Prof. Dr. Stefan Brass
Institut für Informatik
MLU Halle-Wittenberg

20. Oktober 2016

XML und Datenbanken
— 1. Übungsblatt: XML Daten und DTDs —

Dieses Blatt enthält Aufgaben für die erste Übung. Teil a) bis e) sind Präsenzübungen, die nicht abgegeben werden müssen. Falls Sie allerdings die Übung verpassen, sollten Sie einen ausreichend großen Teil der Aufgaben bearbeiten.

Teil f) und g) sind Hausaufgaben. Schicken Sie Lösungen dem Dozenten per EMail: `brass@informatik.uni-halle.de` (mit “`xml16`’ in der Betreff-Zeile, bis 26.10.2016).

a) Wie würden Sie in einer mündlichen Prüfung auf folgende Fragen antworten?

- Beschreiben Sie die Baumstruktur zur Darstellung von XML-Dokumenten im Hauptspeicher (soweit bisher besprochen).
- Was ist die Beziehung zwischen Tags und Elementen?
- Welche Arten von Tags gibt es? Beschreiben Sie jeweils die Syntax.
- Welche Zeichen sind in der Angabe des Attributwertes verboten? Wie kann man sie dennoch hineinbekommen?
- Was sind Unterschiede zwischen XML und HTML?
- Wie sieht eine Element-Deklaration in XML aus? Welche Konstrukte können in einer Inhalts-Spezifikation verwendet werden? Was bedeutet die Einschränkung, dass das Inhaltsmodell deterministisch sein muss? Geben Sie ein Beispiel für ein Inhaltsmodell, das verboten ist.
- Wofür steht “PC” bei “`#PCDATA`”? Warum muss man da ein “`#`” schreiben, aber bei “`EMPTY`” nicht?
- Im Prinzip sind XML Elementdeklarationen und die Regeln einer kontextfreien Grammatik (z.B. in BNF) doch recht ähnlich strukturiert. Was ist der wesentliche Unterschied von XML mit DTDs und konktextfreien Grammatiken hinsichtlich der definierbaren Sprachen?
- Wie werden Attribute in XML DTDs deklariert?
- Wie sieht ein Kommentar in XML aus? Wie kann man sich erklären, dass die Syntax doch recht umständlich ist?
- Wie unterscheiden sich ISO Latin 1 (ISO 8859-1) und UTF-8 bei der Codierung deutscher Umlaute?
- Wie sehen Zeichenreferenzen aus? Spielt die für die XML-Datei verwendete Zeichencodierung hier eine Rolle?

b) Das folgende XML-Dokument enthält mindestens 8 Fehler. Finden Sie möglichst viele davon und korrigieren Sie das Dokument. [Diese Aufgabe hat Dr. Henning Thielemann entwickelt.]

```
<Film 6jahre=freigegeben>
    <titel>Herr Fuchs und Frau Elster</titel>
    <regie>Harald Serowski & Erich Hammer<regie/>
    < altersbeschränkung typ = "für sechsjährige" >
    <beschreibung>
    Wenn es eine Personifizierung von <quote>Neugier</quote> gibt,
    dann ist es <b><i>Frau Elster</b></i>.
    In diesem heiteren Streifen für die ganze Familie
    ist sie wieder voll in ihrem Element.
    ...
    </beschreibung>
</film>
```

Dieses Dokument steht auch unter folgender URI zur Verfügung:

[http://www.informatik.uni-halle.de/˜brass/xml16/examples/err_film.xml]

Sie können Ihre XML-Dokumente unter einer der folgenden Adressen prüfen lassen:

- [http://www.validome.org/xml/]
  (Sie müssen hier den Schalter “Nur auf Wohlgeformtheit prüfen” setzen. Dieser wird leider nur bei manchen Browsern angezeigt, z.B. nicht bei Firefox. Mit Chromium funktioniert es dagegen. Dies scheint ein Fehler der Webseite zu sein, die unterschiedliches HTML an verschiedene Browser ausliefert.)
- [http://www.xmlvalidation.com/]
  (Da es keine DTD gibt, zeigt dieser Dienst am Ende an, dass das Element “`Film`” nicht definiert ist. Hier gibt es keine Prüfung nur auf Wohlgeformtheit.)
- [http://www.w3schools.com/xml/xml_validator.asp]
- Für die Prüfung auf Wohlgeformtheit, die wir hier benötigen, können Sie das Dokument aber auch einfach mit der Endung “`.xml`” speichern (z.B. in Ihr Verzeichnis “`public_html`”) und dann in einem Browser wie dem Firefox öffnen.

c) Schauen Sie sich eine XML-Datei im Browser an, z.B.

[https://www.ibiblio.org/xml/examples/shakespeare/much_ado.xml]

d) Schreiben Sie eine XML-Datei für ein vereinfachtes Vorlesungsverzeichnis der Informatik. Es muss eine Liste von Lehrveranstaltungen abgespeichert werden können. Jede Lehrveranstaltung hat

- einen Titel (kann nicht weggelassen werden),
- optional eine Web-Adresse,
- einen Typ (Vorlesung, Seminar oder Praktikum),
- null oder mehr Dozenten, und ein oder mehr Termine.

Für Termine sind Wochentag, Start-Uhrzeit, End-Uhrzeit und Raum abzuspeichern. Für Dozenten ist nur der Name abzuspeichern.

Verfassen Sie eine XML-Datei mit mindestens zwei Lehrveranstaltungen, von denen mindestens eine zwei Termine hat. Prüfen Sie Ihre Datei mit einem XML Validator (zunächst nur die Wohlgeformtheit, eine DTD für die eigentliche Validierung haben Sie ja noch nicht).

e) Entwickeln Sie nun eine DTD zu Ihrer XML Datei. Betten Sie die DTD in die Datendatei ein, so wie in folgendem Beispiel (für eine andere Anwendung), und validieren Sie die Datei.

```
<?xml version="1.0"?>
<!-- Kommentar: Beispiel -->
<!DOCTYPE BOOKLIST [
    <!ELEMENT BOOKLIST (BOOK)*>
    <!ELEMENT BOOK (AUTHOR+, TITLE, PUBL?, NOTE?)>
    <!ATTLIST BOOK ISBN CDATA #REQUIRED
              PAGES CDATA #IMPLIED>
    ...
]>
<BOOKLIST>
    <BOOK ISBN="0-13-014714-1" PAGES="1074">
        <AUTHOR FIRST="Paul" LAST="Prescod"/>
        <AUTHOR FIRST="Charles" MI="F." LAST="Goldfarb"/>
        <TITLE>The XML Handbook - 2nd Edition</TITLE>
        <PUBL YEAR="1999" MO="11">Prentice Hall</PUBL>
        <NOTE>Contains CD.</NOTE>
    </BOOK>
...
</BOOKLIST>
```


Hausaufgabe

f) Eine Kirchengemeinde möchte ihre Veranstaltungen und die an der Organisation beteiligten Personen in einer XML-Datei erfassen. Entwerfen Sie dafür eine DTD, die folgende Anforderungen erfüllt:

- Zu jeder Person ist ein eindeutiges Kürzel (eine ID), sowie Name, Telefonnummer und EMail-Adresse zu erfassen. Kürzel und Name sind notwendig, Telefonnummer und EMail-Adresse können fehlen.
- Zu jeder Veranstaltung sind Art (z.B. Gottesdienst, Konzert, Familienfrühstück), Datum, Uhrzeit und eine optionale Zusatzinformation (Text, z.B. beim Konzert gespielte Stücke) zu speichern.
- Außerdem sind zu jeder Veranstaltung die Personen zu speichern, die an der Organisation dieser Veranstaltung beteiligt sind (Veranstaltungs-Team). Dies können beliebig viele Personen sein (auch null).

Zur Verdeutlichung hier eine relationale Repräsentation einiger Beispieldaten (die Veranstaltungs-Nummer ist in der XML-Darstellung eventuell nicht nötig):

PERSONEN			
ID	NAME	TEL	EMAIL
Weihe	Pfarrerin Regina Weihe	1234	weihe@gmx.de
tz	Tom Zierenberg	5678	

VERANSTALTUNGEN				
NR	DATUM	UHRZEIT	ART	INFO
1	24.12.2015	15:00	Gottesdienst	
2	24.12.2015	22:00	Konzert	Buxtehude u.a.

TEAM	
NR	ID
1	Weihe
1	tz
2	tz

Schreiben Sie eine syntaktisch korrekte XML-Datei mit den Daten aus den Tabellen (ggf. ohne die Veranstaltungsnummer, eventuell auch ohne die Personen-ID, falls die Beziehung klar ist). Die Datei braucht keine `DOCTYPE`-Deklaration zu enthalten. Sie können davon ausgehen, dass UTF-8 als Zeichencodierung verwendet wird.

g) Schreiben Sie syntaktisch korrekte Element- und Attributlisten-Deklarationen. Die Datentypen der Attribute sind nicht sehr kritisch, sollen aber natürlich die oben genannten Beispielwerte erlauben. Selbstverständlich sollte Ihre Datei aus f) bezüglich dieser DTD korrekt sein (validiert werden können).