

Institutionen för datavetenskap  
Linköpings universitet

# TENTAMEN

## TDDD43 Advanced Data Models and Databases

March 26, 2013, 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)

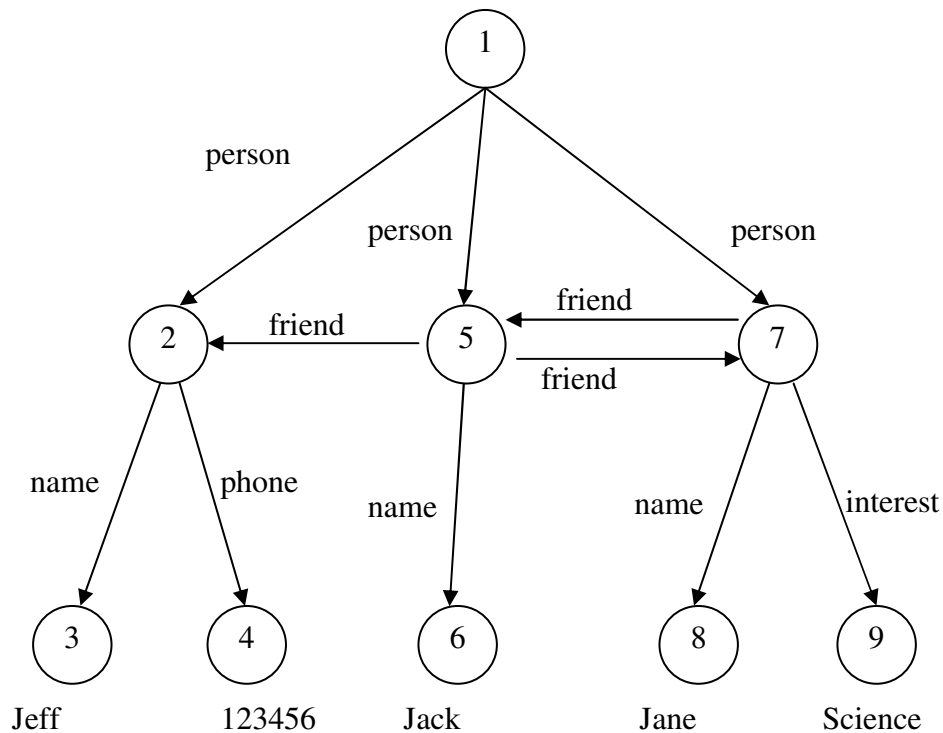
Study the following XML file.

- a) What is the result of executing the following XPath expressions on the XML file?
  - i) `//country[@name="Denmark"]/members`
  - ii) `/geodata/name`
  - iii) `//countries[country/@id=3]`
- b) Express “*Find the population of the country whose capital name is Berlin*” 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 and enzyme, but not protein. Show the completed DNF (disjunctive normal form) of the query

- Explain tf and idf in the vector model.

## 5. Description logics (2 + 3 = 5p)

- (i) Represent the following information in description logics: *LFC is a football team. A football team is a team with at least 11 members and all members are football players. Teams that have at least 10 members are large teams. Lotta Rohlin is a member of LFC.*
- (ii) Given a domain of objects  $D = \{P1, P2, P3, P4, P5, P6, P7, P8, P9, P10, P11, P12, P13, P14, P15, T1, T2, T3\}$ . Define an interpretation function that together with  $D$ , gives an interpretation that is a model for the knowledge base you created in (i).

## 6. Integration (2p)

Compare the global as view approach versus the local as view approach for integration of data sources regarding the following issues:

- easy/hard to add new data sources
- easy/hard query processing

Explain your answer for each issue with an example.

## 7. Ontology alignment (1 + 2 = 3p)

- Explain the notion (what is it?) of ontology alignment and why we want to do this.
- Given the concept 'cell death' in ontology O1 and the concept 'apoptosis' in ontology O2. What kinds of matchers would give this pair of concepts a high similarity value and in which cases? Discuss the situations. You may make assumptions about the content, structure etc. of the ontologies.

