CHALMERS UNIVERSITY OF TECHNOLOGY
UNIVERSITY OF GOTHENBURG
Department of Computer Science and Engineering

Examination in Databases, TDA357/DIT620

Monday 13 April 2015, 14:00-18:00

Examiner:	Graham Kemp (telephone 772 54 11, room 6475 EDIT) The examiner will visit the exam room at 14:45.
Results:	Will be published by 5 May at the latest.
Exam review:	See course web page for time and place: http://www.cse.chalmers.se/edu/year/2014/course/TDA357/HT2014/
Grades:	Grades for Chalmers students (TDA357) are normally determined as follows: $\geq$ 48 for grade 5; $\geq$ 36 for grade 4; $\geq$ 24 for grade 3. Grades for GU students (DIT620) are normally determined as follows: $\geq$ 42 for grade VG; $\geq$ 24 for grade G.
Help material:	One A4 sheet with hand-written notes. You may write on both sides of that sheet. If you bring a sheet, it 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 the page number and question number on every page.

Question 1. Consider the following domain description.
12 p

An organisation arranges several running races (marathons, half-marathons, 10 km races, etc.) each year and wants to use a database to manage information about its races, the runners and their results. Each race is identified by its name (e.g. GöteborgsVarvet) and its year. The race date and distance will be stored for each race. A runner can enter many different races. The first time a runner uses the organisation's computerised entry system they are assigned a unique runner identification number, and they must provide the system with their name, year of birth and an e-mail address (for receiving race information). Each time they enter a race they must give the name of their running club (over time, a runner might change clubs several times, so the runner is asked to give the name of their current club on each occasion when they enter a race), and they are assigned a competitor number for that race. Competitor numbers are unique within a race, but different runners could have the same competitor number (e.g. number 1234) in different races. The competitor's fee status for the race (paid or not paid) should be recorded in the system. In each race the runners are assigned to start blocks, each of which is identified by its start block number. The start time for each start block should be stored in the system (e.g. start block 1 starts at 15:00, start block 2 at 15:10, etc.). The system needs to record which runners are in which start block for each race. The database will also be used to record information about runners' results, including split times. For example, in a 10 km race, times for each runner might be recorded as they cross the start line (their actual start time will typically be different to their block's start time, since not every runner in the same start block in a large race can cross the start line at the same second), at the half-way point, and at the finish line. Each time is recorded together with the distance at which it was recorded (for the 10 km race example just described, the split times will be at distances 0 km, 5 km and 10 km).

a) Draw an E-R diagram that correctly models this domain.
(6p)

b) Translate this E-R diagram into a set of relations, clearly marking all references and keys.
(6p)

Question 2. 10 p

a) Suppose we have relation $S(A, B, C, D)$ with functional dependencies $A \to B$, $A \to C$, $A \to D$, $CD \to A$.

i) Is relation S in BCNF? Explain your answer. (2p)

ii) Suppose we now add attribute E to relation S, giving relation $T(A, B, C, D, E)$, with the same four FDs: $A \to B$, $A \to C$, $A \to D$, $CD \to A$.
Is relation T in BCNF? Explain your answer. (2p)

iii) Are relations S and T in third normal form (3NF)? Explain your answer. (3p)

b) Suppose we have relation $R(A, B, C, D, E)$ with functional dependencies $A \to B$, $AC \to D$, $AC \to E$, $D \to C$.
Decompose relation R to BCNF. Show each step in the normalisation process, and at each step indicate which functional dependency is being used. Indicate keys and references for the resulting relations. (3p)

Question 3. 10 p

A database is used to store information about published articles, their authors, and citations between articles.

This database has the following relations:

Articles(articleId, title, journal, year, startPage, endPage)
ArticleAuthors(article, num, authorId)
Citations(cited, citedBy)

Each article is identified by its *articleId*. Attribute *title* is the title of the article and *journal* is the name of the journal where the article appears. Attributes *year*, *startPage* and *endPage* have integer values. Attribute *year* is the article's year of publication.

Each author has a *authorId*, and each author can be an author (or co-author) of several articles. Each article can have several co-authors, and integer-valued attribute *num* is used to record who is the first author, who is the second author, etc., in the article's author list. Relation *ArticleAuthors* has two keys.

Each article will normally cite several earlier articles (these are usually listed in a reference list at the end of an article). Relation *Citations* records which newer articles (attribute *citedBy*) cite which older articles (attribute *cited*).

a) Suggest references for these relations.
Write SQL statements that create these relations with constraints in a DBMS.
(4p)

b) An article cannot cite another article that was published more recently (i.e. the year of publication of the cited article cannot be greater than its own year of publication).
Write an assertion that checks this.
(2p)

c) Suppose view W is defined as follows:

```
CREATE VIEW W as
  SELECT authorId, article
  FROM   ArticleAuthors;
```

Implement a trigger that, when we delete a row from this view, removes any corresponding row from the ArticleAuthors table and adjusts the author number of any co-authors with higher author numbers than the deleted author.
(4p)

Question 4. 6 p

Assume the same relations as in Question 3:

Articles(articleId, title, journal, year, startPage, endPage)
ArticleAuthors(article, num, authorId)
Citations(cited, citedBy)

a) Write a relational algebra expression that finds all articles (article identifier and title) published in the journal "Nature" where author "1234" is an author.
(3p)

b) Write a relational algebra expression that finds the article identifier and title of each article, together with a count of the number of times that article has been cited.
(3p)

Question 5. Assume the same relations as in Question 3:

9 p

Articles(articleId, title, journal, year, startPage, endPage)
ArticleAuthors(article, num, authorId)
Citations(cited, citedBy)

a) Write an SQL query that finds all articles (article identifier and title) published in the journal "Nature" where author "1234" is an author.
(3p)

b) Define view SelfCitations(cited, citedBy) that contains occurrences where the cited and citedBy articles have at least one author in common.
(3p)

c) Write an SQL query that finds the article identifier and title of all articles that have never been cited.
(3p)

Question 6. A large store operates a customer loyalty scheme. Customers who are members of this scheme get a loyalty scheme card that is connected to their customer account number. There is one customer account per household/family, and each member of the household/family can have a loyalty scheme card that is connected to the same customer account (for example, Ann Smith and Bob Smith each have a loyalty scheme card that is connected to customer account number '987').

6 p

When a customer makes a purchase and presents their loyalty scheme card at the sales point, their customer account number is recorded in a database table together with information about the product purchased: *Purchases(productId, quantity, totalPrice, customerAccountNumber)*.

Sometimes the store makes special offers to members of the loyalty scheme. For example a child's bicycle (with productId 'p345') might be offered for a 20% discount, but the offer is limited to one bicycle per customer account. Information about offers made and whether they have been used are stored in relation *Offers(productId, customerAccountNumber, discount, used)*, e.g. ('p345', 987, 20, 'no').

At the sales point, the system checks the *Offers* table to see whether a discount can be applied and sets the *used* attribute to 'yes' before the customer completes the purchase.

Suppose Ann Smith and Bob Smith independently go to the store to buy a child's bicycle. They go to different sales points and, unaware of each other, each tries to buy a bicycle at a discounted price.

- Discuss whether both Ann and Bob could be able to buy a bicycle at a discounted price.
- Discuss whether there's a situation where only Ann buys a bicycle, but does not get the discounted price.

You should mention isolation levels in your answer.

(6p)

Question 7. Consider the following piece of XML:

7 p

```
<Question7>
 <Articles>
  <Article articleId="a1" title="t1" year="2013"/>
  <Article articleId="a2" title="t2" year="2012"/>
  <Article articleId="a3" title="t3" year="2013"/>
 </Articles>
 <Authors>
  <Author article="a1" num="1" name="Jones" />
  <Author article="a1" num="2" name="Green" />
  <Author article="a1" num="3" name="Black" />
  <Author article="a2" num="1" name="Jones" />
  <Author article="a3" num="1" name="Green" />
  <Author article="a3" num="2" name="Jones" />
 </Authors>
 <Citations>
  <Citation cited="a2" citedBy="a3" />
  <Citation cited="a1" citedBy="a3" />
  <Citation cited="a2" citedBy="a1" />
 </Citations>
</Question7>
```

a) Write a Document Type Definition (DTD) for the XML that is given above.
(2p)

b) Write an XPath expression that finds Citation elements where the citation is from article a3.
(1p)

c) Assuming that the XML shown above is in file *exam.xml*, write an XQuery expression that finds how many authors wrote article a1.
(1p)

d) Assuming that the XML shown above is in file *exam.xml*, write an XQuery expression that constructs the following result:

```
<Result>
  <Cites older="t2" newer="t3"/>
  <Cites older="t1" newer="t3"/>
  <Cites older="t2" newer="t1"/>
</Result>
```

(The result has one Cites element for each Citation element in *exam.xml*, and the attributes are the titles of the two related articles, not the article identifiers.)
(3p)