

Grundlagen Semantic Web
Seminar für Computerlinguistik, Universität Heidelberg

Sebastian Rudolph

Wintersemester 2008/09

<http://semantic-web-grundlagen.de>

Übung 1: XML, XML Schema, URIs

Aufgabe 1.1 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):

```
<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</i></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 1.2 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 1.3 Entscheiden Sie, welche der folgenden Zeichenketten gültige URIs beziehungsweise URI-Referenzen sind.

1. <http://www.ietf.org/rfc/rfc3986.txt>
2. mak@aifb.uni-karlsruhe.de
3. <file:///etc/passwd>
4. <tel:+358-555-1234567>
5. ftp://localhost/Meine_Bücher/Algorithmen_in_C++#Kapitel-4
6. <http://de.wikipedia.org/wiki/Σ>

Aufgabe 1.4 XML Schema bietet eine Vielzahl vordefinierter Datentypen. Genaue Informationen über diese Typen können leicht online in den offiziellen W3C-Dokumenten nachgelesen werden <http://www.w3.org/TR/xmlschema-2/>. Verwenden Sie dieses Dokument zur Beantwortung der folgenden Fragen, und machen Sie sich dazu zunächst grob mit der Struktur der Spezifikation vertraut. Verwenden Sie Textsuche und Inhaltsverzeichnis um gezielt Informationen zu finden – keinesfalls sollten Sie die gesamte Spezifikation durchlesen!

- (a) Informieren Sie sich über die Datentypen `xsd:decimal`, `xsd:dateTime`, und `xsd:anyURI`, so dass Sie jeweils grob in eigenen Worten den lexikalischen Bereich und Wertebereich beschreiben können. Geben Sie jeweils ein (nicht schon in der Spezifikation verwendetes) Beispiel für einen syntaktisch korrekten Wert dieser Typen an.
- (b) XML Schema definiert eine Reihe von numerischen Datentypen. Informieren Sie sich über die Datentypen `xsd:int`, `xsd:decimal`, und `xsd:float`. Geben Sie für jeden dieser Datentypen einen möglichst charakteristischen Beispielwert an.
- (c) Worin besteht der Unterschied zwischen den Datentypen `xsd:int` und `xsd:integer`?

Aufgabe 1.5 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 1.6 Ist der folgende Ausschnitt einer XML Datei zur DTD aus der vorigen Aufgabe konform? Begründen Sie Ihre Entscheidung.

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

```

Aufgabe 1.7 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 1.8 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.

Aufgabe 1.9 Namensräume sind für die Serialisierung von XML-Dokumenten von großer Bedeutung, nicht nur als abkürzende Schreibweise.

- (a) Bestimmen Sie für das folgende XML-Dokumente, welche Elemente durch die Verwendung von Namensräumen jeweils ausgedrückt werden:

```

<?xml version="1.0" encoding="UTF-8"?>
<Eins xmlns ="Namensraum"
  xmlns:a="http://example.org/"
  xmlns:b="B1#">
  <a:Zwei xmlns:b="B2#">
    <b:Drei>
      <a:δ xmlns:a="http://example.org/2/">
    </b:Drei>
  </a:Zwei>
</Eins>

```

(b) Könnte man die in Aufgabe (a) dargestellte Baumstruktur auch ohne die Verwendung von Namensräumen darstellen?

Hinweis: Ein moderner Browser genügt, um XML-Dateien auf ihre Wohlgeformtheit hin zu untersuchen (indem man sie einfach öffnet). Sie können die Antwort auf diese Frage also auch „experimentell“ ermitteln.