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

27. Oktober 2016

Grundlagen des WWW
— 2. Übungsblatt: XML —

Bei Teil a) handelt es sich um eine Präsenzübung, die nicht abgegeben werden muss. Die anderen Teile sind Hausaufgaben. Bitte schicken Sie eine Lösung per EMail an den Dozenten (mit "`xml16`" in der Betreff-Zeile, bis spätestens 03.11.2016, 16:00). Schicken Sie bitte keine leeren EMails nur mit Anhang, diese landen im Spam-Ordner.

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

- Was ist der Unterschied zwischen wohlgeformten ("well-formed") und gültigen ("valid") XML?
- Was sind die wichtigsten Datentypen für Attribute in XML DTDs? Nennen Sie möglichst mindestens fünf.
- Wie kann man Texte/Daten in eine XML-Datei einbetten, die möglicherweise "`<`" und andere Spezialzeichen enthalten? Wie würden Sie vorgehen, wenn diese Zeichen sehr selten sind? Wie würden Sie bei längeren Abschnitten vorgehen, in denen die Zeichen relativ häufig sind (z.B. Java Programmcode)?
- Was ist eine CDATA Section?
- Was sind Entities (genauer: "general, parsed, internal entities")? Was wäre eine typische Anwendung? Wie sieht eine Deklaration aus? Wie ist die Syntax einer Entity-Referenz (Aufruf)?
- Nennen Sie die fünf vordefinierten Entities in XML.
- Was sind Parameter Entities? Wie unterscheiden sich Deklaration und Benutzung von normalen (allgemeinen) Entities? Warum macht es Sinn, die beiden Arten von Entities von einander zu unterscheiden? Worin liegt der Vorteil für den Anwender?
- Wie sieht eine Processing Instruction syntaktisch aus? Was ist das überhaupt? Warum gibt es die Ähnlichkeit mit der XML Deklaration?
- Welche Möglichkeit gibt es syntaktisch für eine DOCTYPE Deklaration?
- Wie kann man eine eins-zu-viele Beziehung aus einem ER-Diagramm in eine XML DTD übersetzen? Nennen Sie mindestens zwei Möglichkeiten. Diskutieren Sie ggf. Einschränkungen und Probleme.
- Wie kann man eine viele-zu-viele Beziehung übersetzen?
- Wie geht man allgemein bei einem beliebigen ER-Diagramm vor?

Hausaufgabe

b) 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).

c) Schauen Sie sich die XHTML 1.0 Strict DTD an:

[https://www.w3.org/TR/xhtml1/dtds.html#a_dtd_XHTML-1.0-Strict]

Welches Inhaltsmodell hat das Element "`table`"? Geben Sie eine erlaubte Folge von Kindelementen an (nicht das komplette XML/XHTML, sondern nur die Element-Namen in richtiger Reihenfolge). Welchen Datentyp hat das Attribut `lang` (suchen Sie ggf. nach `i18n`)?

d) Entwickeln Sie eine DTD und eine passende XML-Daten-Datei zur Verwaltung von Information über Kunden und Bestellungen eines kleinen Internetshops:

- Die Basisdaten eines Kunden sind Kunden-Nr, Anrede (Herr oder Frau), Titel, Vorname, Name, Strasse, PLZ, Stadt, EMail-Adresse und Telefonnummer. Titel und Telefonnummer sind optional, alles andere muss angegeben werden.
- Produkte haben eine Bestellnummer, eine Bezeichnung und einen aktuellen Preis.
- Ggf. sollen zu Produkten auch alte Preise gespeichert werden, mit der Information, bis zu welchem Datum der Preis aktuell war. Damit die gesamte Preis-Historie erfasst werden kann, kann ein Produkt auch mehrere alte Preise haben.
- Zu einer Bestellung eines Kunden muss das Eingangsdatum erfasst werden. Natürlich gehört eine Bestellung zu genau einem Kunden.

- Eine Bestellung hat mehrere Positionen (mindestens eine). Eine Bestellposition ist für genau ein Produkt. Optional muss es möglich sein, eine Stückzahl zu erfassen. Wenn keine Stückzahl angegeben ist, soll die Anwendung die Stückzahl "1" geliefert bekommen.

Ihre XML-Datei soll mindestens zwei Kunden, zwei Produkte, zwei alte Preise, drei Bestellungen und vier Bestellpositionen enthalten.

Bitte lassen Sie Ihre Datei bezüglich Ihrer DTD validieren (wie auf dem letzten Blatt angegeben). Es sei auch auf die AltovaXML Community Edition verwiesen:

[http://www.softpedia.com/get/Internet/Other-Internet-Related/AltovaXML.shtml]

Das Programm hat nur eine einfache Kommandozeilen-Schnittstelle (für eine GUI und Zusatzfeatures muss man den Altova XMLSpy kaufen). Aber es enthält einen Validator, einen XSLT Prozessor, und eine XQuery-Implementierung.