Introduction to Databases (winter term 2005/2006)

Assignment 10

hand in on January 23, 2006 during the tutorial

On our homepage we have published the intermediate results of the tutorials (up to assignment 8). Please check whether our data is correct and contact us if you encounter an error.

Task 10.1 (Some simple questions on XML) (2 Points)

a) For each XML document one can calculate the corresponding XML Information Set. How does this correspondence help in comparing two XML documents?

b) Consider a relational database (containing multiple tables with multiple rows). Sketch, how this data could be represented using XML. Consider both mapping possibilities (data inside nested elements or in attributes).

Task 10.2 (Relational, Object-Oriented and XML Data Model) (5 Points)
Consider the following ER diagram describing companies, their departments and their employees. Please note, that some attributes may remain null (e.g. a company might not have a web page, it might not have a stock price, etc.).

a) Map the above ER diagram to a relational database model in formal notation.

b) Map the diagram to an object oriented database model specified in ODL.

c) So far, we only consider schemaless XML documents. Give an example XML document that stores the information denoted in the ER diagram for two companies. Remember that not all presented attributes carry information for each entity!

Task 10.3 (RDF) (4 Points)
Model the following informally described knowledge by formulating appropriate RDF statements:

- We consider five music songs. The first three songs (We are the champions, Bohemian Rhapsody, Flash) are interpreted by Queen, the fourth song (Englishman in New York) by Sting and the last (Yesterday) by The Beatles.

- There are two ‘mini-samplers’. The first (with the name ‘Greatest Songs I’) contains the first, third and the last song (in that order). The second sampler (‘All Favourites’) contains the first, the second and the fourth song (in that order).

- The Sting song sounds similar to Yesterday. This is at least Joe Smith’s opinion.

- Joe Smith is 30 years old and lives in Aachen.

- His favourite songs are the first, the third and the last song.

- Freddy Mercury is the lead singer of Queen, Brian May is the guitarist.
To uniquely identify the resources in your statements, you can invent URIs of a common namespace (e.g. you could bind the namespace prefix i5 to the address: http://www-i5.informatik.rwth-aachen.de/EBD05/MusicRDF).

a) Construct an RDF graph visualisation of your model.

b) Give an XML serialization of this graph.

Task 10.4 (XPath) (2 Points)
Consider a single XML document that stores information about books in a library. An example of such a file that shows the structure we assume is as follows:

```xml
<?xml version="1.0" encoding="UTF-8"?>
<bib>
  <book isbn="3-8274-1488-1" price="100">
    <name>Databases</name>
    <author>George Smith</author>
    <author>Joe Black</author>
    <chapter>
      <title>Introduction</title>
      <NoPages>16</NoPages>
    </chapter>
    <chapter>
      <title>Basics of Query Languages</title>
      <NoPages>50</NoPages>
    </chapter>
    <chapter>
      <title>The SQL Standard</title>
      <NoPages>40</NoPages>
    </chapter>
    <editor>Mike Blue</editor>
  </book>
  <book isbn="3-8274-3213-1" price="80">
    <name>Watching TV</name>
    <author>Freddy TV Junkie</author>
    <chapter>
      <title>Introduction</title>
      <NoPages>30</NoPages>
    </chapter>
    <chapter>
      <title>Switching On TV</title>
      <NoPages>50</NoPages>
    </chapter>
    <chapter>
      <title>Switching Off TV</title>
      <NoPages>2</NoPages>
    </chapter>
  </book>
</bib>
```

Formulate suitable (absolute) XPath expressions that return a sequence of nodes of the XML document satisfying the following information demands:

- All books that cost below 50$.
- All authors who wrote a book with more than 5 chapters
- All authors and editors of books with a chapter that is longer than 50 pages.
- All authors who wrote a book that has chapter with the title ‘Evaluation’ between the chapters ‘Introduction’ and ‘Conclusion’.