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

02. Mai 2016

Grundlagen des WWW
— 4. Übungsblatt: SGML und XML —

Dieses Blatt enthält Aufgaben für die vierte Übung. Es handelt sich um Präsenzübungen, die nicht abgegeben werden müssen. Falls Sie allerdings die Übung verpassen, sollten Sie einen ausreichend großen Teil der Aufgaben bearbeiten und dem Dozenten per EMail schicken (mit “`www16`” in der Betreff-Zeile, bis spätestens 07.05.2016).

a) Das folgende XML-Dokument enthält mindestens 8 Fehler. Finden Sie möglichst viele davon und korrigieren Sie das Dokument.

```
<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/www16/film.xmlerr]

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.

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

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

c) 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,
- 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.

Nächste Woche werden Sie eine DTD für diese Anwendung entwickeln müssen. Da DTDs in der Vorlesung noch nicht besprochen wurden, sollen Sie diese Woche nur eine Beispiel-Daten-Datei entwickeln (man kann zur DTD-Entwicklung auch so vorgehen, dass man zunächst ein Beispiel für Daten-Dateien schreibt). Verfassen Sie also eine XML-Datei mit mindestens zwei Lehrveranstaltungen, von denen mindestens eine zwei Termine hat. Prüfen Sie Ihre Datei mit einem XML Validator (wieder nur die Wohlgeformtheit, eine DTD für die eigentliche Validierung haben Sie ja noch nicht).

d) Geben Sie beim HTML Validator des W3C

[https://validator.w3.org/]

die Webadressen der Universitäten Halle und Magdeburg ein. Können Sie einen Teil der angezeigten Fehler schon verstehen? Schauen Sie sich auch den Quellcode an, bei Firefox geht dies z.B. mit “Control-U” (oder über das Menu unter “Developer Tools”, “Page Source”).

Man sollte hierfür u.a. wissen, dass die “Unordered List” `ul` eine Liste von “List Items” “`li` enthalten muss (wie in der Vorlesung gezeigt). Nichts sonst ist erlaubt.

Außerdem ist zu bedenken, dass das Zeichen “`&`” immer eine Zeichenreferenz oder eine Entity-Referenz einleitet, auch im Inneren von Attributwerten. Eine Liste von in HTML definierten Entities finden Sie z.B. unter

[https://www.w3.org/TR/html4/sgml/entities.html]

e) Wie würden Sie in einer mündlichen Prüfung die Beziehung von XML zu SGML beschreiben?

- Nennen Sie einige wichtige Einschränkungen von XML gegenüber SGML.
- Warum waren diese Einschränkungen wünschenswert?
- warum hat man die Einschränkungen als nicht so schlimm empfunden?

f) Schauen Sie sich die XML 1.0 Spezifikation an:

[https://www.w3.org/TR/REC-xml/]

Alternativ können Sie auch "The Annotated XML Specification" von Tim Bray verwenden:

[http://www.xml.com/axml/testaxml.htm]

Finden Sie heraus, welche Zeichen als Startzeichen eines Bezeichners (z.B. als Namen von Elementen) erlaubt sind. (Tipp: Es steht in Abschnitt 2.3 "Common Syntactic Constructs".) Es reicht, wenn Sie sich auf ASCII-Zeichen konzentrieren. Es wäre aber auch spannend, die Frage für deutsche Umlaute zu klären. Die offizielle Quelle für die Codes der Umlaute im Unicode ist

[http://unicode.org/charts/PDF/U0080.pdf]

Es eignet sich aber jede ISO Latin 1 Tabelle (die Zahlwerte der Zeichen-Codes sind identisch, die Codierung mittels UTF-8 nicht).