

## Middleware for Heterogeneous and Distributed Information Systems – Exercise Sheet 5

Wednesday, November 26, 2008 – 10:00 to 11:30 – Room 48-379

### XML well-formedness

```
1      <?xml version="1.0" encoding="UTF-8"?>
2      <account number=123456789>
3          <balance><dollar>100,000.00</dollar>
4          <!-- Gold Customer -->
5          <depositor>
6              <firstname>John</firstname>
7              <lastname>Doe</lastname>
8          </depositor>
9      </account>
10     <account number="234567890" number="2">
11         <balance><dollar>-200.00</balance></dollar>
12         Account id overdrawn!
13         <depositor>
14             <firstname>Jane</firstName>
15             <lastname>Doe</lastName>
16         </depositor>
17     </account>
```

Figure 1: Sample (incorrect) XML document

An XML document is said to be well-formed if it conforms to the XML syntax rules. The XML document shown in Figure 1 is not well-formed. In what ways are the XML syntax rules violated?

### Document Type Definition (DTD)

A Document Type Definition is used to define the structure of XML documents with a list of legal elements and attributes. Say, you want to specify an XML-based exchange format for lecture catalogs using a DTD. How could the DTD look like considering the following requirements?

A lecture catalog includes none, one, or many lectures. Lectures are given either by a professor or an assistant professor. In either case the name of the lecturer is mentioned. Furthermore a textual description of the lecture is provided. Lectures are given at least once per week. For each teaching event the starting time, the end time, the weekday, and the lecture room is specified. Information about the lecture room consists of the name of the building where the room is located and the room's number.

## XML Schema (XSD)

XML Schema (also referred to as XML Schema Definition) is an alternative to DTD. Specify the schema of the lecture catalog described above by means of an XSD. What advantages of XSD over DTD do you see? Do you see any drawbacks?

## XPath

```
1    <?xml version="1.0" encoding="UTF-8"?>
2    <bibliography>
3      <article>
4        <title>SQL/MED - A Status Report</title>
5        <author>Jim Melton</author>
6        <author>Jan-Eike Michels</author>
7        <author>Vanja Josifovski</author>
8        <author>Krishna G. Kulkarni</author>
9        <author>Peter M. Schwarz</author>
10       <journal>SIGMOD Record</journal>
11       <year>2002</year>
12     </article>
13     <article>
14       <title>Advancements in SQL/XML</title>
15       <author>Andrew Eisenberg</author>
16       <author>Jim Melton</author>
17       <journal>SIGMOD Record</journal>
18       <year>2004</year>
19     </article>
20     <article>
21       <title>XQuery: An XML Query Language</title>
22       <author>Don Chamberlin</author>
23       <journal>IBM Systems Journal</journal>
24       <year>2002</year>
25     </article>
26   </bibliography>
```

**Figure 2: Sample XML document**

Figure 2 shows a sample XML document. Give XPath expressions to retrieve the following information from the document.

1. The title of all papers
2. The names of the authors of the paper “SQL/MED – A Status Report”
3. The title of second paper
4. Articles that were published before 2003
5. Articles with more than three authors

What is returned by the following XPath expressions?

1. `/bibliography/article[position()=1]/year/preceding-sibling::* /text()`
2. `//*`
3. `//author[text()='Don Chamberlin']/parent::journal`
4. `/bibliography/article[position()=2]/following::journal/text()`

## XQuery

XQuery is a general purpose query language for XML data. Consider the following sample queries. What result is produced when these queries are evaluated on the XML document shown in Figure 2?

1. 

```
FOR $a in /bibliography/article
LET $b := $a/year
WHERE $b > 2000
ORDER BY $b
RETURN $a
```
2. 

```
FOR $a IN fn:distinct-values(//author)
RETURN
<author>
  <name>{$a}</name>
  {
    FOR $b IN //article[author = $a]
    RETURN $b/title
  }
</author>
```