been legal? [1]

j) If the first extract in this question had omitted the titl elements, would the XML have been legal? [1]

k) This catalogue entry is converted automatically to HTML for the Oxford website whenever it changes. What two technologies would be most likely to be considered for the conversion? [2]

Question 4

(a) Here is data presented in a serialisation of a data interchange model:

```
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix oa: <http://www.w3.org/ns/oa#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix myrdf: <http://example.org/> .
@prefix armadale: <https://literary-greats.com/WCollins/Armadale/>;

myrdf:anno-001 a oa:Annotation ;
  dcterms:created  2015-10-13T13:00:00+00:00 ^^xsd:dateTime ;
  dcterms:creator myrdf:DL192 ;
  oa:hasBody [
    a oa:TextualBody ;
    rdf:value  Note the use of visual language here.
  ];
  oa:hasTarget [
    a oa:SpecificResource ;
    oa:hasSelector [
      a oa:TextPositionSelector ;
      oa:start  235 ^^xsd:nonNegativeInteger;
      oa:end  300 ^^xsd:nonNegativeInteger ;
  ];
  oa:hasSource armadale:Chapter3/ ;
  oa:motivatedBy oa:commenting ;

myrdf:DL192 a foaf:Person ;
    foaf:name  David Lewis  .
```

   i. What is the model?   [1]

   ii. What is the serialisation format?   [1]

(b) Name two ontologies used in this document.   [3]

(c) For each ontology named in your previous answer, name all the **properties** from the ontology that are used in this document.   [5]

d) This structure is a Web Annotation previously called Open Annotation).
The BODY of the annotation contains a comment on the TARGET, which is

often a part of a SOURCE. In this case, an online chapter of a book is the
source, some part of which is being selected as a target.

A scholar would like to get all the annotations in a database about this
particular chapter – they just want the annotation text and the name of the
author. They have tried the following SPARQL, but it doesn t work. Write
a correct version that will do what they ask.

```
SELECT ?body ?creator
WHERE {
  ?annotation a oa:Annotation .
            ?creator   ;
            oa:hasBody body .
            hasSource
armadale:Chapter3 }
```

[7]

(e) Some Linked Data systems use a backend database to store the data and
for quick retrieval, exporting it as needed. Draw an ER diagram for web
annotations like this. [5]

(f) Identify the tables that you would need for a relational implementation and
list the keys for each. [5]

(g) Give a MySQL query equivalent for the scholar s query you corrected in
question 3(d) [3]

END OF PAPER