



## Data Models and Query Languages Summer 12

Due/Discussion by 21.06.2012

### 5. Sheet: XML, DTDs and Schema

#### Exercise 1 (Namespaces, Well-Formedness - 20 points)

```
<?xml version="1.0"?>
<?DOCTYPE "eth" SYSTEM "eth.dtd"?>
<eth xmlns="http://www.ethz.ch"
    xmlns:xmldb="http://www.dbis.ethz.ch"
    date="11.11.2006"
    xmldb:date="12.11.2006">
  <date text="An <gregorian> date">13.11.2006</date>
  <president number=1>Empty</president>
  <Rektor>Name 2</rektor>
</Doc>
```

- Correct the mistakes in this document to make it well-formed
- What are the namespaces of each attribute and element?
- Try to define the external DTD `eth.dtd` so that the document is valid. Does DTD actually handle namespaces? You can test your results using OxygenXML, the Eclipse XML tools or `xmllint`<sup>1</sup>. To check if an XML file `x.xml` is well-formed, use “`xmllint -format x.xml`”. With “`xmllint -dtdvalid d.dtd x.xml`” you can validate the file against the DTD `d.dtd`.

#### Exercise 2 (DTDs - 20 points)

Given is the following DTD  $D_1$  (from the XQuery Use Cases)

```
<!ELEMENT bib      (book*)>
<!ELEMENT book     (title, (author+ | editor+ ), publisher, price)>
<!ATTLIST book     year CDATA  #REQUIRED >
<!ELEMENT author   (last, first)>
<!ELEMENT editor   (last, first, affiliation)>
<!ELEMENT title    (#PCDATA)>      <!ELEMENT last      (#PCDATA)>
<!ELEMENT first    (#PCDATA)>      <!ELEMENT affiliation (#PCDATA)>
<!ELEMENT publisher (#PCDATA)>      <!ELEMENT price     (#PCDATA)>
```

- Provide each time the smallest XML document which is valid according to this DTD, and contains at least nodes (1) `bib`, (2) `book`, and (3) `bib`, `author` as well as `editor`. Size is measured in terms of characters, excluding whitespace.
- What changes to this DTD are necessary to ensure that `bib` is always being used as root element?<sup>2</sup>

<sup>1</sup><http://xmlsoft.org/>

<sup>2</sup>Caution: `xmllint` might not support this!

- c) In a web application the rule for the `book`-Node is changed as follows:

```
<!ELEMENT book ( (title, author+, publisher, price)
                  | (title, editor+, price, publisher) )>
```

What kinds of problems does this change possibly cause? And how could they be fixed?

### Exercise 3 (XML Schema: Constraints - 10 points)

Model tree structures (tree, node) with unique node ids:

- Provide a DTD and an XML Schema definition for a tree structure so that each has a document-wide unique ID.
- Provide an XML Schema definition for a tree structure, so that all children (i.e., direct descendants) of a node have a unique ID.

### Exercise 4 (XML-Schema: Lottery Numbers - 20 points)

- Provide an element definition for a List of positive numbers. Example:  
`<zahlen>11 45 27 17 21 55 100 356</zahlen>`
- Provide an XML Schema document *lotterie.xsd* so that the following XML document is valid. In particular, the element *zahlen* may only contain lists of 6 Integer numbers from to range of 1 to 49.

```
<?xml version="1.0"?>
<lottozahlen xmlns="http://www.lotterie.de.org"
              xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
              xsi:schemaLocation="http://www.lotterie.de/lotterie.xsd">
  <ziehung>
    <woche>July 1</woche>
    <zahlen>21 3 47 1 30 12</zahlen>
  </ziehung>
  <ziehung>
    <woche>July 8</woche>
    <zahlen>35 31 4 7 8 22</zahlen>
  </ziehung>
  <ziehung>
    <woche>July 15</woche>
    <zahlen>1 40 19 35 44 11</zahlen>
  </ziehung>
</lottozahlen>
```