

XML und Datenbanken

— 8. Übungsblatt: XPath —

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, 12.12.2022, 16⁰⁰, über die Übungsplattform in StudIP ab. Schreiben Sie die Lösungen in eine `.txt`- oder `.xpath`- 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 g) bis i), 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 Fragen aus g) zu beantworten.

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/xml22/examples/empdept.xml>]
- [<http://users.informatik.uni-halle.de/~brass/xml22/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>
```

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

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

Ich empfehle aber auch, BaseX auf dem eigenen Computer zu installieren (es kann XQuery-Anfragen auswerten, und XPath ist eine Teilmenge von XQuery):

[<https://basex.org/>]

- a) Drucken Sie die Namen (ENAME) aller Angestellten in einer Abteilung in "DALLAS" (LOC).
- b) Geben Sie alle Daten des Angestellten "FORD" (ENAME) aus.
- c) Was ist der Beruf (JOB) des Angestellten "SCOTT"?
- d) Welche Abteilungen (DNAME) haben Angestellte mit dem Beruf (JOB) "ANALYST"?
- e) 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.)
- f) Welche Abteilung hat keine Angestellten?

Wiederholungsaufgaben

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

g) Was 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).

h) Ich möchte Sie auf die Spezifikationen des W3C hinweisen. Zur XML Syntax:

- Extensible Markup Language (XML) 1.0 (Fifth Edition):
[<https://www.w3.org/TR/xml/>]
- Extensible Markup Language (XML) 1.1 (Second Edition):
[<https://www.w3.org/TR/xml11/>]
- The Annotated XML Specification (Tim Bray):
[<https://www.xml.com/axml/axml.html>]

Zum Datenmodell:

- XQuery 1.0 and XPath 2.0 Data Model (XDM) (Second Edition):
[<https://www.w3.org/TR/2010/REC-xpath-datamodel-20101214/>]
- XQuery and XPath Data Model 3.1:
[<https://www.w3.org/TR/xpath-datamodel-31/>]

Zu XPath:

- XML Path Language (XPath), Version 1.0 (alt, für viele Browser noch relevant):
[<https://www.w3.org/TR/1999/REC-xpath-19991116/>]
- XML Path Language (XPath) 2.0 (Second Edition):
[<https://www.w3.org/TR/xpath20/>]
- XML Path Language (XPath) 3.1:
[<https://www.w3.org/TR/xpath-31/>]

Für Interessierte

- i) Hier finden Sie eine Erläuterung zum XQuery and XPath Data Model von IBM für die Db2 Datenbank (mit XQuery Unterstützung):
[<https://www.ibm.com/docs/en/db2/11.5?topic=overview-xquery-xpath-data-model>]