Information page for written examinations at Linköping University

Examination date	2016-08-27
Room (1)	TER2
Time	8-12
Course code	TDDD43
Exam code	TEN1
Course name Exam name	Advanced Data Models and Databases (Datamodeller och databaser, avancerad kurs) Written examination (Skriftlig tentamen)
Department	IDA
Number of questions in the examination	8
Teacher responsible/contact person during the exam time	Patrick Lambrix
Contact number during the exam time	2605
Visit to the examination room approximately	9:30, 11:00
Name and contact details to the course administrator (name + phone nr + mail)	Madeleine Häger Dahlqvist, 282360 madeleine.hager.dahlqvist@liu.se
Equipment permitted	Dictionary
Other important information	
Number of exams in the bag	

TENTAMEN

TDDD43 Advanced Data Models and Databases

August 27, 2016, 8-12

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)   //players
        ii)   /teams//item
        iii)   //teams[item/@id=2]
    b) Express *"Find the name of everyone who plays in LHC Dam."* as an XQuery query.

```xml
<?xml version="1.0" encoding="UTF-8"?>
<Linköpingsportslist>
  <persons>
    <person  id="1"  name="Niklas" job="hockey" tool="stick" />
    <person  id="2"  name="Sebastian" job="hockey" tool="puck"/>
    <person  id="3"  name="Stina" job="football"  tool="ball"/>
    <person  id="4"  name="Florence" 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="2016-2017"/>
    < plays-in who="2" what="1" when="2016-2017"/>
    < plays-in who="3" what="3" when="2016"/>
    < plays-in who="4" what="2" when="2016-2017"/>
  </players>
</Linköpingsportslist>
```

2. NoSQL databases (2 + 1 = 3p)

P1, P2 and P3 are three distributed processes. The events following below have occurred during the processes and the values for their vector clocks are given:
P1: A (0,0,0); B (1,0,0); C (2,0,0); D (3,0,0); E (4,0,2)
P2: F (0,0,0); G (1,1,0); H (2,2,0); I (2,3,3)
P3: J (0,0,0); K (0,0,1); L (0,0,2); M (0,0,3)

a. Illustrate the ordering of the events in each process and the communication between the processes (sent and received messages) on a figure where a separate axis represents the physical time for every process; the time increases from left to right.

b. Give an example of concurrent events, (write down a pair of events which are concurrent and explain why they are concurrent).

3. Linked Data (0.5 + 0.5 + 1 = 2p)

a. The data model for RDF represents data as triples (subject, predicate, object). What is stored in subject, predicate and object, respectively?

b. What are URIs?

c. Given the following SPARQL query, describe the result of the query in natural language.

```
SELECT ?name
FROM <http://ida.liu.se/LHC-DB>
WHERE
      ?x <http://example.org/definitions/plays-for> <http://example.org/definitions/Linköping-Hockey-Club>
      ?x <http://example.org/definitions/has-team-role> <http://example.org/definitions/forward>
      ?x <http://example.org/definitions/has-name> ?name
```

4. Description logics (1 + 2 + 1 = 4p)

a. Give the semantics of the union and exists operators. (Set notation!)

b. Is *FORALL eats.plant* subsumed by (is-a) *EXISTS eats.plant*?
   Prove your answer using a tableau algorithm.

c. When is a concept C satisfiable with respect to a Tbox T?
   When are two concepts C and D disjoint with respect to a Tbox T?

5. Information Retrieval (2 + 2 = 4p)

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.

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

        1. Explain tf and idf in the vector model.

        2. In which cases is a weight w_{ij} in a document vector equal to 0?

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

    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'.

        1. Give the document representations for Document 1 and Document 2 according to the Boolean model.

        2. Represent the query for all documents containing gene or protein, but not signal. Compute then the completed DNF (disjunctive normal form) of the query.

6. Ontologies (2 + 1 = 3p)

    a. Describe from a knowledge representation point of view the different kinds of ontologies. For each kind, state which ontology components are represented.

    b. Give an example of a modeling defect and an example of a semantic defect in ontologies.

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

    a. Define the notions of precision and recall in the context of ontology alignment.

    b. What is the influence on precision and recall for matchers based on string matching? Explain why.

    c. Give an example of a structure-based matcher and an example of an instance-based matcher. Explain what they do.

8. Data Guides (3p)

Draw a strong data guide for the data model below.