

Written exam in

TDDD43 Databases and Data Models Advanced course

Location: G33, G35, G36

Date: 2009-12-18

Time: 14-18

Equipment: None

Requirements:

15 points is required for passing the exam

Approx. 20 points is required for grade 4

Approx. 25 points is required for grade 5

The results is announced within two weeks

Teacher in charge: Lena Strömbäck, 0709-39 6776, 013-28 23 24 ,
Patrick Lambrix, 282605

Use one sheet for each question.

Write your AID number on each sheet.

Only write on one side of each sheet.

Give relevant answers to the question.

Points can be deducted for answers that are not answers to the question.

1. XML modeling and querying (5p)

The XML documents at the end of this exam show two ways of modeling Santa's gift records.

- a) Compare the two models and for each of them, state at least one reason or situation where it is preferable over the other. (2p)
- b) Express "*Who has got any gift from Lena*" as an XPath query for the second example model. (1p)
- c) Express "*Who has got any gift from Lena*" as an XQuery query for the first example model. (2p)

2. XML technologies (2p)

What is XSLT? What can it be used for? Give one example of the use of XSLT. You only need to exemplify in general terms, not give any examples of code.

3. XML storage (4p)

- a) What are the main ideas behind Native XML storage? (1p)
- b) What are the main characteristics behind shredding XML into a RDBMS? (1p)
- c) Give an example shredding for the first example of Santa's gift list used in exercise 1. (2p)

4. Data Guides (3p)

In figures 1 and 2 at the end of this exam (simplified and slightly modified) data from a SwissProt data source is given. (The dashed nodes in the figures refer to nodes in the other figure. Both figures together make up the data source.) Draw a strong data guide for this data. Use the OEM model.

5. OODBMS modeling (4p)

In a hospital information about patients is stored. This information contains the patient record with information about the patient's diseases and treatments, the letters of doctors requesting lab tests, the reports of the lab tests, communications (letters and e-mails) between health personnel concerning the patient, and transcripts of telephone meetings.

Health personnel consists of doctors and lab personnel. Doctors can read and write all the patient information. Lab personnel can read the requests for lab tests and read and write the lab test reports. Communications can be read and written by all health personnel. Secretaries can read and write transcripts of telephone meetings.

In addition, the hospital has policy documents. These can be read by all employees of the hospital.

- a) Draw the subject, authorization object and authorization type hierarchies. (1p)
- b) Draw an authorization model using implicit/explicit, weak/strong and positive/negative authorizations for each of the following: doctors, nurses, secretaries. Assume that when a positive authorization is not stated in the text that the authorization is not given. (3p)

6. Ontologies (2p)

Describe the 4 components of ontologies and give examples for each.

7. Description logics (4p)

Define the following concepts using description logics:

C1: team for which all members are IDA-employees (1p)

C2: team for which there exists a member that is an IDA-employee (1p)

Is there an *is-a* relation between C2 and C1, i.e. C2 *isa* C1? Prove your answer using a tableau algorithm. (2p)

8. Integration (4p)

Discuss the global as view approach versus the local as view approach regarding the mappings between the global model and the content of the local sources as well as regarding query processing. Exemplify all notions (global model, mapping – LAV and GAV, query processing LAV – GAV) with examples. You can (but don't have to) use the examples given in figure 1-4 at the end of this exam in your examples.

9. Ontology alignment (2p)

Explain the notion (what is it?), and the basic intuitions (why do this?) and ideas (how to do this?) behind the double threshold filtering technique in ontology alignment.

```

<santasGiftList>
  <persons>
    <person id="1" name="Ludvig" age="7" />
    <person id="2" name="Filip" age="12" />
    <person id="3" name="Lisa" age="2" />
    <person id="4" name="Lena" age="40" />
  </persons>
  <items>
    <item id="1" type="Doll">
      <description>A nice little doll for small children</description>
    </item>
    <item id="1" type="Lego toolbox">
      <description>For the more advanced builder</description>
      <ageRecommendation>12</ageRecommendation>
      <varning>Contain small parts. Not suitable for small children</varning>
    </item>
  </items>
  <gifts>
    <gift from="4" to="3" what="1"/>
    <gift from="4" to="1" what="2"/>
    <gift from="2" to="4" what="2"/>
  </gifts>
</santasGiftList>

```

Example 1: Santas gift List version 1

```

<santasGiftList>
  <giver id="4" name="Lena" age="40">
    <gift>
      <item id="1" type="Doll">
        <description>A nice little doll for small children</description>
      </item>
      <to id="3" name="Lisa" age="2" />
    </gift>
    <gift>
      <item id="1" type="Lego toolbox">
        <description>For the more advanced builder</description>
        <ageRecommendation>12</ageRecommendation>
        <varning>Contain small parts. Not suitable for small children</varning>
      </item>
      <to id="3" name="Ludvig" age="7" />
    </gift>
  </giver>
  <giver id="2" name="Filip" age="412">
    <gift>
      <item id="1" type="Lego toolbox">
        <description>For the more advanced builder</description>
        <ageRecommendation>12</ageRecommendation>
        <varning>Contain small parts. Not suitable for small children</varning>
      </item>
      <to id="4" name="Lena" age="40" />
    </gift>
  </giver>
</santasGiftList>

```

Example 2: Santas gift list version 2

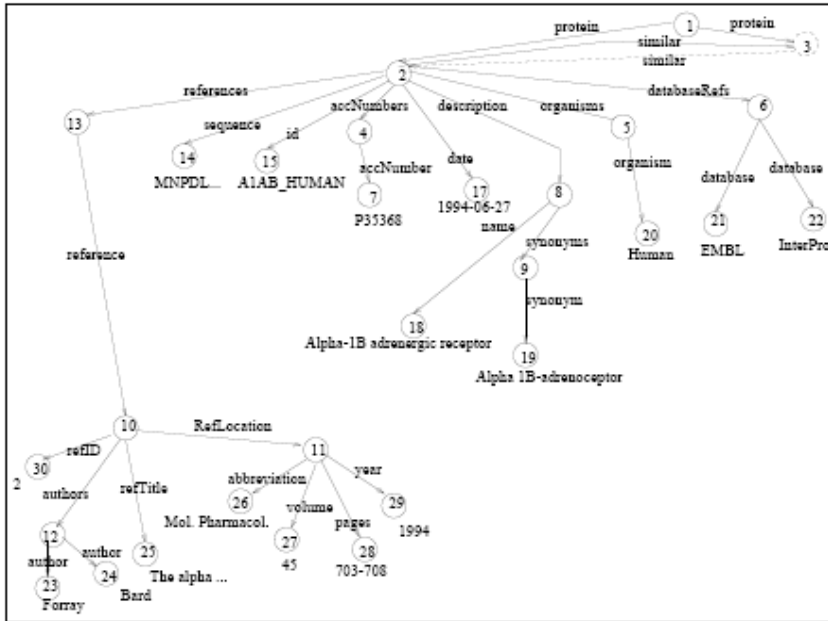


Figure 1: SwissProt db -1

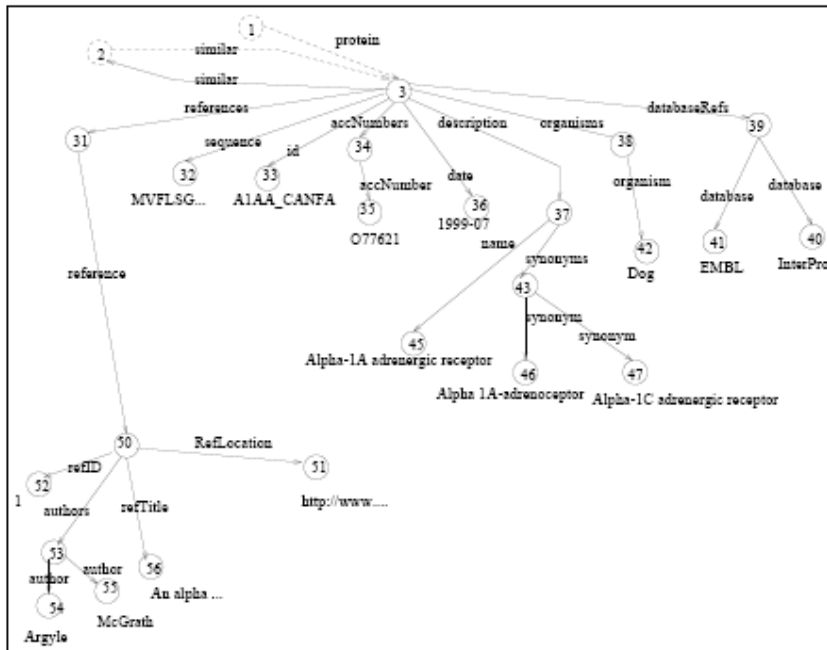


Figure 2: SwissProt db -2

```

<!-- *****
PIR-International Protein Sequence Database (PSD)
*****-->

<!-- ProteinEntry: the root element. -->
<!ELEMENT ProteinEntry (header,protein,organism,reference*,sequence)>

<!-- header: database information. -->
<!ELEMENT header (uid,accession*,created_date,seq-update_date)>

<!ELEMENT accession(#PCDATA)> <!-- accession number -->
<!ELEMENT created_date (#PCDATA)> <!-- date (DD-MMM-YYYY) -->
<!ELEMENT seq-update_date (#PCDATA)> <!-- date (DD-MMM-YYYY) -->

<!-- protein: the protein-names. -->
<!ELEMENT protein(name,alt-name*,contains*)>

<!ELEMENT name (#PCDATA)> <!-- protein name -->
<!ELEMENT alt-name (#PCDATA)> <!-- alternate protein name -->
<!ELEMENT contains (#PCDATA)> <!-- activity name -->

<!-- organism: identification of the biological source. -->
<!ELEMENT organism (source,common-name?,note*)>

<!ELEMENT source (#PCDATA)> <!-- source name -->
<!ELEMENT common-name(#PCDATA)> <!-- common name -->

<!-- reference -->
<!ELEMENT reference(refinfo,note*)>

<!-- refinfo: identification of the literature source. -->
<!ELEMENT refinfo(authors,citation,title?,xrefs?)>

<!ELEMENT authors(author+)> <!-- list of authors -->
<!ELEMENT author (#PCDATA)> <!-- author name -->
<!ELEMENT citation (#PCDATA)> <!-- citation name -->
<!ELEMENT title(#PCDATA)> <!-- title text -->

<!ELEMENT xrefs(xref+)> <!-- cross-references -->
<!ELEMENT xref (db,uid)> <!-- a cross-reference -->
<!ELEMENT db (#PCDATA)> <!-- database tag -->

<!-- sequence: the amino acid sequence. -->
<!ELEMENT sequence (#PCDATA)> <!-- amino acid symbols and
punctuation -->

<!-- General elements. Elements that can be contained in several
other elements. -->
<!ELEMENT note (#PCDATA)> <!-- note text -->
<!ELEMENT uid(#PCDATA)> <!-- entry identifier -->

```

Figure 3: DTD for the biological data source PIR-PSD

SwissProt concept	PIR-PSD concept
id	uid
accNumbers	accession
date	created_date
description	protein
synonym	alt_name
refTitle	title
refLocation	citation
databaseRefs	xrefs
database	db

Figure 4: Matching concepts