

DTD, Infoset, XML Validation

Exercise 1: Well-Formed XML

1.1. Correct the following XML fragment to be well-formed:

```
<?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>
```

Exercise 2: Validation XML: DTD and XML Schema

Given the following DTD (an equivalent XML Schema can be used; DTD used just because of less verbosity):

```
<!ELEMENT movies (Movie+)>
<!ELEMENT Movie ( title, year, _director, (comment | newcomment)+)>
<!ATTLIST Movie id ID #REQUIRED>
<!ELEMENT title (#PCDATA)>
<!ELEMENT year (#PCDATA)>
<!ELEMENT _director (#PCDATA)>
<!ATTLIST _director name CDATA #IMPLIED>
<!ELEMENT comment (#PCDATA)>
<!ELEMENT newcomment (#PCDATA)>
<!ATTLIST comment lang CDATA #IMPLIED>
```

(Simple DTD Tutorial: <http://www.w3schools.com/dtd/default.asp>)

2.1. Correct the document from Exercise 1 to conform to this DTD.

2.2* Describe the InfoSet of the document. * (possibly not yet presented in the lecture)

2.3. What is the difference between CDATA and PCDATA ?

2.4 Create a XML Schema with additional restrictions over DTD 2.1

(e.g., enforce type of year, enforce unique constraints, unique values)

2.5: What could/should be the result of the following expressions:

(id of first Movie) "+" (year of Movie with id 56225)?

- a. 56225
- b. 562251980
- c. 58205

Can you enforce it using a DTD ? What about an XML Schema ? Who defines the meaning of "+" ?

2.6 Provide 4 reasons to use elements instead of attributes.

Exercise 3:

3.1. Correct the following XML fragment: (e.g., from the point of view of InfoSet?)

```
<?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>13.11.2006</date>
  <president number="1">Empty</president>
  <Rektor>Name 2</rektor>
</Doc>
```

3.2. Is this correct 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>
  Cristian
</Html>
```

3.3. What is **the value and type** of the following expressions from the document INFOSET ? What is the number of items corresponding to the tree part ?

1) eth

|
date

2) eth

|
xmldb:date (xmlns:xmldb="http://www.dbis.ethz.ch")