CHALMERS UNIVERSITY OF TECHNOLOGY
Department of Computer Science and Engineering

Examination in Databases, TDA357/DIT620

Thursday 27 March 2008, 08:30-12:30, Building V

Examiner:	Graham Kemp (telephone 772 5411, room 6475 EDIT) The examiner will visit the exam room at 09:30 and 11:30.
Results:	Will be published by the middle of April at the latest.
Exam review:	Thursday 10 April 2008, see course web page for time and place http://www.cs.chalmers.se/Cs/Grundutb/Kurser/dbas/DbasHT2007/index.cgi
Grades:	Grades for Chalmers students (TDA357) are normally determined as follows: ≥ 48 for grade 5; ≥ 36 for grade 4; ≥ 24 for grade 3. Grades for GU students (DIT620) are normally determined as follows: ≥ 42 for grade VG; ≥ 24 for grade G.
Help material:	One A4 sheet with hand-written notes. You may write on both sides of that sheet. That sheet must be handed in with your answers to the exam questions. English language dictionaries are allowed.

Specific instructions:

- Please answer in English where possible. You may clarify your answers in Swedish if you are not confident you have expressed yourself correctly in English.
- Begin the answer to each question on a new page.
- Write clearly; unreadable = wrong!
- Fewer points are given for unnecessarily complicated solutions.
- Indicate clearly if you make any assumptions that are not given in the question.
- Write your personal number on every page.
- Write the page number and question number on every page.

Question 1. Consider the following domain description.

12 p

In a database of experimentally determined protein structures, each structure is identified by a 4-character code. Each structure consists of one or more protein chains, where each chain is uniquely identified within the structure by a one-character chain identifier. The length of each chain is recorded in the database. Each protein chain consists of many residues which, in turn, consist of several atoms. Within each residue the atoms have unique names (e.g. 'CA', 'OD1', etc.) and the x, y and z coordinates of the centre of each atom is stored in the database. Each residue in a protein chain is uniquely identified by its position within the chain (typically an integer value between 1 and 300). We want to know which residues are near to each other in space, so a 'neighbour' relationship is used for this purpose. Some of the structures in the database (called X-ray structures) have been determined by an experimental technique called X-ray crystallography. For these structures, and only these structures, we want to record a 'resolution' value, which is a real number that measures an important property of the experiment. Some of the other structures (called NMR structures) have been determined by a technique called NMR spectroscopy. Each NMR structure is determined from a set of experiments, each of which has a unique name (e.g. 'COSY', 'TOCSY', 'NOESY'). For each NMR structure, we want to record which kinds of experiment were used to determine that structure.

a) Draw an E-R diagram that correctly models this domain. (8p)

b) Translate this E-R diagram into a set of relations, clearly marking references and keys. (4p)

Question 2. 12 p

a) Suppose we have relation $R(A, B, C, D)$ and functional dependencies $AC \rightarrow D$, $D \rightarrow B$, $B \rightarrow A$.
By considering the closures of all subsets of attributes, find **all** non-trivial FDs, superkeys and keys. (4p)

b) Information about hotels and room bookings is stored in the relation:

$$R(name, city, address, parkingspaces, bookingnumber, roomnumber, arrival, departure, rate, guest, roomtype, roomstandard)$$

The following functional dependencies hold for this relation:

i) $name, city \rightarrow address, parkingspaces$
ii) $city, name, roomtype, roomstandard \rightarrow rate$
iii) $name, city, roomnumber \rightarrow roomtype$
iv) $name, city, roomnumber \rightarrow roomstandard$
v) $bookingnumber \rightarrow name, city, roomnumber$
vi) $bookingnumber \rightarrow arrival, departure, guestname$

Normalise this relation into BCNF.
Show each step in the normalisation process, and at each step indicate which functional dependency is being used. (4p)

c) Suppose relation $R(A, B, C, D)$ is as follows:

A	**B**	**C**	**D**
a1	b1	c1	d1
a1	b1	c1	d2
a1	b2	c2	d1
a1	b2	c2	d2

i) Explain, with reasons, which of the following multi-valued dependencies hold: $A \twoheadrightarrow B$, $A \twoheadrightarrow C$, $A \twoheadrightarrow D$, $A \twoheadrightarrow BC$, $A \twoheadrightarrow BD$, $A \twoheadrightarrow CD$ (2p)
ii) Decompose this relation to 4th normal form, and show the rows in the resulting relations. (2p)

Question 3. 7 p

a) Consider the following relations:

r:

A	B	C
1	x	k
2	y	k
3	y	m

s:

C	D
k	5
m	5
m	3

What is the result of evaluating the following relational algebra expressions? (3p)

i) $\pi_{B,D}(r \bowtie s)$

ii) $\sigma_{A=D}(r \times s)$

b) A car dealer uses a relational database to record information about car sales.

$Sales(day, month, year, \underline{regno}, price)$

Relation *Sales* contains the *day* (1-31), *month* (1-12) and *year* in which the car with registration number *regno* was sold for *price*.

i) Write a relational algebra expression that computes, for each month in year 2007, the average sale price during that month, and sorts the results by average sale price. (2p)

ii) Translate the relational algebra expression in your answer to part (i) to SQL. (2p)

Question 4. 8 p

The regional health service uses a relational database to record information about the doctors who work at the health centres in the region, and appointments that patients have with doctors. The database has the following relations:

$Doctors(\underline{person\_number}, health\_centre, city)$

$Appointments(\underline{appointment\_id}, patient, doctor, day, month, year)$
$doctor \rightarrow Doctors.person\_number$

Each tuple in relation *Doctors* contains a doctor's personal identification number, the name of the health centre where that doctor works and the name of the city where the health centre is located. Each tuple in relation *Appointments* contains a unique appointment identifier, the personal identification number of the patient, the personal identification number of the doctor, the day of the month (1-31), the month (1-12) and the year.

a) Write an SQL query that finds the names of the health centres where the patient with identification number 6006064444 had appointments in year 2000 or more recently. (2p)

b) Write an SQL query that finds the personal identification number of patients who had appointments with all of the doctors at the Gibraltargatan health centre in Göteborg in 2007. (2p)

c) Doctors are sometimes the patients of other doctors. Write an SQL query that finds the personal identification number of those doctors who have been the patients of doctors who work in the same city as themselves. (2p)

d) Create a view $AppointmentsPerMonth(health\_centre, month, year, num\_appointments)$ which gives the total number of appointments at each health centre during each month. (2p)

Question 5. The following relation schemas describe relations that will hold information about a university's employees, departments and programmes.
6 p

$Employees(\underline{persNum}, dept)$
$dept \rightarrow Departments.name$

$Departments(\underline{name}, head)$
$head \rightarrow Employees.persNum$

$Programmes(\underline{name}, ownedBy, director)$
$ownedBy \rightarrow Departments.name$
$director \rightarrow Employees.persNum$

Write SQL statements that create the relations with constraints as tables in a DBMS. Motivate the update and delete policies that you choose for attributes *head*, *ownedBy* and *director*.

Your solution should include code that checks the following constraints:

- the head of a department must be an employee of that department;
- each programme director must be employed in the department that owns the programme.

(6p)

Question 6. Suppose we have relation $Products(code, inStock, onOrder)$, and that this relation contains the following row:
4 p

code	inStock	onOrder
P123	5	0

Suppose that this row is modified by transactions T1 (selling two items and reordering 10 items if fewer than 5 items are in stock or on order) and T2 (accepting a delivery of new items) as follows:

T1:

$Step\ T1_A:\ get\ inStock\ of\ \text{“}P123\text{”}\ into\ X$
$Step\ T1_B:\ set\ inStock\ of\ \text{“}P123\text{”}\ to\ X-2$
$Step\ T1_C:\ get\ onOrder\ of\ \text{“}P123\text{”}\ into\ Y$
$Step\ T1_D:\ if\ X+Y<5\ set\ onOrder\ of\ \text{“}P123\text{”}\ to\ Y+10;$

T2:

$Step\ T2_A:\ get\ onOrder\ of\ \text{“}P123\text{”}\ into\ A$
$Step\ T2_B:\ get\ inStock\ of\ \text{“}P123\text{”}\ into\ B$
$Step\ T2_C:\ set\ inStock\ of\ \text{“}P123\text{”}\ to\ A+B$
$Step\ T2_D:\ set\ onOrder\ of\ \text{“}P123\text{”}\ to\ 0$

a) What are the possible outcomes if neither T1 nor T2 is run as a serializable transaction? Explain your answer. (2p)

b) What are the possible outcomes if T1 is run as a serializable transaction, but T2 is not? Explain your answer. (2p)

Question 7. Consider the following piece of XML.

6 p

```
<?xml version="1.0" standalone="no">

<Universities>
 <City name="Gothenburg">
  <University name="CTH" type="Technical" />
  <University name="GU"  type="Comprehensive" />
 </City>
 <City name="Stockholm">
  <University name="KTH" type="Technical" />
  <University name="SU"  type="Comprehensive" />
  <University name="KI"  type="Medical" />
 </City>
 <City name="Uppsala">
  <University name="UU"  type="Comprehensive" />
 </City>
</Universities>
```

a) Write a Document Type Definition (DTD) for this XML example. (2p)

b) Write XPath expressions that find:

- the names of all universities;
- all University elements for the technical universities. (2p)

c) Write an XQuery expression that gives the city name and university name of the technical universities, i.e.

```
<Result>Gothenburg: CTH</Result>
<Result>Stockholm: KTH</Result>
```

(2p)

Question 8. Suppose relation R contains the following rows:

5 p

a1	d	80
a1	k	20
a1	r	30
a2	m	60
a4	g	90
a5	d	60
a6	m	40
a7	c	80
a8	k	60
a8	s	30

a) Draw a picture that shows a *sparse primary index* on the first column of relation R. (2p)

b) Draw a picture that shows a *secondary index* on one of the other columns of relation R. (3p)