# Information page for written examinations at Linköping University

| | |
|---|---|
| **Examination date** | 2012-12-20 |
| **Room (2)**<br>If the exam is given in different rooms you have to attach an information paper for each room and mark intended place | U14 U15 |
| **Time** | 14-18 |
| **Course code** | TDDD43 |
| **Exam code** | TEN1 |
| **Course name**<br>**Exam name** | Datamodeller och databaser, avancerad kurs<br>Skriftlig tentamen |
| **Department** | IDA |
| **Number of questions in the examination** | 7 |
| **Teacher responsible/contact person during the exam time** | Patrick Lambrix / Fang Wei-Kleiner |
| **Contact number during the exam time** | 2605 / 4604 |
| **Visit to the examination room approx.** | 15:15; 16:45 |
| **Name and contact details to the course administrator**<br>(name + phone nr + mail) | Madeleine Häger Dahlqvist, madeleine.hager.dahlqvist@liu.se, 2360 |
| **Equipment permitted** | dictionary |
| **Other important information** | For pass: half of max points. |
| **Which type of paper should be used, cross-ruled or lined** | |
| **Number of exams in the bag** | |

Institutionen för datavetenskap
Linköpings universitet

# TENTAMEN
# TDDD43 Advanced Data Models and Databases

## December 20, 2012, 14-18

*Grades:* For a pass grade you need to obtain 50% of the total points.

*Instructions:* In addition to the instructions on the cover page:

- Write clearly.
- Start the answers to a question on a new page.
- If you make assumptions that are not given in a question, then clearly describe these assumptions. (Of course, these assumptions cannot change the exercise.)
- Give relevant answers to the questions. Points can be deducted for answers that are not answers to the question.
- Answer in English.

LYCKA TILL!

## 1. XML querying (3 + 2 = 5p)

Study the following XML file.

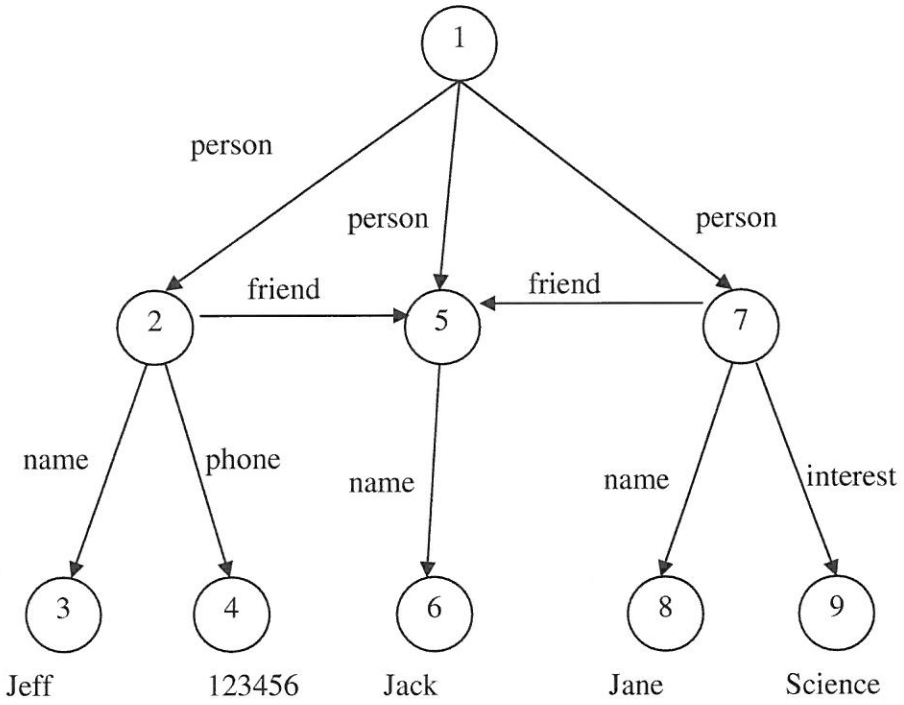
a) What is the result of executing the following XPath expressions on the XML file?
   i) /cities
   ii) //members//member
   iii) //countries[country/@id=3]

b) Express *"Find the population of the country whose capital's name is Stockholm"* as an XQuery query.

```
<?xml version="1.0" encoding="UTF-8"?>
<geodata>
  <countries>
    <country id = "1" name="Sweden" capital="c1">
      <population>9,540,065</population>
    </country>
    <country id = "2" name="Denmark" capital="c3">
      <members>
        <member>UN</member>
        <member>EU</member>
      </members>
    </country>
    <country id = "3" name="Germany" capital="c2">
      <population>81,726,000</population>
    </country>
  </countries>
  <cities>
    <city id="c1">
      <name>Stockholm</name>
    </city>
    <city id="c2">
      <name>Berlin</name>
    </city>
    <city id="c3">
      <name>Copenhagen </name>
    </city>
  </cities>
</geodata>
```

## 2. Data Guides (3p)

Draw a strong data guide for the data model below.




## 3. NoSQL databases (2p)

Consider the figure on the last page. Let P1, P2 and P3 be three processes and each of them maintain a vector clock with the initial value of (0,0,0).
Fill in the values of the vector clock of each process for the events illustrated with black points. Hand in the last page with your answer.

## 4. Information Retrieval (1 + 1 + 1 = 3p)

- Assume the boolean model for information retrieval. Assume we are interested in the words 'gene', 'enzyme', 'protein' and 'signal'.
    a. Show how to represent documents in the boolean model
    b. Represent the query for all documents containing gene or enzyme, but not protein. Show the completed DNF (disjunctive normal form) of the query
- Explain tf and idf in the vector model.

## 5. Ontology alignment (2 + 2 = 4p)

- Describe a framework for ontology alignment and explain the functionality of the different components.

- Give 4 different kinds of matchers. For each kind of matcher give an example and explain briefly what it does.

## 6. Description logics (1 + 2 = 3p)

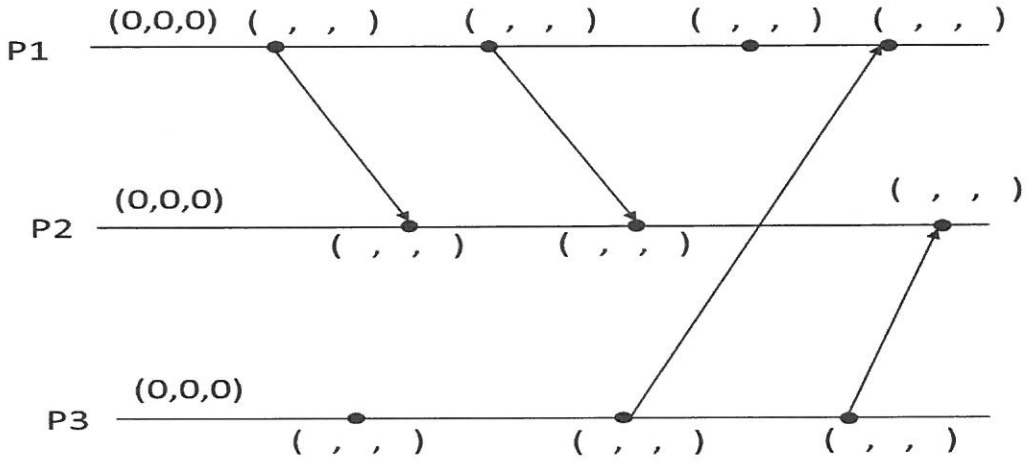
- Define the semantics of
  - *FORALL WorksAt.Company*
  - *EXISTS WorksAt.Company*

- Does *EXISTS WorksAt.Company* subsume *FORALL WorksAt.Company*?
  Prove your answer using a tableau algorithm.

## 7. Integration (4p)

Data source 1 has information about used cars and contains model, year, price, seller information. Data source 2 contains information about luxury cars for which the price is at least 500 000 euro and contains model, year, price, seller information. Data source 3 contains reviews for cars based on the model and the year. Assume the global schema defines the following relations: Car(model, price, year) and Car-review(model, year, review)

- Give local as view mappings for all data sources with respect to the given global schema.
- Describe the query: Return reviews for cars from 2010 with price lower than 10 000 euro.
- Apply then the bucket algorithm to generate the queries to the local data sources.


P1
(0,0,0)
( , , )
( , , )
( , , )
( , , )
P2
(0,0,0)
( , , )
( , , )
( , , )
P3
(0,0,0)
( , , )
( , , )
( , , )