UNIVERSITY OF LONDON

CM3010

BSc EXAMINATION

COMPUTER SCIENCE

This is a mock exam.

This mock exam provides you with examples of the level and style of long answer questions you will encounter in the real exam. It does not provide examples of multiple choice questions, though you will encounter those in the real exam.

Please note that aspects that are about the context and actual writing of the exam are subject to change. These could include: date/time; number of questions and choice available; balance between multiple choice and longform questions; permitted materials; exam platform details; use of calculators; marks awarded; etc.

Full details of the above for your real exams will be communicated in your admission notice, sent out closer to the time of your exams.

Part A

Question 1

Candidates should answer the **TEN** Multiple Choice Questions (MCQs) in Part A.

NOTE: these are not included in the mock paper

Part B

Candidates should answer **BOTH** questions from Part B.

IMPORTANT: full marks will only be awarded for the correct answer with ALL steps shown. For each of the following questions in Part B, make sure to show all the steps of your solution.

Question 2

The Bodleian Library of Oxford University stores its main catalogue using a standard called MARC. When I navigate to a particular book in the web interface and choose to view the source record, I see the following:

```
leader  00000nam a22002772i 4500
001     990118204070107026
005     20011114100255.0
008     860101s1954    enka   j      000 1 eng d
035     ##$a(UkOxU)011820407
035     ##$a(UkOxU)011820407BIB01
904     ##$aMatched
010     ##$aGB54-13352
035     ##$aOCLC ocm06935463 from D960307M
040     ##$aKWW $cKWW $dOCL $dUKM $dEQO
041     1#$aengswe
082     ##$a[Fic]
090     ##$aPZ7.L6585 $bPi5
092     ##$aD0503666694
100     1#$aLindgren, Astrid, $d1907-2002.
240     10$aPippi Långstrump. $lEnglish
245     10$aPippi Longstocking / $c[translated from the Swedish by Edna Hurup
; illustrated by Richard Kennedy]
260     ##$aLondon : $bOxford, $c[1954]
300     ##$a120 p. : $bill. ; $c21 cm
500     ##$aTranslation of Pippi Långstrump
520     ##$aEscapades of a lucky little girl who lives with a horse and a
monkey--but without any parents--at the edge of a Swedish village
700     1#$aKennedy, Richard.
```

Each row in this represents a data key value pair, although some are a little more complicated. The opening 8 lines (up to code 035) are mostly catalogue IDs. Code 040 refers to who was responsible for cataloguing, so $a introduces the original source of the entry, $c the transcribing agency (the same in this case) and $d refers to providers of modifications since the entry was created.
Code 041 refers to the language. The 1 at the beginning means that the book is in translation, and then the language codes are given after $a.

(a) The MARC standard specifies a binary format for sharing data from this sort of catalogue, but what sort of database system would be best for storing and retrieving it? Justify your answer.

[8 marks]

(b) Code point 100 indicates the main personal name associated with the record. The leading 1 refers to the format of the name – it specifies that surname is given first.

$a indicates the name string, while $d specifies the dates. How would you represent this information in your recommended database system? What problems might arise?

[6 marks]

(c) Suggest Entities, Attributes and Relationships for an ER model that would represent bibliographic items such as this book. Try to include everything you can see in the record above, and any additional information you might think would be useful.

[6 marks]

(d) The development of MARC was co-ordinated by the US Library of Congress. They created an XML format and schema for sharing MARC data. A fragment of MARC XML, derived from the MARC above, is given below:

```
<datafield tag="100" ind1="1" ind2=" ">
  <subfield code="a">Lindgren, Astrid,</subfield>
  <subfield code="d">1907-2002.</subfield>
</datafield>
```

Here, the main MARC field code (in this case, 100) is specified in the attribute 'tag', the first two characters of that field's value are encoded as 'ind1' and 'ind2' and the subfields are given their own elements with an attribute 'code' that matches the $ code in the source.

(i) Give an XPath expression for retrieving information on whether items are translations (see earlier for the relevant MARC code).

[2 marks]

An alternative approach to encoding MARC data in XML could have produced code like this:

```
<mainPerson>
  <name format="surnameFirst">Lindgren, Astrid,</name>
  <dates>1907-2002</dates>
</mainPerson>
```

(ii) What difference do you think this would make for the schema and the functionality of the system? Why might they have chosen not to do this (and to use datafield and subfield instead)?

[4 marks]

(e) The Library of Congress have moved towards developing a Linked Data standard called BIBFRAME to supersede MARC. What benefits and what risks might they expect from this move?

[4 marks]

Question 3

(a) Academic conferences are often organised through online database applications. These cover the review process and conference registration. The review process tracks who has submitted papers, assigning each paper to multiple reviewers, getting reviewer scores and feedback, and a program committee deciding which papers to accept. The conference registration process will cover who is registering, what should be printed on name tags, when they will attend (i.e. on which days they will come for a conference that spans several days) and any extras, such as workshops or dinner.
Develop an ER model for a basic conference system. List entities, attributes and relationships for your model (you don't need to draw anything, just answer as text).

[14 marks]

(b) Imagine you have converted your model to the relational model and implemented it as a MySQL database. Give a query that would return the number of dinner tickets ordered.

[4 marks]

(c) Reviews are often 'double blind'. This means that a reviewer doesn't know who authored the paper they are reviewing, and the author doesn't know who reviewed their paper. Give any GRANT statements for your database (and list any changes you'd have to make to your model) that would help support that anonymity.

[4 marks]

(d) A colleague looks at your database and suggests that, because some conferences can be very large with a thousand or more attendees, you should consider denormalising some tables to improve performance. How would you respond? Justify your answer.

[4 marks]

(e) Another colleague suggests that you replace the relational database with an XML database. What advantages or disadvantages might this offer?

[4 marks]

END OF PAPER