

## XML und Datenbanken — 2. Übungsblatt: XML DTDs —

### 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, 24.10.2022, 16<sup>00</sup>, über die Übungsplattform in StudIP ab. Schreiben Sie die Lösungen jeweils in eine `.xml`-Datei. Es gibt 6 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) Erstellen Sie eine DTD für Ihr XML Dokument mit den Daten von Kassenbons gemäß Aufgabe b) vom letzten Übungsblatt. Sie dürfen auch Ihre XML-Datei bei Bedarf auch ändern (insbesondere sollten Sie Redundanzen vermeiden, wie in der letzten Übung besprochen). Geben Sie das XML-Dokument mit der eingebetteten DTD ab. Benutzen Sie unbedingt einen Validator, und schreiben Sie in einen Kommentar, welchen. Falls der Validator die DTD in einer getrennten Datei erwartet, können Sie in der Abgabe die DTD auch oben als Kommentar einfügen.

Falls es Bedingungen gibt, die die DTD nicht überwachen kann, die aber offensichtlich gelten sollten, schreiben Sie auch diese als Kommentar in Ihre Datei. Natürlich wäre es günstig, die Struktur der Datei und die Spezifikationen in der DTD so zu wählen, dass möglichst wenig zusätzliche Integritätsbedingungen anderweitig zu prüfen wären.

Hier nochmals die Beispieldaten vom letzten Übungsblatt (je nach Repräsentation der Daten können Sie auf die ID der Ware und/oder die Einkaufs-Nummer 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

- b) Ein Feuerwerker verkauft zu Silvester auch Feuerwerksartikel. Sein Angebot hat er in einer XML-Datei erfasst (siehe auch nächste Seite):

[<http://www.informatik.uni-halle.de/~brass/xml22/examples/angebot.xml>]

Um seine Verkaufspreise zu überprüfen, hat er in der Datei auch einige Konkurrenzpreise notiert. Entwickeln Sie eine passende DTD. Das Angebot soll mindestens einen Artikel enthalten. Konkurrenz-Preise muss es dagegen nicht unbedingt geben.

Das Attribut `Art` soll nur folgende Werte annehmen können:

- `Batt` (Batterie),
- `Rak` (Rakete),
- `Font` (Fontäne/Vulkan) und
- `Sonst` (Sonstiges).

Falls es nicht angegeben ist, soll der Wert „`Sonst`“ automatisch eingesetzt werden.

```
<?xml version="1.0">
<Angebot>
  <Artikel Marke="Funke" Name="Weißblinker-Pink"
    Nr="FC30-16-8" Art="Batt"
    Preis="27.00">
    <Text>
      16 Schuss-Batterie im Kaliber 30mm.
      Feuertöpfe und Buketts aus weißen Blinkern
      mit pinken Dahliensternen,
      mit weißem Blinkschweifaufstieg.
    </Text>
    <Konkurrenz Shop="Röder" Preis="28.95"/>
    <Konkurrenz Shop="Pyroland" Preis="29.74"/>
    <Konkurrenz Shop="pyroweb.de" Preis="33.29"
      Hinweis="Nicht lagernd"/>
    <Konkurrenz Shop="pyrohandel.de" Preis="24.50"
      Hinweis="Derzeit nicht lieferbar"/>
  </Artikel>
  <Artikel Marke="Jorge" Name="Eden" Art="Batt"
    Preis="22.00">
    <Text>
      36 Schuss Crossette Batterie mit bunten Crossette Sternen
      im Kaliber 20mm: rot, grün, blau, purpur, lemon und gelb.
    </Text>
    <Konkurrenz Shop="Röder" Preis="22.99"/>
    <Konkurrenz Shop="talfeuerwerk.de" Preis="19.95"
      Hinweis="Ausverkauft"/>
    <Konkurrenz Shop="Lichtenrader Feuerwerksverkauf"
      Preis="19.50"/>
  </Artikel>
  <Artikel Marke="JGW Berckholtz" Name="Pulsar"
    Preis="6.00">
    <Text>
      Feuertopf mit "Rauschgold" und weißen Blinksternen.
    </Text>
  </Artikel>
</Angebot>
```

## Wiederholungsaufgaben

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

c) Wie würden Sie in einer mündlichen Prüfung auf folgende Fragen antworten?

- Wie sieht ein Kommentar in XML aus? Wie kann man sich erklären, dass die Syntax doch recht umständlich ist?
- Wie unterscheiden sich ISO Latin 1 (ISO 8859-1) und UTF-8 bei der Codierung deutscher Umlaute?
- Was ist der Unterschied zwischen wohlgeformten (“well-formed”) und gültigen (“valid”) XML?
- Wofür steht die Abkürzung „DTD“?
- Nennen Sie mindestens eine Möglichkeit, um ein XML Dokument zu validieren (d.h. einen Validator, also ein Programm, mit kurzer Erläuterung zur Verwendung).
- Wie sieht eine Element-Deklaration in XML aus? Welche Konstrukte können in einer Inhalts-Spezifikation verwendet werden? Was bedeutet die Einschränkung, dass das Inhaltsmodell deterministisch sein muss? Geben Sie ein Beispiel für ein Inhaltsmodell, das verboten ist.
- Wofür steht “PC” bei “#PCDATA”? Warum muss man da ein “#” schreiben, aber bei “EMPTY” nicht?
- Im Prinzip sind XML Elementdeklarationen und die Regeln einer kontextfreien Grammatik (z.B. in BNF) doch recht ähnlich strukturiert. Was ist der wesentliche Unterschied von XML mit DTDs und konktextfreien Grammatiken hinsichtlich der definierbaren Sprachen?
- Wie werden Attribute in XML DTDs deklariert?
- Was sind die wichtigsten Datentypen für Attribute in XML DTDs? Nennen Sie möglichst mindestens fünf.
- Wie sieht eine Processing Instruction syntaktisch aus? Was ist das überhaupt? Warum gibt es die Ähnlichkeit mit der XML Deklaration?
- Welche Möglichkeit gibt es syntaktisch für eine DOCTYPE Deklaration?

d) Schauen Sie sich die XML 1.0 Spezifikation an:

[<https://www.w3.org/TR/REC-xml/>]

Alternativ können Sie auch “The Annotated XML Specification” von Tim Bray verwenden:

[<http://www.xml.com/axml/testaxml.htm>]

Finden Sie heraus, welche Zeichen als Startzeichen eines Bezeichners (z.B. als Namen von Elementen) erlaubt sind. (Tipp: Es steht in Abschnitt 2.3 “Common Syntactic Constructs”. Falls Sie die ältere “Annotated XML Specification” verwenden, müssen Sie noch dem Link nach Anhang B folgen.) Es reicht, wenn Sie sich auf ASCII-Zeichen konzentrieren. Es wäre aber auch spannend, die Frage für deutsche Umlaute zu klären. Die offizielle Quelle für die Codes der Umlaute im Unicode ist

[<http://unicode.org/charts/PDF/U0080.pdf>]

Es eignet sich aber jede ISO Latin 1 Tabelle (die Zahlwerte der Zeichen-Codes sind identisch, die Codierung mittels UTF-8 nicht). In den älteren Ausgaben des XML Standards sind auch die ASCII Buchstaben über ihre Codes angegeben:

[<http://unicode.org/charts/PDF/U0000.pdf>]