

Information Retrieval - Sommer 2006
Dipl.-Inform. Sascha Kriewel, LF 137
kriewel@is.informatik.uni-duisburg.de

Übungsblatt 13 XML und XML-Anfragesprachen

Aufgabe 35: XML – DTD und XML Schema

- (a) Finde über die Suchfunktion der Uni-Bibliothek ein Buch in der UB, z.B. „Modern Information Retrieval“ von Baeza-Yates und Ribeiro-Neto. Lass Dir die Detailansicht zum Buch anzeigen, und erstelle eine passende DTD zu diesen Daten. Gebe die Metadaten zu Deinem Buch als DTD-konformes XML-Dokument an.

Der XML-Standard:

<http://www.w3.org/TR/2004/REC-xml11-20040204/>

- (b) Benutze statt einer DTD nun XML-Schema. Nutze dabei die zusätzlichen Möglichkeiten dieses Standards und erstelle ggf. sinnvolle Datentypen für die Elemente.

Der XML-Schema-Standard:

<http://www.w3c.org/TR/xmlschema-0/>

Aufgabe 36: XML-Anfragesprachen – XQuery

Die Informatik-Arbeitsgruppe für Datenbanken und Informationssysteme der Universität Göttingen stellt auf ihrer Webseite Länderdaten zu europäischen Ländern zur Verfügung, die aus einer Reihe von Datenquellen integriert wurden. Benutze die Daten unter

<http://dbis.informatik.uni-goettingen.de/Mondial/mondial-europe.xml>

mit der DTD

<http://dbis.informatik.uni-goettingen.de/Mondial/mondial.dtd>

Erstelle zu mindestens dreien der folgenden Aufgaben Anfragen in der XQuery-Syntax erstellt werden.

- (a) Gib eine Liste aller in Deutschland verlaufenden Flüsse aus.
- (b) Gib eine Liste der in Europa ausgeübten Religionen aus, und zu jeder Religion als Kindelemente die Länder (mit Prozentanteil), in denen diese verbreitet ist.
- (c) Gib eine Liste der Länder aus, die mehr Einwohner haben als Nordrhein-Westfalen (ermittle die Einwohnerzahl von NRW aus den XML-Daten statt eine Konstante zu benutzen).

- (d) Gib die Namen der europäischen Länder mit überdurchschnittlich grossem Bevölkerungswachstum aus.
- (e) Gib eine Liste der Organisationen aus, in denen Deutschland (Ländercode D) Mitglied ist. Jedes Organisationselement des Ergebnisses soll den Namen der Organisation enthalten und ein Attribut `abk` mit dem Akronym der Organisation.

Der XQuery-Standard:

<http://www.w3.org/TR/xquery/>

Aufgabe 37: XML-Anfragesprachen – XPath

Hier findest Du den kompletten Text von „Romeo and Juliet“ in XML-Form:

http://www.ibiblio.org/xml/examples/shakespeare/r_and_j.xml

Erstelle passende XPath-Ausdrücke, um mindestens drei der folgenden Anfragen zu beantworten. Benutze dabei, wenn nötig, die generalisierten Lokalisierungsoperatoren (Achsen).

- (a) Finde die letzte Szene des Stückes.
- (b) Welche Franziskaner treten im Stück auf? Es gibt Personengruppen unter den `Personae`. Um an den eigentlichen Textinhalt (`#PCDATA`) von Elementen zu kommen, benutzt man `ELEMENT/text()`.
- (c) Finde die Zeilen, die Romeo Juliet direkt entgegnet, d.h. Romeo antwortet Juliet.
- (d) In wievielen Szenen tritt Romeo auf, aber nicht Juliet? Zum Zählen gibt es die Funktion `count()`; z.B. evaluiert `count(//ACT)` zu 5.
- (e) Wer stirbt in dem Stück? Gehe vereinfacht davon aus, dass auf den letzten Auftritt des Sterbenden eine `STAGEDIR` folgt, die den Text „dies“ oder „Dies“ enthält. Die XPath-Funktion hierzu ist `contains(text(), "something")`.

Der XPath-Standard:

<http://www.w3c.org/TR/xpath>

Wer kontrollieren möchte, ob der XPath-Ausdruck auch das Richtige tut: Es gibt eine Reihe von Werkzeugen für verschiedene Plattformen, um XPath-Ausdrücke über XML-Dokumenten auszuwerten. Ich persönlich mag *xmllint* (<http://xmlsoft.org/xmllint.html>), das auch einen Shellmodus hat, in dem man interaktiv XPath-Ausdrücke angeben und innerhalb des Dokuments wie in einem Dateibrowser navigieren kann. Eine Windows-Portierung ist unter <http://www.zlatkovic.com/libxml.en.html> zu bekommen. Das Werkzeug ist in dem Paket `libxml2-2.6.24.win32.zip` enthalten.