Information page for written examinations at Linköping University

Examination date	2019-08-31
Room (1)	TER3(1)
Time	8-12
Edu. code	TDDD43
Module	TEN1
Edu. code name **Module name**	Advanced Data Models and Databases (Datamodeller och databaser, avancerad kurs) Written examination (Skriftlig tentamen)
Department	IDA
Number of questions in the examination	9
Teacher responsible/contact person during the exam time	Patrick Lambrix, (Olaf Hartig)
Contact number during the exam time	2605
Visit to the examination room approximately	10:00
Name and contact details to the course administrator (name + phone nr + mail)	
Equipment permitted	Dictionary
Other important information	
Number of exams in the bag	

Institutionen för datavetenskap
Linköpings universitet

TENTAMEN

TDDD43 Advanced Data Models and Databases

August 31, 2019, 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 (1 + 1 + 2 = 4p)

Consider the following XML document:

```
<?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="axe"/>
    </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>
```

a) What is the result of executing the following XPath expression on the XML document?

//item[@type='Chopping things']/usable-tool

b) For an XML document such as the given one, write an XPath expression that returns the name of every person that is of the family "hobbit".

c) For an XML document such as the given one, express the following query using XQuery: Find the name of every person who booked a "Cooking" activity. (The only constant that is allowed to be used in the query is the string "Cooking")

2. RDF, SPARQL (1 + 1 = 2p)

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.

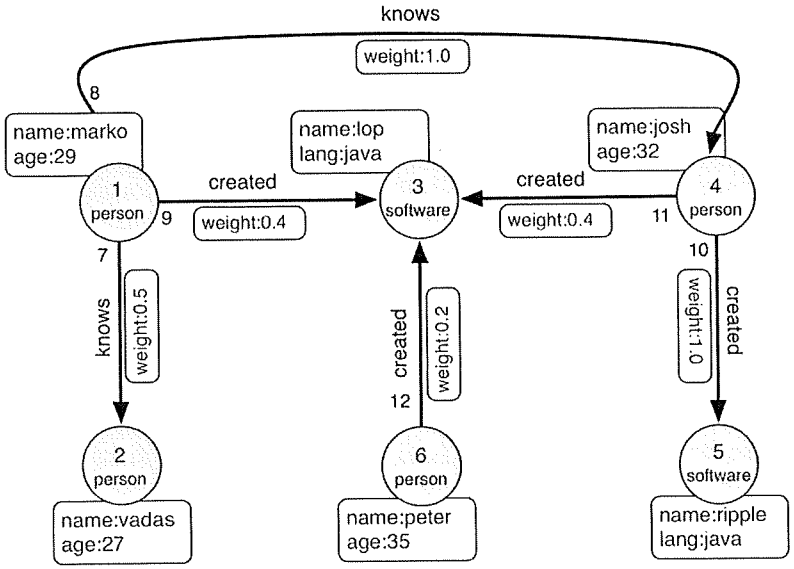
```
SELECT ?x ?n WHERE {
  ?x ex:instructor ?i .
  ?i ex:name ?n
  ?x ex:year ?y .
  FILTER ( ?y < 2014 )
}
```

b) Write another SPARQL query that lists all the years in which the course TDDD12 was given. Hence, the result of this query over the given set of RDF triples should be the following (single-column) table:

```
  ?y
--------
2014
2017
```

3. Graph data (1 + 1 = 2p)

a) Consider the Property Graph as illustrated in the following figure.



What is the result of evaluating the following Gremlin expression over this Property Graph?

```
g.V().has('age','32').out().values('name')
```

b) Name a reason why the MapReduce system is not suitable for complex, iterative graph algorithms such as the PageRank algorithm.

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

a) What is meant by "read scalability"?

b) Revisit the set of RDF triples given above. Represent all of this data as a key-value database.

5. Information Retrieval (2 + 0.5 + 0.5 = 3p)

Assume that we use the vector model for information retrieval. Assume that we are only interested in the words 'dog', 'mouse', 'pig' and 'rooster'. Assume that we have two documents in our document base.

- Document 1 contains 'dog' 5 times, 'mouse' 10 times, 'pig' 0 times and ' rooster' 8 times.
- Document 2 contains 'dog' 0 times, 'mouse' 0 times, 'pig' 7 times and 'rooster' 1 time.

a. Give the document representations for Document 1 and Document 2 according to the vector model **and** show/explain how you computed them.

b. In which cases is a weight wij in a document vector equal to 0?

c. What is the ranking between Document 1 and Document 2 for the query with representation (0,0,1,1)?

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

a. 1. Is it true that 'A or B' is subsumed by 'A and B'? Prove your answer using a tableau algorithm.

2. Is it true that 'A or B' subsumes 'A and B'? Prove your answer using a tableau algorithm.

b. Given the following statement: Teacher(Patrick, databases)

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

1. Teacher(Patrick, databases)?
2. Teacher(Patrick, data mining)?

c. Explain the notions of subsumption and satisfiability. Can subsumption be reduced to (un)satisfiability? Explain your answer.

7. Ontologies (2 + 2 = 4p)

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 4 different kinds of matchers for ontology alignment. For each kind of matcher give an example and explain briefly what it does.

8. Data source integration (3p)

Data source 1 has information about used comics and contains title, series, year, price, seller information. Data source 2 contains information about rare comics for which the price is at least 5 000 euro and contains title, series, year, price, seller information. Data source 3 contains reviews for comics based on the title, series and the year. Assume the global schema defines the following relations: Comics(title, series, price, year) and Comics-review(title, series, year, review)

a. Give local as view mappings for all data sources with respect to the given global schema.

b. Describe the query: Return reviews for comics from 1956 with price lower than 100 euro.

c. Apply then the bucket algorithm to generate the queries to the local data sources.

9. Data Guides (2p)

Draw a strong data guide for the data model below.