IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

EXAMINATIONS 2014

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
BSc Honours Degree in Mathematics and Computer Science Part III
MSci 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

Thursday 1 May 2014, 14:30
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 organisation, there is a record of the city and country in which that organisation is based. 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
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	WFTU	member
CH	CERN	member
R	CERN	observer
B	CERN	member
B	NATO	member
TR	CERN	observer
TR	NATO	member
TR	WFTU	member
GB	C	member
GB	CERN	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
⋮		

is_member(organization) $\overset{fk}{\Rightarrow}$ organization(abbreviation)
is_member(country) $\overset{fk}{\Rightarrow}$ country(code)
borders(country1) $\overset{fk}{\Rightarrow}$ country(code)
borders(country2) $\overset{fk}{\Rightarrow}$ country(code)
encompasses(country) $\overset{fk}{\Rightarrow}$ country(code)
organization(country) $\overset{fk}{\Rightarrow}$ country(code)
located(country) $\overset{fk}{\Rightarrow}$ country(code)

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

a Write an RA query that returns the scheme (city,name) listing each city in located together with the name of the country that city is in.

b Write an RA query that returns the scheme (abbreviation,city,country,established) listing all details of organisations that are based in countries that are members of NATO.

c Consider the following RA query:

$$\pi_{\text{code}}\,\sigma_{\text{population}>10{,}000{,}000}\,\text{country} - (\pi_{\text{country as code}}\,\sigma_{\text{continent}=\text{'Europe'}}\,\text{encompasses} - \pi_{\text{code}}\,\sigma_{\text{population}>10{,}000{,}000}\,\text{country})$$

i) List the result of the query when run on the fragment of data on 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 that returns the scheme (organization) listing those organizations whose members include all those countries that are member of CERN.

i) RA

ii) SQL

iii) Datalog

e Give an RA query, that is equivalent to the following SQL query.

```
SELECT country.capital, country.code
FROM   country
UNION
SELECT organization.city, organization.country
FROM   organization
WHERE  city NOT IN (SELECT capital FROM country)
```

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 country.name,
       country.population
FROM   country
EXCEPT
SELECT country.name,
       country.population
FROM   country
       JOIN organization
       ON   organization.country=country.code
```

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 into an equivalent query that does not use the EXCEPT operator.

b Write an SQL query that returns the scheme (name,member,observer) listing the name of every country, together with the number of organisations that country is a member of, and the number it is an observer of.

c Write an SQL query that returns the scheme (name,neighbours,length) listing the name of every country, together with the number of neighbouring countries, and the total length of the border of the country.

d Write an SQL query that returns the scheme (continent,name,area,rank) listing continents, names of countries with some land area within the continent, the area of the country within the continent, and a ranking of that area within the continent (where the largest country is numbered one). The results must be returned in order of continent and rank.

e Write an SQL query that returns the scheme (organization,city,population) listing the organizations with at least 50 members that might have been established since 1950, together with the city in which the organisation is based, and the total population of all the member countries. The result should be returned in descending order of population.

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

3 a Suppose you have to design a new database to hold information about utility companies providing services to cities in different countries.

Each country will be identified by its ISO code, and have its official name. We record the name, area and population of cities in each country. Whilst the name of each city is unique within each country, the same city name may be used in more than one country. Each utility company is regulated by one country.

Services provided by utility companies are identified by their type (Gas, Water, Sewage, *etc*), and for each service we record a description of the service. We need to record which service is supplied by which utility company to which cities. In general, a service for a city may be supplied by more than one utility. For each supply of a service, we record the start date of the supply, and if it has ended, the end date of the supply. We also record the percentage of the city's current requirements of a service that are provided by the utility.

We record details of many companies, only some of which are utilities. Companies are identified by a registration number, and also have a name. For those companies that are utilities, we must also record the turnover and number of customers. We also record which utilities buy from which companies.

i) Design an ER schema to represent this new database.

ii) Map the ER schema you designed in (i) into a relational schema.

b The following histories describe the sequence of operations performed by three transactions.

$H_1 = r_1[c_{TR}], w_1[c_{TR}], r_1[c_{GB}], w_1[c_{GB}], c_1$
$H_2 = r_2[c_{CH}], r_2[c_{GB}], r_2[c_{CZ}], r_2[c_{TR}], c_2$
$H_3 = r_3[c_{GB}], r_3[c_{CH}], r_3[c_{CZ}], w_3[c_{CZ}], r_3[c_{TR}], w_3[c_{TR}], c_3$

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

$H_a = r_1[c_{TR}], w_1[c_{TR}], r_2[c_{CH}], r_2[c_{GB}], r_2[c_{CZ}], r_2[c_{TR}], c_2, r_1[c_{GB}], w_1[c_{GB}], c_1$

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

$H_b = r_3[c_{GB}], r_2[c_{CH}], r_2[c_{GB}], r_3[c_{CH}], r_2[c_{CZ}], r_2[c_{TR}], r_3[c_{CZ}], w_3[c_{CZ}], r_3[c_{TR}], w_3[c_{TR}], c_3, c_2$

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

$H_c = r_1[c_{TR}], r_3[c_{GB}], r_3[c_{CH}], r_3[c_{CZ}], w_3[c_{CZ}], r_3[c_{TR}], w_3[c_{TR}], w_1[c_{TR}], r_1[c_{GB}], c_3, w_1[c_{GB}], c_1$

iv) Give a concurrent execution of T_1, T_2, T_3 which produces a deadlock involving all three transactions, and draw a waits-for graph for the deadlock state.

The two parts carry equal marks.

4a Suppose that a relation $R(A, B, C, D, E, F, G, H)$ has the functional dependencies:
$S = \{ABGH \rightarrow BC, B \rightarrow BDEF, C \rightarrow H, E \rightarrow D, F \rightarrow BD, G \rightarrow AB\}$.

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.

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

UNDO	$w_4[c_R, \text{population} = 148,179,000]$
UNDO	$w_1[c_R, \text{population} = 148,180,000]$
UNDO	$w_1[c_{CH}, \text{population} = 7,207,060]$
UNDO	$w_4[c_{CH}, \text{population} = 7,208,999]$
UNDO	$w_2[c_B, \text{population} = 11,017,777]$
UNDO	$w_3[c_{ET}, \text{population} = 64,020,123]$
UNDO	$w_2[c_{CH}, \text{population} = 7,210,000]$
LOG	c_1
UNDO	$w_2[c_R, \text{population} = 150,100,000]$
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) Considering the time just after when c_1 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, 70%, and 30% of the marks.