Universität Freiburg
Institut für Informatik
Prof. Dr. Peter Fischer
Lecuture XML and Databases
Winter Semester 2011/12

Exercise Sheet 1

XML, Well-Formed XML

Prerequisite:

Working with XML is greatly simplified using the right tools. As an XML editor, we can recommend the following platforms:

- Oxygen XML. A trial version is available via http://www.oxygenxml.com
- Eclipse XML Tools. Part of the WTP tools. WTP tools are included in the Java EE version of Eclipse, or can be installed via http://www.eclipse.org/webtools/

Exercise 1: One document, several XML representations

1.1. Perform the following tasks

a) Create a sample document in Microsoft Word containing this exercise with about the same formatting, and save it as **sample.doc**.

b) Copy the contents (as raw text) of **sample.doc** into the editor and specify its structure using XML tags. (Title, sections). Also mark specific styles (such as italic) inside the text. Save it as **doc1.xml**.

c) Open **sample.doc** in Word and save it as XML (Save as... XML Document (*.xml)) with the name **doc2.xml**.

d) Open **sample.doc** in OpenOffice and save it in the OpenOffice format. Change the extension of the file from .write to .zip. Extract from the zip file the **content.xml** file and rename it to **doc3.xml**.

1.2. Open **doc1.xml, doc2.xml and doc3.xml** in the editor. Check that they are well-formed (oXygen tells you). Which format is better? Microsoft Office, OpenOffice, yours?

1.3. Is this data structured, unstructured or semi-structured?

Exercise 2: Well-Formed XML

2.1. Correct the following XML document to be well-formed (hint: use the XML editor):

```
<?xml version="1.0" encoding="utf-16"?>
<movies>
  <movie id=”56225”>
    <title>Love Story</title>
    <title></title>
    <year>1980</year>
    <_director name='Coppola'></_director>
    <comment text=”Five start” text=”Average”/>
    <xml>Introduce XML content</xml>
    <newcomment text="An <important> text">Oscar</newcomment>
    <comment lang=de>&copy; 1980 Warner Bros.</comment>
    <!-- Famous movie of the --80s -->
  </Movie>
</movies>
```

2.2. Is this correct in HTML? How about in XHTML? Why?

```
<html>
  <head>
    <title>Untitled</title>
  </head>
  Dear jane <br>
  <p>You are invited at the weekly meeting
  <p>Yours sincerely, <br>
  John
</html>
```