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

10. Januar 2019

XML und Datenbanken
— 10. 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 16.01.2019). 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?

- XPath hat keine reservierten Worte, d.h. alle in XML erlaubten Bezeichner für Elementnamen können auch in XPath verwendet werden (ohne spezielle Kennzeichnung). Nennen Sie einige Beispiele, wie Schlüsselworte in XPath von normalen Elementnamen unterschieden werden können.
- Was bedeutet der XPath-Ausdruck “`x-1`”?
- Wie werden Variablen in XPath gekennzeichnet?
- Wenn man in XPath eine Zeichenketten-Konstante mit `'` als Begrenzungszeichen schreibt, wie kann man dieses Zeichen als Daten der Zeichenkette einfügen? Warum funktion die XML Lösung mit “`'`” nicht in XPath? (Betrachten Sie dabei den Fall, dass der XPath-Ausdruck ein Attributwert in XML ist, wie z.B. bei einem XSLT-Stylesheet.)
- Wie schreibt man einen Datumswert in XPath, z.B. den 24.12.2018?
- Wie schreibt man boolesche Werte in XPath?
- Wie kann man in XPath auf die drei wichtigsten Komponenten des Kontexts zugreifen (Kontext Element, Kontext Position und Kontext Größe)?
- Was macht der Komma-Operator “`,`” in XPath? Vergleichen Sie es mit dem Vereinigungs-Operator “`|`” (den man auch “`union`” schreiben kann).
- Was ist “Atomisierung”? Wo wird das angewendet? Mit welcher Funktion können Sie es explizit aufrufen?
- Was ist der Haupt-Unterschied zwischen allgemeinen Vergleichsoperatoren wie “`=`” und den Wert-Vergleichsoperatoren wie “`eq`”?
- Sie stellen mit XPath Anfragen an ein Dokument, das nicht validiert wurde. Welches Problem ergibt sich dann bei einer Anfrage wie “`//RESULT[@POINTS gt 8]`” Welchen Typ hat der Attributwert hier? Wie können Sie das Problem lösen?
- Geben Sie ein Beispiel, in dem gleichzeitig `$x = 1` und `$x != 1` gelten.

- Zeigen Sie an einem einfachen Beispiel, dass “`=`” die Transitivität verletzen kann.
- Wie kann man in XPath prüfen, ob zwei Knoten identisch sind?
- Wie kann man abfragen, ob ein Knoten in der Dokument-Anordnung vor einem anderen Knoten steht?
- Wie schreibt man den Allquantor und den Existenzquantor in XPath? Erklären Sie Syntax und Semantik. Warum ist es ein Problem, wenn die Bedingung einen Laufzeitfehler für einige der Werte erzeugt, über denen der Quantor läuft?
- Beschreiben Sie Syntax und Semantik von `for`-Ausdrücken in XPath.
- Normalerweise sind Schleifen typische Kennzeichen von imperativen Programmiersprachen. Vergleichen Sie `for`-Schleifen in Sprachen wie Java mit den `for`-Ausdrücken in XPath. Erklären Sie, warum XPath nicht imperativ ist. Wie könnten `for`-Ausdrücke von XPath auf parallelen Prozessoren ausgewertet werden? Warum sind typische Schleifen in Java viel schwieriger zu parallelisieren?
- Erklären Sie die Syntax und Semantik von `if`-Ausdrücken in Xpath. Warum gibt es keine Version ohne `else`? Was ist der Unterschied zu `if`-Statements in Java?

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 Anfrage in Xpath, die die `SID` von “Ann Smith” liefert. Es reicht, das `SID`-Element auszuwählen. Sie können aber auch `data(...)` verwenden, um die Atomisierung zu erzwingen.

- Drucken Sie die Nachnamen aller Studenten, die mehr als 8 Punkte für Hausaufgabe 1 bekommen haben.

  Diese Aufgabe verlangt einen (Semi-)Join. Sie können den `some`-Quantor verwenden, um eine Variable an einen der benötigten Knoten zu binden, und dann den impliziten Kontext für den anderen Knoten verwenden.

- Was ist der Fehler in

  ```
  //EXERCISE[some $r in //RESULT satisfies POINTS = MAXPT]
  ```

- Schreiben Sie eine Anfrage, die alle Hausaufgaben-Punkte der Studentin 101 liefert. Schachteln Sie dann diese Anfrage in die Xpath-Funktion `sum(...)`.

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]

Die DTD ist hier abgedruckt, da sie eine recht kompakte Übersicht der Struktur des Dokuments ist:

```
<!ELEMENT CDDB (composers, cds?, soloists?)>
<!ATTLIST CDDB  xmlns:xsi CDATA #IMPLIED
                xsi:noNamespaceSchemaLocation CDATA #IMPLIED>

<!ELEMENT composers    (composer*)>

<!ELEMENT composer     (pieceOfMusic*)>
<!ATTLIST composer     cno        NMTOKEN #REQUIRED
                       firstName  CDATA   #REQUIRED
                       name       CDATA   #REQUIRED
                       born       NMTOKEN #IMPLIED
                       died       NMTOKEN #IMPLIED>
```



```
<!ELEMENT pieceOfMusic (recording*)>
<!ATTLIST pieceOfMusic pno        NMTOKEN #REQUIRED
                       title      CDATA   #REQUIRED
                       key        CDATA   #IMPLIED
                       opus       CDATA   #IMPLIED>

<!ELEMENT recording    EMPTY>
<!ATTLIST recording    rno        NMTOKEN #REQUIRED
                       orchestra  CDATA   #IMPLIED
                       conductor  CDATA   #IMPLIED>

<!ELEMENT cds          (cd*)>

<!ELEMENT cd           (track*)>
<!ATTLIST cd           cdno       NMTOKEN #REQUIRED
                       name       CDATA   #REQUIRED
                       producer   CDATA   #REQUIRED
                       numDiscs   NMTOKEN "1"
                       totalTime  NMTOKEN #IMPLIED>

<!ELEMENT track        EMPTY>
<!ATTLIST track        rno        NMTOKEN #REQUIRED>

<!ELEMENT soloists     (soloist*)>

<!ELEMENT soloist      (performance*)>
<!ATTLIST soloist      name       CDATA   #REQUIRED>

<!ELEMENT performance  EMPTY>
<!ATTLIST performance  rno        NMTOKEN #REQUIRED
                       instrument CDATA   #IMPLIED>
```

Schreiben Sie die folgenden Anfragen in XPath und testen Sie Ihre Anfragen mit einer XPath-Implementierung. Beschreiben Sie in Ihrer Hausaufgabe kurz, mit welchem System Sie die Abfragen ausprobiert haben, und ob es dabei irgendwelche Schwierigkeiten gab.

c) Geben Sie die Namen aller Solisten aus, die Violine spielen. Da ein Solist eventuell mehrere Instrumente spielt (z.B. "Violine" und "Viola"), steht das Instrument im Element `solist_aufnahme`, das in `solist` geschachtelt ist.

d) Geben Sie alle CDs aus (Element-Knoten), die mehr als 3 Einzel-CDs enthalten (also mindestens den Wert 4 im Attribut `anz_cds` haben).

e) Welche CDs (Attribut `name`) enthalten Stücke von "Wolfgang Amadeus Mozart"?

f) Selektieren Sie alle Komponisten, bei denen keine Opus-Nummern verwendet werden, d.h. keines ihrer Stücke hat das Attribut "`opus`".