

Übung zur Lehrveranstaltung

## Intelligente Systeme im WWW

Dr. Pascal Hitzler, Dr. Sebastian Rudolph, Dr. Raphael Volz  
Sommersemester 2007

### Übung 1: XML (30.04.2007)

---

*Wichtige Mitteilungen zum Inhalt und Ablauf der Vorlesung werden mit Hilfe unserer Mailingliste [iswww@lists.uni-karlsruhe.de](mailto:iswww@lists.uni-karlsruhe.de) bekannt gegeben. Es wird empfohlen, sich dafür via <https://www.lists.uni-karlsruhe.de/sympa/info/iswww> anzumelden (siehe auch Vorlesungswebseite).*

**Aufgabe 1.1** Beantworten sie stichpunktartig die folgenden zwei Fragen zum allgemeinen Verständnis.

- (a) Warum stellt XML eine solch wichtige Entwicklung dar?
- (b) Worin besteht der Unterschied zwischen DTD und Schema?

**Aufgabe 1.2** Entscheiden Sie, ob die folgenden XML-Dokumente (bzw. Teile davon) *wohlgeformt* sind und schlagen Sie gegebenenfalls eine Korrektur vor.

- (a) Ein Auszug aus einer früheren Startseite des AIFB ([www.aifb.uni-karlsruhe.de](http://www.aifb.uni-karlsruhe.de)):

```
<table width="100%">
<tr>
<td width="33%" align="left">
<a href="mailto:webmaster@aifb.uni-karlsruhe.de">
  <font class="footnote">&copy; AIFB 2001-2005</font></a></td>
<td width="33%" align="center">
&nbsp;<td width="33%" align="right">
<a href="http://www.aifb.uni-karlsruhe.de/about.html">
  <font class="footnote">About our SEMantic portAL</font></a></td>
</tr>
</table>
```

- (b) Ein vollständiges HTML-Dokument:

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE html PUBLIC
"-//W3C/DTD XHTML 1.0 Transitional//EN"
```

```
"http://www.w3.org/TR/xhtml1/DTD/xhtml1/transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
<head><title>Kontrollieren Sie Ihr HTML!</title></head>
<body>
  <h3>HTML-Check des W3C</h3>
  <b>Kostenloser HTML-Check:<br/><i>http://validator.w3.org/</b></i>
</body></html>
```

(c) Dieses Dokument hält sich nicht an die Empfehlung, alle Tags klein zu schreiben:

```
<Liste laenge=4>
  <Listeneintrag pos=1>
    <Element/>Lorem ipsum<element/>
  </Listeneintrag>
  <listeneintrag pos=3/>
</Liste>
```

**Aufgabe 1.3** Auf den Folien zur Vorlesung wurden bereits einige Kriterien für ein wohlgeformtes XML-Dokument genannt: jedes XML-Dokument enthält genau eine Wurzel und auf jedes öffnende Tag folgt ein entsprechendes schließendes Tag. Finden Sie, ausgehend von den obigen Beispielen, drei weitere Kriterien.

**Aufgabe 1.4** Geben Sie eine natürlichsprachliche Beschreibung für die folgende DTD an.

```
<!DOCTYPE kochbuch [
<!ELEMENT kochbuch (rezept)+ >
<!ELEMENT rezept (rezepttyp, titel, zutat+, arbeitsschritt+) >
<!ELEMENT titel (#PCDATA) >
<!ELEMENT zutat (#PCDATA) >
<!ELEMENT arbeitsschritt (#PCDATA) >
<!ATTLIST arbeitsschritt nummer CDATA #REQUIRED >
<!ELEMENT rezepttyp EMPTY >
<!ATTLIST rezepttyp name CDATA #REQUIRED >
]>
```

**Aufgabe 1.5** Ist der folgende Ausschnitt einer XML Datei zur DTD aus der vorigen Aufgabe konform? Begründen Sie Ihre Entscheidung.

```
<kochbuch>
  <rezept>
    <rezepttyp name="Torten"/>
    <titel>Schwarzwaldler Kirschtorte</titel>
    <arbeitsschritt nummer="19">Schokostreusel draufmachen</arbeitsschritt>
```

```
<zutat>Sahne</zutat>
<zutat>Schokostreusel</zutat>
</rezept>
</kochbuch>
```

**Aufgabe 1.6** Geben Sie die korrekte Übersetzung der folgenden Elemente von XML Schema nach DTD an.

- (a) `<element name="foo" minOccurs="0" maxOccurs="unbounded">`
- (b) `<element name="bar" minOccurs="0">`
- (c) `<element name="baz" maxOccurs="unbounded">`

**Aufgabe 1.7** Betrachten Sie die folgende XML-Datei:

```
<bib>
  <book year="1994">
    <title>TCP/IP Illustrated</title>
    <author><last>Stevens</last><first>W.</first></author>
    <publisher>Addison-Wesley</publisher>
    <price> 65.95</price>
  </book>
  <book year="1992">
    <title>Advanced Programming in the Unix environment</title>
    <author><last>Stevens</last><first>W.</first></author>
    <publisher>Addison-Wesley</publisher>
    <price>65.95</price>
  </book>
  <book year="2000">
    <title>Data on the Web</title>
    <author><last>Abiteboul</last><first>Serge</first></author>
    <author><last>Buneman</last><first>Peter</first></author>
    <author><last>Suciu</last><first>Dan</first></author>
    <publisher>Morgan Kaufmann Publishers</publisher>
    <price> 39.95</price>
  </book>
  <book year="1999">
    <title>The Economics of Technology and Content for Digital TV</title>
    <editor>
      <last>Gerbarg</last><first>Darcy</first>
      <affiliation>CITI</affiliation>
    </editor>
    <publisher>Kluwer Academic Publishers</publisher>
    <price>129.95</price>
```

```
</book>  
</bib>
```

- (a) Geben Sie eine DTD an, für die diese XML-Datei gültig ist.
- (b) Geben Sie ein XML Schema an, für die diese XML-Datei gültig ist.

```
<!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 )>
```

- (b) Die Lösung zu (b) ist als gesonderte XML-Datei auf den Webseiten der Vorlesung verfügbar.