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

17. Januar 2019

XML und Datenbanken — 11. Übungsblatt: XPath —

Bei Teil a) und b) handelt es sich um Präsenzübungen, die nicht abgegeben werden müssen. Sie sollten über die Wiederholungsfragen a) allerdings vor der nächsten Übung selbst nachdenken. Teil c) bis e) sind Hausaufgaben. Bitte schicken Sie eine Lösung per EMail an den Dozenten (mit "`xml18`" in der Betreff-Zeile, bis zum 23.01.2019). 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 XPath antworten?

- Nennen Sie einige Unterschiede zwischen XPath 1.0 und XPath 2.0. D.h. welche XPath 2.0 Konstrukte können nicht in einem XSLT Stylesheet verwendet werden, das auch in Browsern funktionieren soll, die nur XPath 1.0 unterstützen? (Das ist noch der Stand der Technik.)
- Was ist die Bedeutung der folgenden Bedingung in XPath, d.h. für welche Attributwerte ist sie wahr?

```
@weekday = ('Sat', 'Sun')
```

- Warum könnte der Ausdruck "`$x idiv 1`" nützlich sein? (Bedenken Sie, dass das Ergebnis von `idiv` immer eine ganze Zahl ist, während die Argumente von beliebigem numerischen Typ sein können.)
- Was ist der Unterschied zwischen den beiden folgenden XPath Ausdrücken?
  - `if @GUEST_STUDENT = true() then ... else ...`
  - `if @GUEST_STUDENT then ... else ...`
- Und was ist der Unterschied zwischen diesen XPath Ausdrücken?
  - `true`
  - `true()`
- Gibt es einen Unterschied zwischen diesen XPath Ausdrücken?
  - `not /A`
  - `not(/A)`
- Was ist die Bedeutung von "`not(*)`"?
- Wie können Sie prüfen, ob der Wert des Attributes `A` des Kontext-Knotens eine gültige lexikalische Repräsentation einer ganzen Zahl (`xs:integer`) ist (d.h. eine Folge von Ziffern, ggf. mit Vorzeichen)? Schreiben Sie einen XPath Ausdruck, der in diesem Fall den Zahlwert liefert, und sonst die Zahl `-1`.

- Was ist der Unterschied zwischen den beiden folgenden XPath-Bedingungen?
  - `@A castable as xs:positiveInteger`
  - `@A instance of xs:positiveInteger`
- Und was ist der Unterschied zwischen diesen XPath-Ausdrücken?
  - `@A cast as xs:positiveInteger`
  - `@A treat as xs:positiveInteger`
- Was sind die Vor- und Nachteile von statischer Typprüfung in XPath?
- Nennen Sie zehn wichtige XPath-Funktionen. Wie würden Sie ein Handbuch über XPath-Funktionen strukturieren, d.h. welche Kapitelüberschriften hätte es? Zum Vergleich: "XPath and XQuery Functions and Operators 3.1":

[https://www.w3.org/TR/xpath-functions-31/].

Präsenzaufgaben

b) Gegeben sei wieder die Punktedatenbank mit den Daten in Elementen:

```
<?xml version='1.0' encoding='ISO-8859-1'?>
<GRADES-DB>
    <STUDENTS>
        <STUDENT>
            <SID>101</SID>
            <FIRST>Ann</FIRST>
            <LAST>Smith</LAST>
        </STUDENT>
        ...
    <RESULTS>
        <RESULT>
            <SID>101</SID>
            <CAT>H</CAT>
            <ENO>1</ENO>
            <POINTS>10</POINTS>
        </RESULT>
        ...
</GRADES-DB>
```

[http://users.informatik.uni-halle.de/~brass/xml18/examples/ex2.xml]

- Schreiben Sie eine XPath-Anfrage, um die durchschnittliche Punktzahl für Hausaufgabe 1 zu berechnen.
- Was ist die Bedeutung der folgenden XPath-Anfrage?

```
for $p in max(//POINTS) return //RESULT[POINTS=$p]
```

- Liefert diese Anfrage immer das gleiche Ergebnis?

```
//RESULT[not(number(POINTS) < //RESULT/POINTS)]
```

  Diese Anfrage benutzt nur Konstrukte von XPath 1.0.

- Warum funktioniert die Anfrage nicht ohne "`number`", d.h.

```
//RESULT[not(POINTS < //RESULT/POINTS)]
```

  Die vorkommenden Datenwerte sind 5, 7, 8, 9, 10, 12. Sie stellen fest, dass die Anfrage nur `RESULT`-Elemente mit 9 Punkten liefert. Warum?

- Schreiben Sie eine XPath-Anfrage, um den Nachnamen aller Studenten mit maximaler Punktzahl für Hausaufgabe 1 zu drucken. Entwickeln Sie zwei Varianten: Einmal für das tatsächlich erreichte Maximum, und einmal für die im `MAXPT`-Element gespeicherte Maximalpunktzahl der Aufgabe. (Sie können es auch mit Hausaufgabe 2 testen, bei der kein Student `MAXPT` Punkte erreicht hat.)
- Drucken Sie die Nachnamen aller Studenten, die gar keine Hausaufgaben abgegeben haben.
- Drucken Sie die Nachnamen aller Studenten, die mindestens 9 Punkte sowohl für Hausaufgabe 1 als auch für Hausaufgabe 2 bekommen haben.
- Was ist die durchschnittlich erreichte Gesamtpunktzahl für Hausausaufgaben, d.h. was ist der Durchschnitt der Summe der Hausaufgabenpunkte über allen Studenten, die mindestens eine Hausaufgabe abgegeben haben.

Sie können die Anfragen z.B. mit folgender Webseite ausprobieren:

[http://www.freeformatter.com/xpath-tester.html]

Hausaufgabe

Verwenden Sie nochmals die XML-Datensammlung für klassische Musik-CDs:

- Daten:
  [http://www.informatik.uni-halle.de/˜brass/xml18/examples/cd.xml]
- XML Schema:
  [http://www.informatik.uni-halle.de/˜brass/xml18/examples/cd.xsd]
- DTD:
  [http://www.informatik.uni-halle.de/˜brass/xml18/examples/cd.dtd]

Das Dokument enthält die folgenden Elemente:

- `CDDB`: `(composers, cds?, soloists?)`

- `composers`: `(composer*)`
- `composer`: `(pieceOfMusic*)`, attributes: `cno`, `firstName`, `name`, `born`, `died`.
- `pieceOfMusic`: `(recording*)`, attributes: `pno`, `title`, `key`, `opus`.
- `recording`: empty content, attributes: `rno`, `orchestra`, `conductor`.
- `cds`: `(cd*)`.
- `cd`: `(track*)`, attributes: `cdno`, `name`, `producer`, `numDiscs`, `totalTime`.
- `track`: empty content, attribute: `rno`.
- `soloists`: `(soloist*)`.
- `soloist`: `(performance*)`, attribute: `name`.
- `performance`: empty content, attributes: `rno`, `instrument`.

c) Schreiben Sie die folgenden Anfragen in XPath und testen Sie Ihre Anfragen mit einer XPath-Implementierung.

   - Gesucht ist der früheste Komponist in dem Dokument, d.h. der Komponist mit dem minimalen Geburtsjahr (Attribute `born`). Selektieren Sie das ganze `composer` Element.
   - Drucken Sie die Nachnamen aller Komponisten, die eine Sinfonie geschrieben haben, d.h. der Titel des Musikstücks enthält eine der Teilzeichenketten "`Sinfon`" or "`Symphon`". Stellen Sie sicher, dass jeder Komponist nur ein Mal ausgegeben wird, selbst wenn er mehr als eine Sinfonie geschrieben hat. (Für verschiedene Komponisten mit gleichem Nachnamen wäre es gut, wenn der Nachname doppelt ausgegeben wird, Sie dürfen aber auch einfach alle Duplikate entfernen.) Wenn Sie wollen, können Sie den Vergleich der Zeichenketten so machen, dass Groß-/Kleinschreibung nicht berücksichtigt wird. Das ist aber nicht verlangt.
   - Welche CDs enthalten Musikstücke von mehr als einem Komponisten? Drucken Sie den Namen der CS.
   - Was ist die durchschnittliche Anzahl Tracks auf einer CD? Um auch Mehrfach-CDs korrekt zu behandeln, teilen Sie die Anzahl durch den Wert des Attributs `numDisks`. Beachten Sie, dass ohne Validierung der Defaultwert "`1`" nicht zur Verfügung steht, sondern das Attribut für normale CDs einfach fehlt.