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

20. Dezember 2018

XML und Datenbanken
— 9. 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 09.01.2019). 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?

- Was ist der Zweck von XPath? Wo wird es verwendet? Nennen Sie andere XML Technologien, die XPath verwenden.
- Was berechnet XPath, d.h. was ist die Basisdatenstruktur bzw. der Ergebnistyp von XPath Ausdrücken?
- Nennen Sie einige wichtige Komponenten des Kontextes. Warum gibt es bei XPath 2.0 einen statischen und einen dynamischen Kontext? Nennen Sie zwei Vorteile von statischer Typprüfung.
- Wenn eine Sprache wie XPath (oder auch relationale Algebra)
  - eine "Kernsyntax" hat (bei der relationalen Algebra nur bis Basisoperationen), und
  - eine "volle Syntax" mit Abkürzungen, die in die "Kernsyntax" übersetzt werden können

  was muss man dann definieren, um Syntax und Semantik vollständig zu erklären?
- Wo werden Achsen in XPath verwendet? Nennen Sie einige Achsen von XPath. Welchen Unterschied macht es, ob eine Achse eine Vorwärtsachse ist oder eine Rückwärtsachse?
- Was für Knotentests gibt es? Warum reicht es, dass man nur entweder Name oder Typ angeben kann?
- Warum wird der Typ-Test auf Text-Knoten `text()` geschrieben, und nicht einfach `text`? D.h. warum sind die Klammern nötig? Was würde passieren, wenn man sie weglässt?
- Nennen Sie einige Beispiele für Sequenztypen in XPath 2.0.
- Wie sind Pfadausdrücke in XPath aufgebaut? Was sind die Komponenten eines einzelnen Schritts in der vollen Syntax? Was sind die wichtigsten Abkürzungen?
- Beschreiben Sie genau, wie `E1/E2` ausgewertet wird.

- Beschreiben Sie genau, wie `E1[E2]` ausgewertet wird.
- Was ist der "effektive boolesche Wert" einer Sequenz? Wo wird er gebraucht?
- Geben Sie ein Beispiel für einen XPath Ausdruck mit einer vermutlich unerwarteten Semantik (einen möglichen Fehler).

Präsenzaufgaben

b) Die folgenden XPath-Anfragen beziehen sich auf eine XML-Datei mit Daten über Angestellte ("employees"), Abteilungen ("departments"), und Gehaltsstufen ("salary grades"). Die XML Datei wurde aus einer relationalen Beispiel-Datenbank von Oracle abgeleitet. Die Daten-Datei und eine zugehörige DTD finden sich hier:

- [http://users.informatik.uni-halle.de/˜brass/xml18/examples/empdept.xml]
- [http://users.informatik.uni-halle.de/˜brass/xml18/examples/empdept.dtd]

Die DTD ist hier als Referenz abgedruckt:

```
<!ELEMENT EMPDEPT (DEPT*, SALGRADE*)>

<!ELEMENT DEPT (EMP*)>
<!ATTLIST DEPT  DEPTNO   NMTOKEN #REQUIRED
                DNAME    CDATA   #REQUIRED
                LOC      CDATA   #REQUIRED>

<!ELEMENT EMP EMPTY>
<!ATTLIST EMP   EMPNO    NMTOKEN #REQUIRED
                ENAME    CDATA   #REQUIRED
                JOB      CDATA   #REQUIRED
                MGR      NMTOKEN #IMPLIED
                HIREDATE CDATA   #REQUIRED
                SAL      NMTOKEN #REQUIRED
                COMM     NMTOKEN #IMPLIED>

<!ELEMENT SALGRADE EMPTY>
<!ATTLIST SALGRADE
                GRADE    NMTOKEN #REQUIRED
                LOSAL    NMTOKEN #REQUIRED
                HISAL    NMTOKEN #REQUIRED>
```

Schreiben Sie die folgenden Anfragen in XPath:

- Drucken Sie die Namen (`ENAME`) aller Angestellten in einer Abteilung in "`DALLAS`" (`LOC`).

- Geben Sie alle Daten des Angestellten "`FORD`" (`ENAME`) aus.
- Was ist der Beruf (`JOB`) des Angestellten "`SCOTT`"?
- Welche Abteilungen (`DNAME`) haben Angestellte mit dem Beruf (`JOB`) "`ANALYST`"?
- Wer hat den Präsidenten der Firma als direkten Vorgesetzten? Die Angestellten-Nummer `EMPNO` des direkten Vorgesetzten ist im Attribut `MGR` ("manager") gespeichert, "`PRESIDENT`" ist ein `JOB`.)
- Welche Abteilung hat keine Angestellten?

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

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

Hausaufgabe

Laden Sie die folgenden Dateien herunter, die Daten über eine CD-Sammlung für klassische Musik enthalten:

- 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 (und ggf., wie Sie diese gelöst haben). Eine Möglichkeit ist es, die BaseX XML Datenbank herunterzuladen:

[http://basex.org/home/]

c) Gesucht sind alle Kommentarknoten im ganzen Dokument.

d) Gesucht sind die Namen (Attributwert) von allen im Dokument vorkommenden Orchestern.

e) Selektieren Sie alle Stücke (Elementknoten) von "`Bach`" in "`E-dur`".