Institutionen för datavetenskap
Linköpings universitet

TENTAMEN

TDDD43 Advanced Data Models and Databases

January 14, 2016, 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 + 2 = 7p)

Study the following XML file:

a) What is the result of executing the following XPath expressions on the XML file?
   i) /players
   ii) //teams//item
   iii) //persons[person/@id=4]

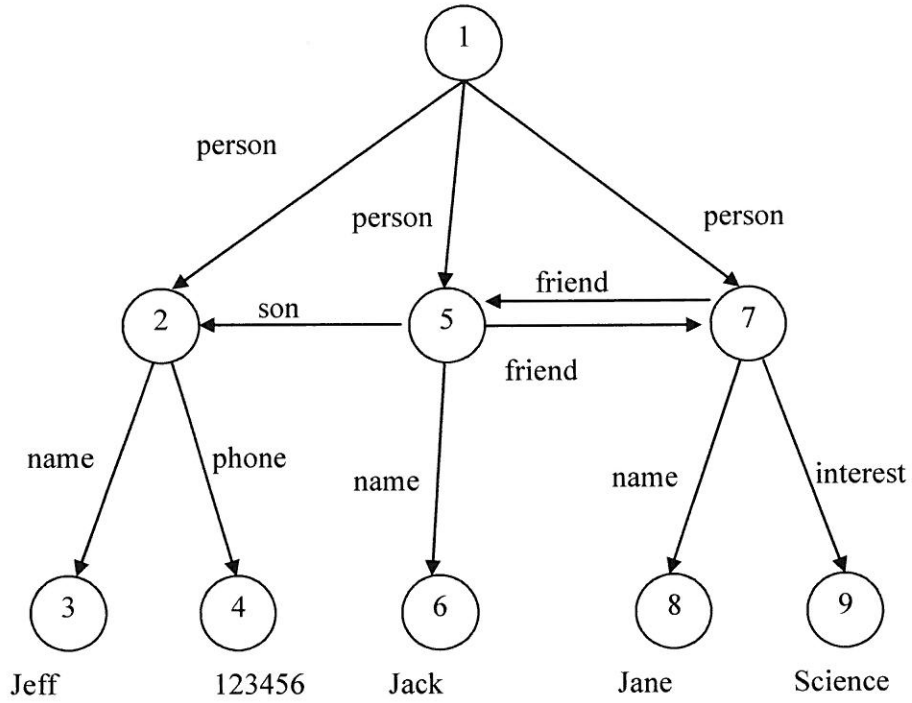
b) Express *"Find the name of everyone who plays in LFC."* as an XQuery query.

c) Model the data in the XML file using RDF.

```
<?xml version="1.0" encoding="UTF-8"?>
<Linköpingsportslist>
  <persons>
    <person id="1" name="Magnus" job="hockey" tool="stick" />
    <person id="2" name="Jonas" job="hockey" tool="puck"/>
    <person id="3" name="Lotta" job="football" tool="ball"/>
    <person id="4" name="Kim" job="hockey" tool="mask"/>
  </persons>
  <teams>
    <item id="1" type="LHC">
        <usable-tool>stick</usable-tool>
        <usable-tool>puck</usable-tool>
        <usable-tool>mask</usable-tool>
   </item>
    <item id="2" type="LHC Dam">
        <usable-tool>stick</usable-tool>
        <usable-tool>puck</usable-tool>
        <usable-tool>mask</usable-tool>
   </item>
    <item id="3" type="LFC">
        <usable-tool>ball</usable-tool>
    </item>
  </teams>
  <players>
    <plays-in who="1" what="1" when="2014-2015"/>
    < plays-in who="2" what="1" when="2014-2015"/>
    < plays-in who="3" what="3" when="2015"/>
    < plays-in who="4" what="2" when="2014-2015"/>
  </players>
</Linköpingsportslist>
```

2. Data Guides (3p)

Draw a strong data guide for the data model below.



3. NoSQL databases (2p)

Explain how the matrix-vector multiplication task can be solved with the Map-Reduce workflow. Describe the operations performed during the Map and Reduce steps together with the inputs to and outputs from them.

4. **Information Retrieval (2 + 2 + 2 = 6p)**

   a. Assume that we use the boolean model for information retrieval. Assume that we are only interested in the words 'gene', 'enzyme', 'protein' and 'signal'.

      1. Give the query that represents all documents that contain 'gene', but not 'signal'.
      2. Compute the completed DNF (disjunctive normal form) of the query - make sure to show all steps in the computation.

   b. Assume that we use the vector model for information retrieval. Assume that we are only interested in the words 'gene', 'enzyme', 'protein' and 'signal'. Assume that we have two documents in our document base. Document 1 contains 'enzyme' 5 times, 'gene' 10 times, 'protein' 0 times and 'signal' 8 times. Document 2 contains 'enzyme' 0 times, 'gene' 0 times, 'protein' 7 times and 'signal' 1 time.

      1. Explain tf and idf in the vector model.
      2. Give the document representations for Document 1 and Document 2 according to the tf-idf model.

   c. Compare the boolean and vector models for information retrieval regarding D, Q, F and R (document model, query model, framework and ranking).

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

   a. Define the semantics of:

      - *EXISTS PlaysAt.Team*
      - *FORALL PlaysAt.Team*
      - *Employee AND Team-member*
      - *ATLEAST 3 PlaysAt*

   b. Does the following hold?

      Employee AND FORALL PlaysAt.Team
      is-a (is subsumed by)
      Employee AND EXISTS PlaysAt.Team

      Prove your answer using a tableau algorithm.

   c. Given the following statement:

      Teacher(Patrick, databases)

      What is the result of the following queries using open, respectively closed world semantics:

      - Teacher(Patrick, databases)?
      - Teacher(Patrick, data mining)?

6. **Ontologies (2 + 2 + 2 = 6p)**

   a. Give and explain 4 OBO Foundry principles.

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

   c. Give the different kinds of defects that can occur in an ontology and exemplify.