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

25. November 2016

XML und Datenbanken
— 6. Übungsblatt: XML Schema, XDM, DOM —

Teil a) und b) müssen nicht abgegeben werden. Sie sollten sich allerdings zu Hause mit diesen Fragen beschäftigen, um den Stoff der Vorlesung zu wiederholen.

Teil c) und d) sind Hausaufgaben. Bitte schicken Sie eine Lösung per EMail an den Dozenten (mit "`xml16`" in der Betreff-Zeile, möglichst bis 01.12.2016, 16:00). Schicken Sie bitte keine leeren EMails nur mit Anhang, diese landen im Spam-Ordner. Nennen Sie die Datei(en) bitte "Nachname_Vorname", ggf. mit weiterem Suffix.

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

- Welche Möglichkeiten gibt es, um
  - Von einem "Simple Type" einen "Simple Type" abzuleiten?
  - Von einem "Simple Type" einen "Complex Type" abzuleiten?
  - Von einem "Complex Type" einen "Complex Type" abzuleiten?

  Geben Sie jeweils ein Beispiel.
- Vergleichen Sie abgeleitete Typen in XML Schema mit Unterklassen in Java.
- Angenommen, Sie schreiben ein Anwendungsprogramm, das XML Daten verarbeitet. Sie haben dafür ein Schema angegeben. Wenn Sie erlauben, das andere Leute Ihr Schema mit `include` benutzen, und erweitern, womit müssen Sie dann bei der Programmierung rechnen? Die erweiterten Schemata können natürlich keine Typen, Elemente oder Attribute umdefinieren, die Sie schon definiert haben. Sie können aber neue hinzufügen, auch Untertypen.
- Was ist eine "Substitution Group" bei Element-Deklarationen?
- Wozu dient "`xsi:type`"?
- Macht es einen Unterschied, ob Sie in einer Element-Deklaration keinen Typ definieren, also das `complexType`-Element weglassen, oder ein leeres `complexType`-Element hinschreiben?

b) Wie würden Sie in einer mündlichen Prüfung auf folgende Fragen zum XPath/XQuery Data Model (XDM) antworten?

- Die wesentliche Datenstruktur, mit der XPath und XQuery arbeiten, ist eine Sequenz. Wie ist das definiert?
- Was sind die sieben Arten von Knoten? Welche Knoten können Kind-Knoten sein, welche Eltern-Knoten?
- Warum braucht man einen extra Dokument-Knoten und verwendet nicht einfach einen Wurzel-Element-Knoten?
- Nennen Sie einige syntaktische Details, die beim Einlesen/Parsen einer XML Datei in die interne XDM-Repräsentation verloren gehen.
- Was ist das Problem mit dem "pattern facet" für Zahl-Datentypen?
- Warum gibt es noch den XML InfoSet Standard, und nicht nur den XDM Standard?
- Welche Unterschiede bestehen zwischen einer XDM-Baumstruktur, die ohne Validierung aufgebaut wird, und einer, die aus einem "Post-Validation InfoSet" aufgebaut wird?
- Beschreiben Sie die Behandlung von Leerplatz zwischen Elementknoten.
- Warum sollen Namenspace-Knoten nicht mehr verwendet werden? Was ist das Problem? (Dies und die folgenden Fragen werden erst in der nächsten Vorlesung definiert.)
- Was ist die "Document Order"?
- Welche Beziehung besteht zwischen
  - "`anyType`",
  - "`anySimpleType`",
  - "`anyAtomicType`" und
  - "`untypedAtomic`"?

  Erklären Sie die Typ-Hierarchie und nennen Sie jeweils ein Beispiel für einen Typ, der noch zum jeweiligen Obertyp gehört, aber nicht zum Untertyp.
- Was ist der String-Value von Element-Knoten? Und der typed value, wenn das Dokument validiert wurde? Hier müssen Sie Elemente mit einfachem Inhalt, Elemente mit reinem Element-Inhalt und Elemente mit "mixed content model" unterscheiden.

Hausaufgabe

c) Definieren Sie ein XML Schema für eine kleine Datenbank von Personen im Institut:

- Alle Personen haben Anrede (Herr/Frau), ggf. Titel, Vorname, Nachname, und optional eine EMail-Adresse.
- Studenten haben zusätzlich ein Studienfach, ein Semester, und eine Matrikelnummer. Falls sie schon einen Abschluss haben (z.B. Bachelor), soll dieser auch gespeichert werden.
- Professoren haben Fachgebiete. Dies können mehrere sein, z.B. "Datenbanken" und "Logische Programmierung".
- Zu Doktoranden soll eine Information über ihre Stelle/Finanzierung festgehalten werden (z.B. "Mitarbeiter", "Graduiertenstipendium") und ein Professor als Betreuer. Wenn Sie wollen, können Sie den Professoren zu diesem Zweck noch einen Schlüsselwert zuweisen, z.B. ein Kürzel. Sie können aber auch annehmen, dass Vorname und Nachname zusammen innerhalb des Instituts für Professoren eindeutig sind.

Da denkbar ist, dass später noch andere Typen von Personen hinzukommen (z.B. Sekretärinnen und Techniker), gestalten Sie Ihr Schema bitte möglichst leicht erweiterbar. Es soll ein äußeres Element "`Institut`" geben, darin direkt für jede Person ein Element.

d) Informieren Sie sich über die Verarbeitung von XML in Java mittels DOM, z.B. mit folgendem Tutorial:

[https://docs.oracle.com/javase/tutorial/jaxp/dom/index.html]

Weitere Tutorials finden Sie z.B. hier:

[https://www.tutorialspoint.com/java_xml/java_dom_parser.htm]
[https://www.mkyong.com/java/how-to-read-xml-file-in-java-dom-parser/]

Die DOM-Spezifikation stammt vom W3C:

[https://www.w3.org/TR/DOM-Level-3-Core/]
[https://www.w3.org/DOM/]

Schreiben Sie oder beschaffen Sie sich ein kleines Beispielprogramm, übersetzen Sie dies, und führen Sie es aus. Es soll mindestens die im Dokument befindlichen Elemente anzeigen (den Element-Namen). Sie können aber auch mehr bzw. alle Knoten ausgeben. Wenn die Baumstruktur z.B. durch Einrückung sichtbar wird, wäre dies ein Plus. Reichen Sie Ihr Programm per EMail ein. Wenn Sie ein Programm aus dem Internet verwenden, müssen Sie die Quelle nennen. Sie müssen es in der Übung erklären können. Achten Sie also gerade bei fremden Programmen darauf, dass Sie diese voll verstanden haben.