

XML und Datenbanken

— 11. Übungsblatt: XQuery II —

Allgemeine Aufgabe

Notieren Sie sich eventuelle Verständnisfragen, so dass wir diese im nächsten Online-Treffen klären können. Da es erfahrungsgemäß sehr sill ist: Nehmen Sie sich bitte die Zeit und denken Sie bewusst über mögliche Fragen nach (wenigstens eine). Ich kann auch Studierende drannehmen, die sich nicht gemeldet haben. Ein Mal würde ich wohl verstehen, dass der Stoff so einfach war, dass Sie einfach keine Fragen haben. Wenn sich das wiederholt, müsste ich dann aber umgekehrt prüfungsähnliche Fragen stellen.

Hausaufgabe

Geben Sie die folgenden Aufgaben bis Montag, 23.01.2023, 16⁰⁰, über die Übungsplattform in StudIP ab. Schreiben Sie die Lösungen in eine `.txt`- oder `.xq`-Datei. Es gibt 3 Punkte pro Anfrage.

Die Abgaben nur stichprobenartig kontrolliert. Wenn Ihre Abgabe nicht kontrolliert wurde, bekommen Sie die volle Punktzahl. Wenn Sie später wegen Plagiaten auffallen, oder bei einer Stichprobe eine fast gar nicht gelöste Aufgabe entdeckt wird, können auch alte Abgaben kontrolliert werden. Dann können auch rückwirkend Punkte abgezogen werden.

Sie benötigen 67% der Hausaufgabenpunkte und eine aktive Mitarbeit in den Übungen für die Studienleistung.

Die „Wiederholungsaufgaben“, also Teil d) und e), sind nicht abzugeben. Beschäftigen Sie sich aber bitte auch mit diesen Aufgaben. Sie müssen damit rechnen, dass Sie beim Online-Treffen gebeten werden, einen Teil des Vorlesungs-Stoffes zu wiederholen und insbesondere eine der Wiederholungsfragen zu beantworten.

Verwenden Sie nochmals die XML-Datensammlung für klassische Musik-CDs (die deutsche Version):

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

Das Dokument enthält die folgenden Elemente:

- CDDB: (komponisten, cds?, solisten?)
- komponisten: (komponist*)
- komponist: (stueck*)
Attribute: knr, vorname, name, geboren, gestorben.
- stueck: (aufnahme*)
Attribute: snr, titel, tonart, opus.
- aufnahme: Leerer Inhalt.
Attribute: aufnnr, orchester, leitung.
- cds: (cd*).
- cd: (CDAufnahme*)
Attribute: cdnr, name, hersteller, anz-cds, gesamtspielzeit.
- CDAufnahme: Leerer Inhalt.
Attribute: aufnnr.
- solisten: (solist*).
- solist: (solist_aufnahme*)
Attribute: name.
- solist_aufnahme: Leerer Inhalt.
Attribute: aufnnr, instrument.

Schreiben Sie die folgenden Anfragen in XQuery und **testen Sie Ihre Lösung mit einer XQuery-Implementierung**.

- a) Erzeugen Sie ein XML-Dokument, das alle Komponisten der CD-Datenbank enthält, aber mit den Attribut-Werten in geschachtelten Elementen. Die Ausgabe soll also so aussehen:

```
<HA13>
  <komponist>
    <vorname>Georg Friedrich</vorname>
    <name>Händel</name>
    <geboren>1685</geboren>
    ...
  </komponist>
  <komponist>
    <vorname>Serge</vorname>
    <name>Prokofiev</name>
    <geboren>1891</geboren>
    ...
  </komponist>
  ...
</HA13>
```

Die Reihenfolge der Elemente ist beliebig. Natürlich ist eine Lösung gewünscht, in der Sie nicht alle Attribute einzeln behandeln.

- b) Erzeugen Sie eine HTML Tabelle mit allen vorkommenden Tonarten zusammen mit der Anzahl Stücke in der jeweiligen Tonart. Die Ausgabe soll folgendermaßen aussehen:

```
<table>
  <tr><th>Tonart</th><th>Anzahl</th></tr>
  <tr><td>A-Dur</td><td>1</td></tr>
  <tr><td>B-Dur</td><td>9</td></tr>
  <tr><td>C-Dur</td><td>7</td></tr>
  <tr><td>D-Dur</td><td>23</td></tr>
  ...
  <tr><td>a-moll</td><td>4</td></tr>
  ...
</table>
```

Ordnen Sie die Ausgabe so, dass alle Dur-Tonarten (Suffix “-Dur”) vor allen Moll-Tonarten (Suffix “-moll”) kommen, innerhalb jeder Gruppe alphabetisch. (Das ist natürlich musikalisch nicht richtig, aber die Information über die Vorzeichen fehlt leider.)

- c) Wählen Sie mindestens eine der SQL-Anfragen aus den Wiederholungsaufgaben d) und geben Sie eine sinnvolle Übersetzung nach XQuery an. Sie können selbst entscheiden, wie das Ergebnis-Dokument genau strukturiert ist.

Wiederholungsaufgaben

Beschäftigen Sie sich mit diesen Aufgaben. Sie brauchen aber nichts abzugeben.

- d) Was würden Sie in einer mündlichen Prüfung auf folgende Fragen zu XQuery antworten? Die Beispiele beziehen sich auf die Datei

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

Diese Datei hat außen ein Element **EMPDEPT**, darin geschachtelt **DEPT**-Elemente für die Abteilungen, und darin wieder geschachtelt **EMP**-Elemente für die Angestellten der jeweiligen Abteilung. Die eigentlichen Daten stehen in Attributen. Da die Zugehörigkeit zu einer Abteilung durch Schachtelung ausgedrückt ist, haben die **EMP**-Elemente kein Attribut **DEPTNO**.

1. Welche Aggregationsfunktionen gibt es in XPath/XQuery? Wie kann eine einfache Aggregationsanfrage von SQL nach XQuery übersetzt werden? Erläutern Sie das an folgendem Beispiel:

```
SELECT SUM(E.SAL)
FROM   EMP E
WHERE  E.JOB = 'CLERK' OR E.JOB = 'MANAGER'
```

2. Wie können **GROUP BY** Anfragen von SQL nach XQuery übersetzt werden? Geben Sie dazu eine Entsprechung zur folgenden SQL-Anfrage in XQuery an:

```
SELECT D.DEPTNO, D.DNAME, MIN(E.SAL), MAX(E.SAL)
FROM   EMP E, DEPT D
WHERE  E.DEPTNO = D.DEPTNO
GROUP BY D.DEPTNO, D.DNAME
```

3. Wie kann man Duplikate in XPath/XQuery entfernen? Ein einfaches Beispiel ist:

```
SELECT DISTINCT E.JOB
FROM   EMP E, DEPT D
WHERE  E.DEPTNO = D.DEPTNO
AND    D.DNAME = 'RESEARCH'
```

4. Ein komplizierteres Beispiel für die Duplikat-Elimination ist:

```
SELECT DISTINCT JOB, SAL
FROM   EMP
```

5. Übersetzen Sie folgende SQL-Anfrage mit `NOT EXISTS` nach XQuery:

```
SELECT D.DNAME
FROM   DEPT D
WHERE  NOT EXISTS(SELECT *
                  FROM   EMP E
                  WHERE  E.DEPTNO = D.DEPTNO
                  AND    D.SAL >= 3000)
```

6. Wie würden Sie folgende Anfrage mit einem äußeren Verbund übersetzen?

```
SELECT BOSS.EMPNO, BOSS.ENAME, COUNT(E.EMPNO)
FROM   EMP BOSS LEFT JOIN EMP E ON BOSS.EMPNO = E.MGR
GROUP BY BOSS.EMPNO, BOSS.ENAME
ORDER BY BOSS.EMPNO
```

Was würde sich bei einem normalen Verbund („`INNER JOIN`“) ändern?

Falls Sie SQL-Anfragen ausprobieren wollen, können Sie das mit PostgreSQL über die Adminer-Schnittstelle machen (Passwort in EMail):

```
[https://dbs.informatik.uni-halle.de/edb?pgsql=db&
username=student_gast&db=postgres&ns=empdept_public]
```

Für Interessierte

e) Schauen Sie sich folgende Seiten über `eXist-db` an:

- [<https://en.wikipedia.org/wiki/EXist>]
- [<http://exist-db.org/exist/apps/homepage/index.html>]
- [<https://exist-db.org/exist/apps/doc/>]
- [<https://glossar.hs-augsburg.de/EXist>]
- [<http://webdam.inria.fr/Jorge/html/wdmpa1.html>]
(das 5. Kapitel ist über `eXist-db`).
- [<http://cluster.cis.drexel.edu:8080/exist/quickstart.xml>]
- [<https://www.youtube.com/watch?v=xvMau2aHRDo>]
(Youtube-Video: First Steps with `eXist-db` 2.0)
- [<https://wwwdh.cs.fau.de/IMMD8/Lectures/DIGIDOK/10a-xquery-exist-2.pdf>]
(Vorlesung, letzter Abschnitt über `eXist-db`)
- [https://ebeshero.github.io/UpTransformation/xquery_setup.html]
(Vorlesung, Anfang über `eXist-db`).
- [<https://www.oreilly.com/library/view/exist/9781449337094/ch04.html>]
Beispiel-Kapitel aus dem Buch „`eXist`“ von Erik Siegel, Adam Retter.