Information page for written examinations at Linköping University

Examination date	2017-01-12
Room (2)	U7(14) U10(20)
Time	14-18
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, (Olaf Hartig, Valentina Ivanova)
Contact number during the exam time	2605
Visit to the examination room approximately	15:30, 17:00
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	

Information page for written examinations at Linköping University

Examination date	2017-01-12
Room (2)	U7(14) U10(20)
Time	14-18
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, (Olaf Hartig, Valentina Ivanova)
Contact number during the exam time	2605
Visit to the examination room approximately	15:30, 17:00
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	

Institutionen för datavetenskap
Linköpings universitet

TENTAMEN

TDDD43 Advanced Data Models and Databases

January 12, 2017, 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 document, respectively?

i) /persons

ii) //activities//item[@id='2']/usable-tool

iii) //activities/item[usable-tool/text()='spoon']/usable-tool/text()

b) Express the following query using XQuery: "*Find the tools used by all persons who have booked an activity for which a dagger is one of the usable tools.*" (The only constant that is allowed to be used in the query is the string "dagger").

```
<?xml version="1.0" encoding="UTF-8"?>
<activitylist>
  <persons>
    <person id="1" name="Frodo" family="hobbit" tool="dagger" />
    <person id="2" name="Samwise" family="hobbit" tool="spoon"/>
    <person id="3" name="Aragorn" family ="human" tool="sword"/>
    <person id="4" name="Legolas" family ="elf" tool="bow"/>
    <person id="5" name="Gimli" family ="dwarf" tool="ax"/>
  </persons>
  <activities>
    <item id="1" type="Chopping things">
        <usable-tool>dagger</usable-tool>
        <usable-tool>axe</usable-tool>
       <usable-tool>sword</usable-tool>
   </item>
    <item id="2" type="Cooking">
       <usable-tool>dagger</usable-tool>
       <usable-tool>spoon</usable-tool>
   </item>
    <item id="3" type="Jewelry">
    </item>
  </activities>
  <bookings>
    <booking who="1" what="3" when="2016-03-31"/>
    <booking who="2" what="2" when="2016-03-31"/>
    <booking who="5" what="1" when="2016-04-01"/>
    <booking who="3" what="1" when="2016-04-02"/>
  </bookings>
</activitylist>
```

2. NoSQL databases (2p)

Explain how vector clocks work and what problems they solve in connection to NoSQL systems. Give examples (use a figure) of causally related and concurrent events and explain how to determine the relationship between the events.

3. RDF, SPARQL, Property Graphs (1 + 2 = 3p)

Consider the following set of RDF triples (prefix declarations omitted).

```
ex:cid651  rdf:type  ex:CourseOccasion .
ex:cid651  ex:code   "TDDD12" .
ex:cid651  ex:year   2014 .
ex:cid651  ex:instructor   ex:AdamSmith .

ex:cid337  rdf:type  ex:CourseOccasion .
ex:cid337  ex:code   "TDDD12" .
ex:cid337  ex:year   2017 .

ex:cid810  rdf:type  ex:CourseOccasion .
ex:cid810  ex:code   "TDDD37" .
ex:cid810  ex:year   2014 .
ex:cid810  ex:instructor   ex:EvaHegen .

ex:cid411  rdf:type  ex:CourseOccasion .
ex:cid411  ex:code   "TDDD37" .
ex:cid411  ex:year   2013 .
ex:cid411  ex:instructor   ex:AdamSmith .

ex:AdamSmith  ex:name  "Adam Smith" .
ex:EvaHegen   ex:name  "Eva Hegen" .
```

a) What is the result of evaluating the following SPARQL query (prefix declarations omitted) over the given set of RDF triples? Represent the results in a tabular form.

```
SELECT ?x ?c WHERE {
 ?x ex:code ?c
 ?x ex:year ?y
 FILTER ( ?y < 2014 )
}
```

b) Draw a Property Graph that captures the same information as represented by the given set of RDF triples. Every RDF triple whose object is a literal must be converted into a vertex property associated with the vertex created for the subject of the triple (hence, literals should not be converted into separate vertexes of the Property Graph).

4. Information Retrieval (2p)

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. 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.

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

a. Represent the following information in description logics: LHC is an ice hockey team. An ice hockey team is a team with at least 22 members and all members are ice hockey players. Teams that have at least 10 members are large teams. Niklas Persson is a member of LHC.

b. Is it true that an animal that eats meat is always an animal that only eats meat? Prove your answer using a tableau algorithm.

6. Ontologies (2 + 1 + 1 + 1 + 1= 6p)

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.

c. Define the notions of precision, recall and f-measure in the context of ontology alignment.

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

e. Give an example of a string-based matcher and an example of a constraint-based matcher. Explain what they do.

7. Data source integration (3p)

Data source 1 has information about comic books and contains information about the series, volume, year, price, and seller information. Data source 2 contains information about rare comic books for which the price is at least 5000 euro and contains series, volume, year, price, and seller information. Data source 3 contains reviews for comic books based on the series and volume. Assume the global schema defines the following relations: Book(series, volume, year, price) and Book-review(series, volume, review).

1. Give local as view mappings for all data sources with respect to the given global schema.
2. Describe the query: Return reviews for comic books from 1979 with price lower than 1000 euro.
3. Apply then the bucket algorithm to generate the queries to the local data sources.

8. Data Guides (2p)

Draw a strong data guide for the data model below.