

IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

## EXAMINATIONS 2016

BEng Honours Degree in Electronic and Information Engineering Part II

MEng Honours Degree in Electronic and Information Engineering Part II

BEng Honours Degree in Mathematics and Computer Science Part II

MEng Honours Degree in Mathematics and Computer Science Part II

BEng Honours Degree in Mathematics and Computer Science Part III

MEng Honours Degree in Mathematics and Computer Science Part III

MSc in Computing Science

for Internal Students of the Imperial College of Science, Technology and Medicine

*This paper is also taken for the relevant examinations for the  
Associateship of the City and Guilds of London Institute*

### PAPER C526

### DATABASES

Wednesday 27 April 2016, 10:00

Duration: 120 minutes

*Answer THREE questions*

Paper contains 4 questions  
Calculators not required

Several parts of the following questions make use of the **mondial** relational database, a fragment of which is listed below. It contains information about countries, and the membership of countries in organisations. For each organization, there is a record of the city and country in which that organisation is based. The **is\_member** records both membership and other associations of a country with an organization. The percentage of the land area of each country that falls within a particular continent is recorded in **encompasses**. The **located** table (which has no key) records the province and country of some cities, and the **borders** table records which countries share a land border.

organization			
abbreviation	city	country?	established?
AL	Cairo	ET	1945-03-22
C	London	GB	1931-12-31
CERN	Geneva	CH	1953-07-01
EU	Brussels	B	1992-02-07
NATO	Brussels	B	1949-09-17
WFTU	Prague	null	1945-10-03
⋮			

country				
name	code	capital	area	population
Czech Republic	CZ	Prague	78,703	10,321,120
Switzerland	CH	Bern	41,290	7,207,060
Russia	R	Moscow	17,075,200	148,178,487
Belgium	B	Brussels	30,510	10,170,241
Turkey	TR	Ankara	780,580	62,484,478
United Kingdom	GB	London	244,820	58,489,975
Egypt	ET	Cairo	1,001,450	63,575,107
⋮				

encompasses		
country	continent	percentage
CZ	Europe	100
CH	Europe	100
R	Europe	20
R	Asia	80
B	Europe	100
TR	Europe	32
TR	Asia	68
GB	Europe	100
ET	Asia	10
ET	Africa	90
⋮		

is_member		
country	organization	type
CZ	CERN	member
CZ	EU	member
CZ	WFTU	member
CH	CERN	member
R	CERN	observer
B	CERN	member
B	EU	member
B	NATO	member
TR	CERN	observer
TR	NATO	member
TR	WFTU	member
GB	C	member
GB	CERN	member
GB	EU	member
GB	NATO	member
⋮		

borders		
country1	country2	length
GR	TR	206
CZ	A	362
CZ	D	646
CZ	PL	658
FL	CH	41
SK	CZ	215
CH	F	573
CH	A	164
CH	D	334
CH	I	740
PL	R	206
UA	R	1,576
B	F	620
B	D	167
B	NL	450
L	B	148
TR	IR	499
⋮		

located		
city	province	country
Cardiff	Cardiff	GB
London	Greater London	GB
Belfast	Northern Ireland	GB
⋮		

$\text{is\_member}(\text{organization}) \overset{fk}{\rightleftharpoons} \text{organization}(\text{abbreviation})$      $\text{encompasses}(\text{country}) \overset{fk}{\rightleftharpoons} \text{country}(\text{code})$   
 $\text{is\_member}(\text{country}) \overset{fk}{\rightleftharpoons} \text{country}(\text{code})$      $\text{organization}(\text{country}) \overset{fk}{\rightleftharpoons} \text{country}(\text{code})$   
 $\text{borders}(\text{country1}) \overset{fk}{\rightleftharpoons} \text{country}(\text{code})$      $\text{located}(\text{country}) \overset{fk}{\rightleftharpoons} \text{country}(\text{code})$   
 $\text{borders}(\text{country2}) \overset{fk}{\rightleftharpoons} \text{country}(\text{code})$

1 The following parts all refer to the **mondial** relational schema on Page 1.

- a Write an RA query that returns the scheme (name,organization) listing the names of countries, and the organizations that they are members of.
- b Write an RA query that returns the scheme (name) listing the names of countries where an organization is based, but that country plays no role (*i.e.* has no entry in is\_member) in the organization.

c Consider the following RA query:

$\pi_{\text{country,organization}} \sigma_{\text{type}='member'} \text{is\_member} \div \pi_{\text{country}} \sigma_{\text{organization}='EU' \wedge \text{type}='member'} \text{is\_member}$

- i) List the result of the query when run on the fragment of data on Page 1, and explain the semantics of the query.
  - ii) Translate the RA query into an equivalent SQL query.
  - iii) Translate the RA query into an equivalent Datalog query.
- d Write a query in each of the following languages returning the scheme (city) listing the names of cities that are both the base for organisations, and the capital city of a country. You should assume that a city name is unique within each country, but the same city name may appear in a different country.
  - i) RA
  - ii) SQL
  - iii) Datalog
- e Suppose the RA query  $\pi_{\text{city}} \text{organization} - \pi_{\text{capital as city}} \text{country}$  has been executed at some point in time to give result Q, after which  $\Delta_o$  has been added to organization to give organization', and  $\Delta_c$  has been added to country to give country'. Give an RA query in terms of Q, organization', country',  $\Delta_o$  and  $\Delta_c$  that returns Q', *i.e.* the result of executing the RA query on the modified database.

*The five parts carry, respectively, 10%, 10%, 35%, 30%, and 15% of the marks.*

2 The following parts all refer to the **mondial** relational schema on Page 1.

a Consider the following SQL query:

```
SELECT organization.abbreviation ,
       encompasses.continent
FROM   organization
       JOIN encompasses
       ON   encompasses.country=organization.country
WHERE  encompasses.continent=ALL
      (SELECT encompasses.continent
       FROM   is_member
       JOIN   encompasses
       ON     encompasses.country=is_member.country
       WHERE  is_member.organization=organization.abbreviation)
```

- i) Briefly explain the semantics of the query, and compute the result of the query on the fragment of data given on Page 1.
  - ii) Rewrite the query listed above to use NOT IN or NOT EXISTS instead of the ALL operator.
- b Write an SQL query that returns the scheme (continent, no\_based) that must list every continent, together with the number of organizations known to be based in the continent.
- c Write an SQL query that returns the scheme (abbreviation, city, europe, asia, africa) listing every organization, the city in which that organization is based, and the number of members found in each of Europe, Asia, and Africa. Note that you should only count a country as being in a continent if at least 50% of its area is within that continent.
- d Write an SQL query that returns the scheme (organization, name, population) listing for each organization with members, the names of the five largest countries (in terms of population) that are members of the organization.
- e Write an SQL query that returns the scheme (name, population, border\_length, neighbour\_population) listing for every country its name and population, together with the total length of any borders that country has, and the total population of any neighbouring countries.

*The five parts carry, respectively, 20%, 15%, 20%, 20%, and 25% of the marks.*

- 3 a Suppose you have to design a new database to hold the following information about the companies that do business with ACME Computing Ltd. These companies may be customers, suppliers, or both. For all companies we record their name and contact email address, and for companies that are VAT registered, we must record their VAT number.

For suppliers, we record the purchasing manager that deals with the supplier, and record a number of types of product that the supplier can supply. We associate to each supplier all the stock items currently being supplied. It is company policy that each stock item may come from only one supplier, and some stock items are manufactured by ACME Computing Ltd itself, and therefore have no supplier. Each stock item has a part number and description, and some stock items have their colour recorded.

For customers, we record the sales manager that deals with the customer, and a credit limit. We also record all orders made by the customer. Each order is on a particular date, is given an order number, and has a reference number given by the customer. An order may have any number of stock items, with the quantity and price for each stock item recorded.

When orders are sent, we record the tracking number of each parcel the order was sent in, and the date on which the parcel was sent. We also wish to record how much of each stock item was put in each parcel, in case the parcel gets lost and an insurance claim must be made. We record the delivery address for each order, from a record of delivery addresses, for which we record the customer name, postcode and address. We identify an address record by the combination of the customer name and postcode.

- i) Design an ER<sup>ADHKLMNOSVW</sup> schema to represent this new database.
  - ii) Map the ER schema you designed in (i) into a relational schema.
- b Suppose that a relation  $R(A, B, C, D, E, F, G, H)$  has the functional dependencies:  
 $S = \{A \rightarrow FG, AB \rightarrow ED, ABH \rightarrow E, B \rightarrow H, CD \rightarrow CE, E \rightarrow D\}.$
- i) Compute a minimum cover  $S_c$  of  $S$ .
  - ii) Identify and justify all the candidate keys of  $R$ .
  - iii) Decompose the relation  $R$  into 3NF, maintaining FDs.
  - iv) Decompose the relation  $R$  into BCNF, and identify which (if any) of the FDs in  $S_c$  are not preserved by the BCNF you have decomposed from  $R$ .

*The two parts carry, respectively, 45%, and 55% of the marks.*

- 4 a The following histories describe the sequence of operations performed respectively by four transactions  $T_1$ – $T_4$ .

$H_1 = r_1[c_{CZ}], r_1[c_R], r_1[c_B], r_1[c_{GB}], c_1$

$H_2 = r_2[c_B], r_2[c_R], r_2[c_{CZ}], w_2[c_{CZ}], c_2$

$H_3 = r_3[c_B], w_3[c_B], r_3[c_{CZ}], w_3[c_{CZ}], c_3$

$H_4 = w_4[c_R], w_4[c_B], w_4[c_{GB}], c_4$

- i) Briefly explain if the following concurrent execution is serialisable and recoverable. If non-serialisable, explain what anomaly occurs.

$H_a = r_1[c_{CZ}], r_1[c_R], r_3[c_B], w_3[c_B], r_3[c_{CZ}], w_3[c_{CZ}],$   
 $r_1[c_B], r_1[c_{GB}], c_1, c_3,$

- ii) Briefly explain if the following concurrent execution is serialisable and recoverable. If non-serialisable, explain what anomaly occurs.

$H_b = r_3[c_B], w_4[c_R], w_4[c_B], w_4[c_{GB}], w_3[c_B],$   
 $r_3[c_{CZ}], w_3[c_{CZ}], c_4, c_3$

- iii) Briefly explain if the following concurrent execution is serialisable and recoverable. If non-serialisable, explain what anomaly occurs.

$H_c = r_1[c_{CZ}], w_4[c_R], r_1[c_R], w_4[c_B], r_1[c_B], w_4[c_{GB}], c_4, r_1[c_{GB}], c_1$

- iv) Briefly explain which (if any) pair of the transactions taken from  $T_1$ – $T_4$  will be serialisable for all concurrent executions of the pair, and also briefly explain which (if any) pair of the transactions taken from  $T_1$ – $T_4$  will be recoverable for all concurrent executions of the pair.

- v) Give a concurrent execution of the four transactions, which produces a deadlock involving all four transactions, and draw a waits-for graph for the deadlock state.

- b The table below lists the contents of a database log, which keeps only UNDO records of updates to the country table on Page 1.

UNDO	$w_1[c_R, \text{population} = 148, 179, 000]$
UNDO	$w_4[c_{TR}, \text{population} = 62, 481, 123]$
UNDO	$w_2[c_{CH}, \text{population} = 7, 392, 444]$
UNDO	$w_2[c_{GB}, \text{population} = 58, 543, 111]$
UNDO	$w_3[c_{CH}, \text{population} = 7, 312, 222]$
UNDO	$w_5[c_{CH}, \text{population} = 7, 210, 000]$
LOG	$c_4$
UNDO	$w_3[c_B, \text{population} = 11, 020, 000]$
LOG	$c_3$

- i) If at the time of recovery the country table on disc was found to have the data listed as on Page 1, describe the actions performed by the recovery procedure, and what population figures will be left after recovery.
- ii) If an additional LOG record were added for  $a_5$  to record the completion of aborting  $T_5$ , would your answer to (i) change, and if so, how does it change?
- iii) Considering the time just after when  $c_4$  occurs, describe and justify which updates from the above log must have been written to disc, which might have been written to disc, and which must not have been written to disc.

*The two parts carry, respectively, 65%, and 35% of the marks.*