

**Klausur “Semistrukturierte Daten und XML”**  
**Sommersemester 2011**  
**Prof. Dr. Wolfgang May**  
**14. Juli 2011, 14-16 Uhr**  
**Bearbeitungszeit: 90 Minuten**

Vorname:

Nachname:

Matrikelnummer:

Bei der Klausur sind **keine Hilfsmittel** (Skripten, Taschenrechner etc.) erlaubt. Handies müssen ausgeschaltet sein. Papier wird gestellt. Benutzen Sie nur die **ausgeteilten**, zusammengehefteten **Blätter** für Ihre Antworten. Schreiben Sie mit blauem/schwarzem Kugelschreiber, Füller etc.; Bleistift ist nicht erlaubt.

Zum **Bestehen** der Klausur sind **45** Punkte hinreichend.

- meine Note soll mit Matrikelnummer so bald wie möglich auf der Vorlesungs-Webseite veröffentlicht werden.
- meine Note soll nicht veröffentlicht werden; ich erfahre sie dann vom Prüfungsamt (bzw. für nicht von einem Prüfungsamt verwaltete Teilnehmer: beim Abholen des Scheins).

	Max. Punkte	Erreichte Punkte
Aufgabe 1 (XML-Entwurf)	30	
Aufgabe 2 (XPath, XQuery)	40	
Aufgabe 3 (XSLT)	20	
Summe	90	

**Note:**

## Themenstellung: Fernbuslinien

Alle Klausuraufgaben basieren auf einem gemeinsamen “Auftrag”: Seit 2011 ist die Einrichtung von *Fernbuslinien* in Deutschland erlaubt (wie sie bereits in vielen Ländern existieren). In der Klausur soll eine Datenbank eines Fernbuslinienbetreibers entworfen werden:

- Der Betreiber will mehrere Fernbuslinien einrichten. Die meisten Linien sammeln ihre Fahrgäste in einer Region schrittweise auf, fahren dann eine längere Strecke am Stück, und lassen die Fahrgäste in der Zielregion an einer oder mehreren Stationen wieder aussteigen.
- Jede Haltestelle hat einen eindeutigen Namen (meistens der Name der Stadt; z.B. “Frankfurt”, aber auch “Frankfurt Flughafen”). Für jede Haltestelle ist der Name sowie die Strasse bzw. nähere Bezeichnung gespeichert (z.B. (Göttingen, ZOB Bahnhofplatz)).

Im den folgenden Aufgaben wird angenommen, dass “in XY” zu lesen ist als “an der Haltestelle, deren Name XY ist”.

- Jede Linie hat eine eindeutige Nummer und besteht aus einer Liste der Haltestellen mit (relativen) Abfahrtszeiten und Fahrpreisen (jeweils bezogen auf den Startort; bei späterem Zustieg wird der entsprechende Teil des Preises abgezogen); vgl. Spezifikation der Linien 2460 (Flensburg-Hamburg-Frankfurt) und 2060 (Hamburg-Frankfurt) auf dem letzten Blatt der Klausur.
- Der Fahrplan gibt für jede Linie Wochentage und die *absoluten* Startzeiten am Startort an. Zum Beispiel fährt Linie 2460 werktags um 5:30, 8:00, 12:00 und 17:30 Uhr in Hamburg ab; samstags/sonntags aber um 8:00, 13:00 und 16:30 Uhr. Nehmen Sie an, dass jede Linie zumindest werktags einmal befahren wird.

Der konkrete Einsatzplan ordnet jeder durchzuführenden Fahrt (Tag, Zeit) einen Bus sowie einen Fahrer zu:

- Jeder Bus hat eine (eindeutige) Nummer.
- Jeder Fahrer hat einen Namen (Annahme: Vor- und Nachname sind zusammen eindeutig).
- Für jeden Tag (Datum) ist jeder an dem Tag durchzuführenden Fahrt ein Bus sowie ein Fahrer zugeordnet. So wird z.B. die erste Fahrt der Linie 2060 (Hamburg – Frankfurt) am 14.07.2011 um 05:30 von Hans Meier mit dem Bus 278 durchgeführt. Derselbe Bus fährt dann um 14:00 als Linie 6020 mit Fahrer Hans Meier wieder nach Hamburg.

### Aufgabe 1 (XML-Entwurf [30 Punkte])

Die Datenspeicherung soll in XML stattfinden (nicht nur, weil das so in dieser Vorlesung behandelt wurde, sondern auch weil es für diesen Anwendungsfall ein sehr sinnvolles Datenformat ist; siehe auch Aufgabe 2d).

- a) Geben Sie die DTD Ihrer Lösung an (15 P).
- b) Geben Sie eine XML-Instanz an, die Ihre Lösung u.a. anhand der in der Spezifikation aufgeführten Beispieldaten illustriert. (10 P)
- c) Geben Sie einen Automaten an, mit dem der Elementtyp validiert werden kann, in dem Sie die einzelnen Buslinien speichern. (5 P)

– hier DTD (und wenn man will XML) angeben –

**Aufgabe 2 (XPath, XQuery [40 Punkte])**

Verwenden Sie für diese Aufgabe die von Ihnen entworfene XML-Datenbasis. Keine der Antworten soll Duplikate enthalten.

**Hinweis (Syntax):** Nehmen Sie für diese Aufgabe an, dass Sie mit dem Datentyp für "Zeit" rechnen (+ minuten, MAX, MIN) und vergleichen können (<, > etc.). Repräsentieren Sie Zeiten als Strings der Form "hh:mm" und Datumswerte als Strings der Form "YYYY-MM-DD". (Dies ist mit XML-Software auch (fast) genauso möglich.)

**Hinweis (Anwendung):** Nehmen Sie für diese Aufgabe an, dass folgendes gilt:

- es gibt keine Fahrten über Mitternacht,
  - in jeder Linie kann an jeder Haltestelle ein- und ausgestiegen werden
- a) Geben Sie eine XQuery-Anfrage an, die die Nummern aller Linien ausgibt, die werktags in Göttingen halten. (3 P)
- b) Erweiterung von a) – Welche Linien halten werktags in Göttingen?  
Für jede Fahrt soll die Liniennummer sowie die jeweilige Abfahrtszeit angegeben werden, und die Antworten sollen nach dieser Zeit geordnet sein, z.B.

```
<antwort linie="2060" zeit="11:10:00"/>
<antwort linie="6020" zeit="13:20:00"/>
<antwort linie="2060" zeit="16:10:00"/>
<antwort linie="6020" zeit="19:20:00"/>          (4 P)
```

- c) Geben Sie eine HTML-Liste aller Linien mit ihren Strecken als Liste der Haltestellen in folgender Form aus: (4 P)

```
<ul>
  :
  <li> 2060
    <ul><li>Hamburg</li> <li>Soltau</li> ...
      <li>Frankfurt Flughafen</li><ul>
    </li>
  <li> 2061 <ul> ... </ul>
  </li>
  :
</ul>
```

- d) Warum ist XML für diese Anwendung wesentlich besser geeignet, als eine relationale Datenbank und SQL? (überlegen Sie sich, wie man Aufgabenteil (c) in SQL lösen müsste). (4 P)
- e) Geben Sie eine Anfrage an, die die Nummern aller Linien ausgibt, mit denen man werktags ohne Umsteigen von Soltau nach Kassel fahren kann. (5 P)

- f) Geben Sie eine XQuery-Anfrage an, die ausgibt, an welchem Ort der Fahrer Hans Meier am 14.7. seinen Dienst beendet. (5 P)
- g) Geben Sie eine XQuery-Anfrage an, die ausgibt, welche Orte man werktags mit 1x umsteigen von Soltau aus erreichen kann. (10 P)

**h) (Updates, abstrakte Sprachsyntax)**

Linie 2060 soll zusätzlich zwischen Hannover und Göttingen noch in Hildesheim (30 Minuten ab Hannover, 3.50E zusätzlich) halten.

Geben Sie ein Statement in einer Syntax an, die auf XQuery aufbaut (Sie können XQuery+Updates verwenden, können sich aber auch prinzipiell an der SQL-Syntax orientieren), das dieses Update auf Ihrem XML-Dokument sinnvoll spezifiziert. (5 P)

Name:

MatNr.:

---



**Aufgabe 3 (XSLT [20 Punkte])**

Geben Sie ein XSLT-Stylesheet an, das eine XML-to-HTML-Transformation ausführt, das für jede Linie den Fahrplan in der folgenden Form als HTML-Tabelle ausgibt:

(Erzeugen Sie die richtige Struktur; Zentrierung, Linien etc. sind unwichtig. Wenn Sie sich bei bestimmten XSL-Kommandos nicht sicher sind, "erfinden" Sie eine *sinnvolle* Syntax.)

<b>2060</b>	Preis			
Hamburg	0	5:30	8:00	14:00
Soltau	6.00	6:20	8:50	14:50
Hannover	12.00	7:10	9:40	15:40
Göttingen	20.00	8:40	11:10	17:10
Kassel	25.00	9:20	11:50	17:50
Gießen	33.00	10:50	13:20	19:20
Frankfurt	39.00	11:40	14:10	20:10
Frankfurt Flughafen	40.00	12:00	14:30	20:30

Betrachten Sie dabei **nur den Werktags-Fahrplan**.





Name:

MatNr.:

[Trennen Sie dieses Blatt am besten vor Beginn der Bearbeitung ab]

Die folgenden Basisdaten für die Linie 2460 (Flensburg nach Frankfurt), 2060 (Hamburg nach Frankfurt mit mehreren Zwischenstops), und 6020 sollen (unter anderem) gespeichert werden:

<b>Linie</b>			
<b>Nr</b>	<b>Ort</b>	<b>Minuten</b>	<b>Preis</b>
:	:	:	:
2460	Flensburg	0	0.00
2460	Schleswig	25	2.00
2460	Rendsburg	55	4.00
2460	Neumünster	85	6.00
2460	Quickborn	120	9.00
2460	Pinneberg	135	10.00
2460	Hamburg	155	12.00
2460	Frankfurt	615	42.00
2460	Frankfurt Flughafen	640	43.00
2060	Hamburg	0	0.00
2060	Soltau	50	6.00
2060	Hannover	100	12.00
2060	Göttingen	190	20.00
2060	Kassel	230	25.00
2060	Gießen	320	33.00
2060	Frankfurt	370	39.00
2060	Frankfurt Flughafen	390	40.00
6020	Frankfurt Flughafen	00	0.00
6020	Frankfurt	20	1.00
:	:	:	:
6020	Hamburg	390	40.00
:	:	:	:

Z.B. Linie 2060 ist 190 Minuten nachdem der Bus in Hamburg losgefahren ist, in Göttingen. Bis dahin kostet die Fahrt 20.00E.



Name:

MatNr.:

---

Trennen Sie dieses Blatt ab und schreiben das XML-Fragment zu Aufgabe (1b) darauf.