

Försättsblad till skriftlig tentamen vid Linköpings universitet



Datum för tentamen	2016-03-31
Sal (2)	G33 <u>G36</u>
Tid	8-12
Kurskod	TDDD43
Provkod	TEN1
Kursnamn/benämning Provnamn/benämning	Datamodeller och databaser, avancerad kurs Skriftlig tentamen
Institution	IDA
Antal uppgifter som ingår i tentamen	8
Jour/Kursansvarig Ange vem som besöker salen	Patrick Lambrix
Telefon under skrivtiden	2605
Besöker salen ca klockan	9:30, 11:00
Kursadministratör/kontaktperson (namn + tfnr + mailaddress)	Madeleine Häger Dahlqvist, 282360 madeleine.hager.dahlqvist@liu.se
Tillåtna hjälpmedel	Dictionary
Övrigt	
Antal exemplar i påsen	

Institutionen för datavetenskap
Linköpings universitet

TENTAMEN

TDDD43 Advanced Data Models and Databases

March 31, 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) /bookings
 - ii) //activities//item
 - iii) //persons[person/@id=1]
- b) Express "Find the name of everyone booked at chopping things." as an XQuery query.

```
<?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>ax</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 consistent hashing works and what problems it solves in connection to NoSQL systems.

3. Linked Data (2p)

- a. What is the data model for RDF?
- 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/personnelDB>
WHERE
  ?x <http://example.org/definitions/teaches> <http://example.org/definitions/database-course>
  ?x <http://example.org/definitions/has-name> ?name
```

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

- 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'. 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.
- b. 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. 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. Given a domain of objects $D = \{X_1, X_2, X_3, \dots, X_{49}, X_{50}\}$. Define an interpretation function that together with D, gives an interpretation that is a model for the knowledge base you created in (a).
- c. Why is it problematic to define the semantics of a cyclic Tbox? How can it be solved?

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

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

- a. Define the notions of precision, recall and f-measure 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 string-based matcher and an example of a constraint-based matcher. Explain what they do.

8. Data Guides (3p)

Draw a strong data guide for the data model below.



