

# XML und Datenbanken

## — 1. Übungsblatt: XML Daten —

### 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 beiden Aufgaben bis Montag, 17.10.2022, 16<sup>00</sup>, über die Übungsplattform in StudIP ab. Schreiben Sie die Lösungen jeweils in eine `.xml`-Datei. Es gibt 4 Punkte für Teil a) und 6 Punkte für Teil b).

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 c) und d), 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 c) zu beantworten.

- a) Das folgende XML-Dokument enthält mindestens 8 Fehler. Finden Sie möglichst viele davon und korrigieren Sie das Dokument. [Aufgabe von Dr. Henning Thielemann]

```
<Film 6jahre=freigegeben>
  <titel>Herr Fuchs und Frau Elster</titel>
  <regie>Harald Serowski & Erich Hammer<regie/>
  < altersbeschränkung typ = "für sechsjährige" >
  <beschreibung>
    Wenn es eine Personifizierung von <quote>Neugier</quote> gibt,
    dann ist es <b><i>Frau Elster</i></b>.
    In diesem heiteren Streifen für die ganze Familie
    ist sie wieder voll in ihrem Element.
    ...
  </beschreibung>
</film>
```

Dieses Dokument steht auch unter folgender URI zur Verfügung:

[[http://www.informatik.uni-halle.de/~brass/xml22/examples/err\\_film.xml](http://www.informatik.uni-halle.de/~brass/xml22/examples/err_film.xml)]

Sie können Ihre XML-Dokumente wie folgt prüfen lassen:

- Die Prüfung auf Wohlgeformtheit ist mit jedem modernen Browser möglich, z.B. dem Firefox. Speichern Sie das Dokument mit der Endung “.xml” und öffnen Sie es mit dem Web Browser. Das Dokument ist wohlgeformt, wenn der Browser die Daten des Dokuments anzeigt. Es ist normal, dass der Browser eine Warnung ausgibt, dass es zu dem XML Dokument kein Stylesheet gibt (und er es deswegen nicht als HTML anzeigen kann).
- [<http://www.xmlvalidation.com/>]  
(Da es keine DTD gibt, zeigt dieser Dienst am Ende an, dass das Element “Film” nicht definiert ist. Hier gibt es keine Prüfung nur auf Wohlgeformtheit.)
- [[http://www.w3schools.com/xml/xml\\_validator.asp](http://www.w3schools.com/xml/xml_validator.asp)]
- Sie könnten auch die AltovaXML Community Edition auf Ihrem PC installieren:  
[<http://www.softpedia.com/get/Internet/Other-Internet-Related/AltovaXML.shtml>]  
Das Programm hat nur eine einfache Kommandozeilen-Schnittstelle, aber es enthält einen Validator (für DTDs und XML Schema), einen XSLT Prozessor, und eine XQuery Implementierung. Für eine graphische Benutzerschnittstelle und weitere Features muss man Altova XMLSpy kaufen.
- Für spätere Aufgaben mit XQuery empfehle ich die XML-Datenbank BaseX auf Ihrem Rechner zu installieren. Wenn Sie eine neue Datenbank aus einer XML Datei erzeugen, wird diese natürlich auch auf Syntaxfehler geprüft.  
[<http://basex.org/>]

b) Angenommen, Sie wollen Ihre Lebensmittel-Einkäufe erfassen und haben sich für eine XML-Datei entschieden. Im Prinzip geht es also um die Speicherung von Kassensbons. Überlegen Sie sich, wie Sie eine XML Datei strukturieren wollen, die folgende Informationen enthält:

- Zu jeder Ware ist ein eindeutiges Kürzel (eine ID), sowie Name und eine Kategorie abzuspeichern.
- Zu jedem Einkauf ist das Datum und der Laden zu speichern.
- Außerdem ist für jeden Einkauf zu speichern, welche Waren Sie in welcher Menge zu welchem Preis gekauft haben. Dabei ist der Preis für ein Stück bzw. ein kg zu speichern. Die Menge kann bei Waren, die einen Preis pro kg haben, auch gebrochene Werte annehmen. Optional kann eine Bemerkung gespeichert werden (z.B. ein Hinweis auf ein besonderes Sonderangebot).

Zur Verdeutlichung hier eine relationale Repräsentation einiger Beispieldaten (auf die ID der Ware und/oder die Einkaufs-Nummer können Sie je nach Repräsentation der Daten eventuell verzichten):

WAREN		
ID	NAME	KATEGORIE
WM	Weidemilch 3.5%	Milch
KZB	Käse-Zwiebel-Brötchen	Brot
BAN	Bananen	Obst
TRH	Trauben hell	Obst

EINKAUF		
NR	DATUM	LADEN
1	11.10.2022	Aldi
2	15.10.2022	NP

GEKAUFT				
NR	ID	PREIS	MENGE	BEM
1	WM	1.45	1	
1	KZB	0.49	2	
1	BAN	1.00	0.5	
1	TRH	4.98	0.4	
2	BAN	1.49	1.2	WWF-Bananen

Schreiben Sie eine syntaktisch korrekte XML-Datei mit den Daten aus den Tabellen oder anderen Daten in gleicher Größenordnung. Die Datei braucht keine DOCTYPE-Deklaration zu enthalten. Sie können davon ausgehen, dass UTF-8 als Zeichencodierung verwendet wird (so dass auch die XML-Deklaration am Anfang entfallen kann).

Falls Sie Verbesserungsvorschläge zum Datenbank-Entwurf haben, können wir das gern in der Übung diskutieren.

## Wiederholungsaufgaben

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

- c) Wie würden Sie in einer mündlichen Prüfung auf folgende Fragen antworten?
- Was sind Kennzeichen semistrukturierter Daten im Vergleich zu einer relationalen Datenbank?
  - Für welche Zwecke ist XML nützlich, selbst wenn man die Daten in einer relationalen Datenbank gespeichert hat?
  - Nennen Sie eine Anwendung von XML.
  - Beschreiben Sie die Baumstruktur zur Darstellung von XML-Dokumenten im Hauptspeicher (soweit bisher besprochen).
  - Was ist die Beziehung zwischen Tags und Elementen?
  - Welche Arten von Tags gibt es? Beschreiben Sie jeweils die Syntax.
  - Welche Zeichen sind in der Angabe des Attributwertes verboten? Wie kann man sie dennoch hineinbekommen?
  - Was sind Unterschiede zwischen XML und HTML?
  - Wie sehen Zeichenreferenzen aus? Spielt die für die XML-Datei verwendete Zeichencodierung hier eine Rolle?
- d) Schauen Sie sich eine XML-Datei im Browser an, z.B.

[[https://www.ibiblio.org/xml/examples/shakespeare/much\\_ado.xml](https://www.ibiblio.org/xml/examples/shakespeare/much_ado.xml)]