UNIVERSITY OF LONDON

CM3010

BSc EXAMINATION

COMPUTER SCIENCE

Databases and Advanced Data Techniques

Release date: Tuesday 7 September 2021 at 12:00 midday British Summer Time

Submission date: Wednesday 8 September 2021 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 **10** 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 a PDF to the appropriate area on the VLE. You are permitted to upload 30 documents. However, we advise you to upload as few documents as possible. Each file uploaded must be accompanied by a coversheet containing your **candidate number**. In addition, your answers must have 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.

UL21/1054

SECTION B

Candidates should answer any **TWO** questions from Section B.

Question 2

The table below is an extract from a bird spotter's records

Species	Date	Number sighted	Conservation status	Nature reserve	Location
Bar-tailed godwit	2021-04-21	31	Least concern	Rainham Marshes	51.5N 0.2E
Wood pigeon	2021-04-21	31	Least concern	Rainham Marshes	51.5N 0.2E
Greater spotted woodpecker	2021-06-13	1	Least concern	Epping Forest	51.6N 0.0E
European turtle dove	2021-06-13	2	Vulnerable	Epping Forest	51.6N 0.0E
Wood pigeon	2021-06-13	2	Least concern	Epping Forest	51.6N 0.0E
Great bustard	2020-04-15	3	Vulnerable	Salisbury Plain	51.1N -1.8W
Bar-tailed godwit	2020-04-20	53	Least concern	Rainham Marshes	51.5N 0.2E

(a) This is the `sightings` table is in a MySQL database. Give a query to retrieve all bird types seen since the first of January 2021. [4]

(b) Is this table in 1NF? Explain your reasoning. [3]

(c) Normalise this data, listing the tables that result and their primary and foreign keys. [7]

(d) What normal form have you reached? Explain your conclusion. [4]

(e) Give a query for your new tables to retrieve bird types **and their conservation status** for birds seen since the first of January 2021. [5]

(f) The bird spotter wants to be sure that their next set of updates go in correctly. Would a **transaction** make a difference? Give example SQL operations to illustrate your argument. [7]

Question 3

Here is an extract of an MEI file.

```
<measure>
    <staff n="2">
        <layer n="1">
            <chord xml:id="d13e1" dur="8" dur.ppq="12"
                   stem.dir="up">
                <note xml:id="d1e101" pname="c" oct="5"/>
                <note xml:id="d1e118" pname="a" oct="4"/>
                <note xml:id="d1e136" pname="c" oct="4"/>
            </chord>
        </layer>
    </staff>
    <staff n="3">
        <layer n="1">
            <chord xml:id="d17e1" dur="8" dur.ppq="12"
                   stem.dir="up">
                <note xml:id="d1e157" pname="f" oct="3"/>
                <note xml:id="d1e174" pname="f" oct="2"/>
            </chord>
        </layer>
    </staff>
</measure>
```

(a) List all the element types you can see in this code. [2]

(b) I am trying to retrieve all chords in the staff with `n` of `2` – that is, they are in the right hand – but I only want chords that contain notes with a `pname` of `f`, but my XPath is incorrect. My attempt is:

```
/staff[n="2"]/layer/chord[note/@pname="c"]
```

Give an XPath expression that **would** work. [3]

(c) A group of developers have decided to evaluate a MongoDB implementation of the MEI model.

i. Translate the first `chord` element in the XML into JSON as effectively as you can. [5]

ii. Imagining the whole data structure was an array of `chord` objects, give a MongoDB `find` command that would return only chords with upward stems that have `f` in one of their notes. [5]

(d) A different group of developers have mapped the MEI model into linked data. The following SPARQL query finds all chords with at least one F in them.

```
SELECT DISTINCT ?chord
WHERE {
  ?chord rdfs:member ?note .
  ?note mei:pitchClass mei:FPitchName .
}
```

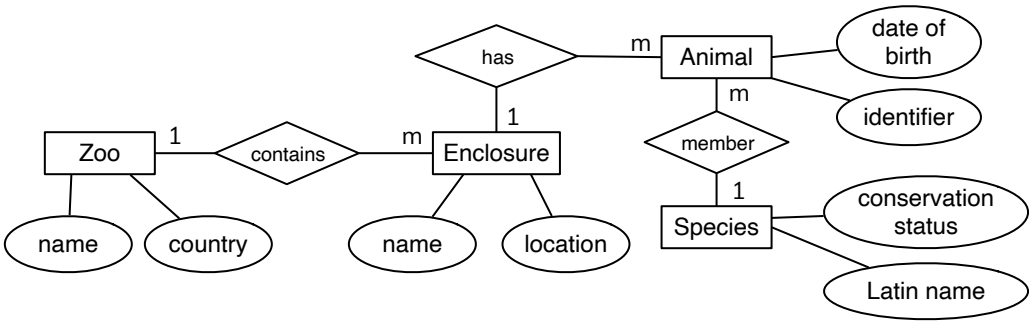
i. `rdfs:member` is defined by the W3 ontology RDF Scheme. Why are we using it here instead of a new `mei:hasNotes` property? [3]

ii. Give some RDF (in whichever serialisation you are most comfortable with) for the first chord element. Invent any new concepts you need in the mei `namespace` [5]

(e) How do these three models – XML, MongoDB/JSON and Linked Data – differ in what they might offer for music notation? What advantages and disadvantages does each have? [7]

Question 4

The E/R diagram below part of the design for a database system for coordinating zoos and other animal collections worldwide.



(a) List the tables and their fields for an SQL implementation of this design. Indicate primary keys for each table. [4]

(b) Give SQL `CREATE TABLE` commands for any TWO of your tables, including any **foreign keys**. [6]

(c) Give a single SQL query to find out how many species are housed in the zoo which has the name ‘Singapore Zoo‘. [5]

(d) Give a single SQL query to find out the date of birth of the oldest animal of the species called ‘Buceros bicornis‘ in each zoo. [5]

(e) Choose ONE of XML or RDF and:

i. BRIEFLY assess the suitability of this model for your chosen technology (i.e. XML or RDF graph). [3]

ii. Give some instance data for the database in your chosen technology. You should aim to cover all or nearly all the entities and attributes in the E/R diagram. [7]

END OF PAPER

.