Institutionen för datavetenskap
Linköpings universitet

TENTAMEN

TDDD43 Advanced Data Models and Databases

January 11, 2019, 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 (1 + 1 + 2 = 4p)

Consider the following XML document:

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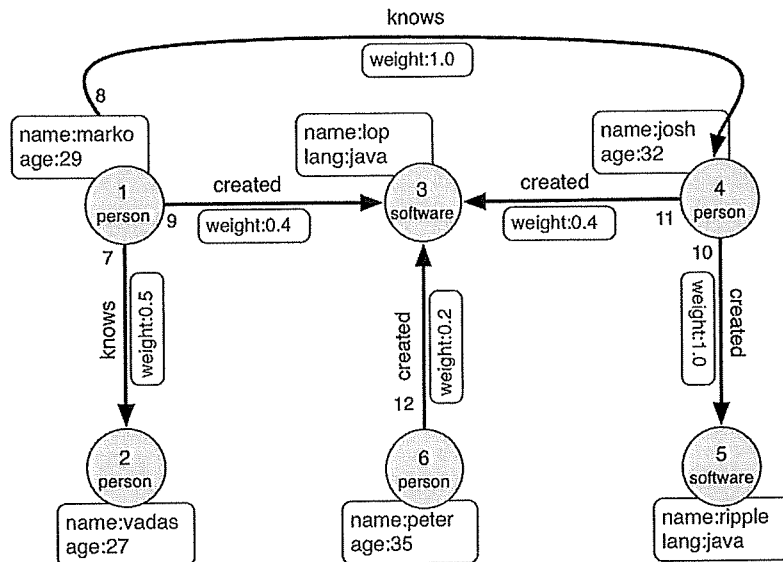
```
SELECT ?x ?n WHERE {
   ?x ex:instructor ?i .
   ?i ex:name ?n
   ?x ex:year ?y .
   FILTER ( ?y < 2014 )
}
```

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 (1 + 1.5 + 0.5 = 3p)

Assume that we use the boolean model for information retrieval. Assume that we are only interested in the words 'gene', 'enzyme', 'protein' and 'signal'. Assume that we have four 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' 6 times and 'signal' 1 time.

* Document 3 contains 'enzyme' 0 times, 'gene' 1 times, 'protein' 6 times and 'signal' 0 times.

* Document 4 contains 'enzyme' 0 times, 'gene' 0 times, 'protein' 5 times and 'signal' 0 times.

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

b. Represent the query for all documents containing enzyme or protein, but not signal. Compute DNF (disjunctive normal form) as well as the completed DNF of the query.

c. Give the result and ranking for the query in question b.

6. Description logics (1.5 + 1.5 + 2 = 5p)

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. Sebastian Karlsson is a member of LHC.

b. Give an interpretation for the knowledge base you created in question *a*.

c. Is it true that a person who plays ice hockey is always a person who only plays ice hockey? Prove your answer using a tableau algorithm.

7. Ontologies (1 + 1 + 2 = 4p)

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

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

c. Given an ontology with the following axioms:
A is sub-concept of B; B is sub-concept of C; A and C are disjoint.
Does the ontology contain a semantic defect? If so, which and how do you repair? If not, explain why there is no defect.

8. Data source integration (3p)

Compare the Global As View approach to 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 and *give an example* for each issue.

9. Data Guides (2p)

Draw a strong data guide for the data model below.