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

22. November 2018

XML und Datenbanken
— 5. Übungsblatt: XML Schema —

Bei Teil a) bis e) handelt es sich um Präsenzübungen, die nicht abgegeben werden müssen. Sie sollten über die Wiederholungsfragen a) allerdings vor der nächsten Übung selbst nachdenken. Teil f) bis h) sind Hausaufgaben. Bitte schicken Sie eine Lösung per EMail an den Dozenten (mit "`xml18`" in der Betreff-Zeile, bis zum 28.11.2018). 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 zu XML Schema antworten?

- Was bedeutet der reguläre Ausdruck "`[ab]*|\d{3}.`"?
- Warum ist es problematisch, die `pattern`-Facette für numerische Typen zu verwenden?
- Nennen Sie einige numerische Typen, die in XML Schema vordefiniert sind.
- Wie definiert man einen "Simple Type" in XML Schema, der dem SQL-Typ `NUMERIC(5,2)` entspricht? Was ist der Basistyp und welche "Facets" müssen benutzt werden?
- Wie definiert man einen Aufzählungstyp in XML Schema?
- Was ist das Format für einen `dateTime` Wert in XML schema? Wie definieren Sie einen Wert für den "`22. November 2018, 9:00 MEZ`" (Mitteleuropäische Zeit, eine Stunde später als UTC)?
- Ist ein Listen-Typ ein einfacher Typ oder ein komplexer Typ in XML Schema? Angenommen, Sie haben eine Liste von ganzen Zahlen als Typ definiert. Wie würde eine lexikalische Repräsentation aussehen?
- Geben Sie ein Beispiel für eine Anwendung eines `union`-Typs.
- Was bedeutet `xs:all` in einem Inhaltsmodell in XML Schema? Nennen Sie einige EInschränkungen für die Anwendung von `xs:all` im Vergleich zu `xs:sequence` und `xs:choice` (oder erklären Sie umgekehrt positiv, was man mit `xs:all` machen kann).
- Was ist eine benannte Modelgruppe ("named model group", `xs:group`)? In welchen Situationen ist sie nützlich? Was ist der Vorteil von benannten Modelgruppen im Vergleich zu Parameter Eintities in DTDs?
- Warum sind globale Attribut-Deklarationen oft problematisch? Was kann man tun, wenn mehrere Attribute mit gleichem Namen und Typ in verschiedenen Elementen benötigt werden?

Präsenzaufgaben

b) Definieren Sie einen Datentyp "`length`" für Längenangaben (mit Einheit). Es soll eine Zeichenkette sein, die zuerst eine Folge von Ziffern enthält, und dann genau einen der Suffixe "`mm`", `cm`", "`m`". Führende Nullen sollen nicht erlaubt sein, aber die Länge `0` (mit einer der Einheiten) muss möglich sein. Schreiben Sie ein minimales Schema und XML Dokument, um Ihre Definition zu testen. Sie können dann z.B. folgenden Validator verwenden:

[http://www.freeformatter.com/xml-validator-xsd.html]

c) Definieren Sie einen Datentyp für Klausurergebnisse, der 0 bis 100 Prozentpunkte erlaubt, und alternativ die Angabe "`NE`" (nicht erschienen). Testen Sie Ihre Definition wieder.

d) Angenommen, Sie wollen die Farben von Feuerwerksartikeln erfassen. Artikel können natürlich auch eine Kombination mehrerer Farben haben. Zur Vereinfachung seien nur folgende Farben betrachtet (es gibt noch viel mehr): `Gold`, `Silver`, `Red`, `Green`, `Yellow`, `Blue`. Definieren Sie einen Aufzählungstyp für diese Farben und anschließend einen Listentyp für das Farbattribut von Feuerwerksartikeln.

e) In Ihrer Firma (für Großfeuerwerkskörper) haben die Produkt-IDs folgenden Aufbau. Definieren Sie einen passenden Typ.

- Erst ein einzelner Großbuchstabe,
- dann eine zweistellige Zahl,
- dann optional ein Punkt "." und eine 1–3 stellige Zahl (für Kaliber-Varianten, Punkt und Zahl müssen immer zusammen vorkommen oder zusammen fehlen), und
- schliesslich noch optional einer der Buchstaben(-Kombinationen): `G` (gold), `S` (silver), `R` (red), `GR` (green), `Y` (yellow), `B` (blue). Nur eine der Farben kann angegeben werden.

Ein möglicher Wert wäre also `S10.75GR`. Dies könne z.B. eine Feuerwerksbombe ("Shell") Typ 10 (z.B. eine Chrysantheme) in Kaliber 75mm mit grünen Spitzen sein. Ein anderer möglicher Wert ist `V01G` (z.B. ein Vulkan in Gold).

Hausaufgabe

f) Ich möchte Benchmark-Ergebnisse wie auf der folgenden Webseite in einer XML-Datei speichern:

[http://users.informatik.uni-halle.de/˜brass/push/bench.html]

Die (etwas vereinfachten) Anforderungen sind wie folgt:

- Es gibt eine Liste von Systemen (Programmen), die verglichen werden sollen. Jedes hat einen kurzen Namen/ID (z.B. "`XSB`"), optional einen langen Namen (z.B. "`XSB 3.6 (Gazpatcho)`"), und ebenfalls optional eine Webadresse (z.B. "[`http://xsb.sourceforge.net/`]"). Der kurze Name ist ein gültiger XML-Bezeichner ohne "`:`".
- Es gibt eine Liste von Benchmarks, jeder mit einem Bezeichner (z.B. "`Join1`"), einem langen Namen ("`Join-1 Benchmark with query a(X,Y)`"), und ein oder mehr Version von Programmcode/Anfragen (z.B. eine Datalog Version und eine SQL-Version).
- Jede Version eines Benchmarks hat einen Namen (z.B. "`SQL version`") und einen möglicherweise längeren Text (z.B. die SQL Anfrage). Natürlich muss klar sein, zu welchem Benchmark eine Version gehört. Wenn Sie wollen, können Sie einer Benchmark-Version auch eine ID zuordnen (zusätzlich zum Namen, der Leerzeichen enthalten kann).
- Zu jeder Messung der Laufzeit (Benchmark-Ergebnis) muss Folgendes gespeichert werden:
  - das System, auf dem der Benchmark ausgeführt wurde,
  - die Version des Benchmarks,
  - das Datum, an dem der Benchmark ausgeführt wurde, und
  - die Laufzeit in Sekunden (mit drei Nachkommastellen, also auf Millisekunden genau). Zur Vereinfachung soll nur die "Total Runtime" (Gesamtlaufzeit) von der obigen Webseite gespeichert werden, nicht die Ladezeit (für die Eingabedaten) und die eigentliche Laufzeit der Anfrageausführung ("execution time").

Entwickeln Sie eine XML Daten-Datei mit mindestens zwei Systemen, zwei Benchmarks (davon einen mit zwei Versionen) und mindestens drei Laufzeit-Messungen.

g) Entwickeln Sie eine DTD für diese Anwendung, so dass Ihre Daten-Datei aus f) bezüglich dieser DTD validiert werden kann.

h) Entwickeln Sie ein XML Schema für diese Anwendung, so dass Ihre Daten-Datei aus f) bezüglich dieses Schemas validiert werden kann.

Hinweis: Bitte validieren Sie Ihre Datei mit DTD und Schema. Abgaben, die Syntaxfehler enthalten, oder bei denen die Validierung nicht klappt, zählen eventuell als "nicht abgegeben" (also kein Punkt).