

Gruppe A

Bitte tragen Sie **sofort** und **leserlich** Namen, Studienkennzahl und Matrikelnummer ein und legen Sie Ihren Studentenausweis bereit.

PRÜFUNG AUS "SEMISTRUKTURIERTE DATEN" 181.135			28. 11. 2008
Kennnr.	Matrikelnr.	Familienname	Vorname

Arbeitszeit: 120 Minuten. Aufgaben sind auf den Angabeblättern zu lösen; Zusatzblätter werden nicht gewertet.

Aufgabe 1:

(12)

Betrachten Sie die folgende DTD **test.dtd**:

```
<?xml version="1.0" encoding="UTF-8"?>
<!ELEMENT A (B*,C*)>
<!ELEMENT B (B?,C,C?)>
<!ELEMENT C (#PCDATA|D|E)*>
<!ATTLIST C F IDREF #IMPLIED>
<!ELEMENT D EMPTY>
<!ELEMENT E (#PCDATA)>
<!ATTLIST E P ID #REQUIRED>
```

Betrachten Sie weiters die acht verschiedenen XML-Dateien, die unten angeführt sind.

Hinweise:

- Gehen Sie davon aus, dass allen folgenden Dateien die Zeile
`<!DOCTYPE A SYSTEM "test.dtd">`
vorangestellt ist.
- Sie können auch davon ausgehen, dass alle folgenden XML-Dateien wohlgeformt sind. Es geht also lediglich darum, ihre Gültigkeit bezüglich **test.dtd** zu entscheiden.

Kreuzen Sie an, welche der folgenden XML-Dateien gültig bezüglich **test.dtd** sind.

1. `<A><C>cde<D>fgh</D></C>` gültig ungültig
2. `<A><C>cde<E P="n1">fgh</E></C>` gültig ungültig
3. `<A><C></C>` gültig ungültig
4. `<A><C/>` gültig ungültig
5. `<A><C F="n1">test</C><C><E P="n1"/></C>` gültig ungültig
6. `<A><C F="n1"><E P="n1"/><E P="n2">test</E></C>` gültig ungültig
7. `<A><C/><C/><C/>` gültig ungültig
8. `<A>` gültig ungültig

(Pro korrekter Antwort 1.5 Punkte, **pro inkorrektter Antwort -1.5 Punkte**, pro nicht beantworteter Frage 0 Punkte, für die gesamte Aufgabe mindestens 0 Punkte)

Vervollständigen Sie auf den nächsten Seiten die XML-Schema Definition **theater.xsd**, sodass das XML-Dokument **theater.xml** (siehe Anhang) bezüglich dieses Schemas gültig ist. (Sie haben auch auf der nächsten Seiten Platz!)

Berücksichtigen Sie dabei:

- Definieren Sie die fehlenden Typen “ensembleTyp” und “repertoireTyp”.
- Sie brauchen sich **nicht** um die Definition des Typs “spielplanTyp” zu kümmern!
- Für Ihre Definition des “ensembleTyp” beachten Sie:
 - Es handelt sich hier um eine beliebige Abfolge von “schauspieler”- und “schauspielerin”-Elementen. Es sollte jedoch insgesamt zumindest ein solches Element angegeben sein!
 - Ein “schauspieler”- bzw. “schauspielerin”-Element enthält genau ein “kennz”-Element und ein “name”-Element (Sie können sich hierfür auch einen zusätzlichen Typ definieren).
- Für Ihre Definition des “repertoireTyp” beachten Sie:
 - Es soll sichergestellt sein, dass zumindest ein “stueck”-element vorhanden ist; jedes “stueck”-Element besitzt genau ein “name”-element, genau ein “autor”-element, mindestens ein “person”-Element und ein “skz”-Attribut.
 - Das “person”-Element soll ein leeres Element mit den entsprechenden Attributen sein.
 - Im “beschreibung”-Element können “ort”- und “bezug”-Elemente sowie zusätzlicher Text beliebig oft und in beliebiger Reihenfolge auftreten.
 - Alle Attribute, die Sie verwenden, sollen verpflichtend sein.

Datei **theater.xsd**:

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">

  <xs:element name="theater">
    <xs:complexType>
      <xs:sequence>
        <xs:element name="ensemble" type="ensembleTyp"/>
        <xs:element name="repertoire" type="repertoireTyp"/>
        <xs:element name="spielplan" type="spielplanTyp"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>

  <xs:complexType name="spielplanTyp">
    <!-- ist nicht Teil der Prüfungsaufgabe! -->
  </xs:complexType>
```

Datei **theater.xsd** (Fortsetzung):

</xs:schema>

Vervollständigen Sie **auf der nächsten Seite** das XSLT-Stylesheet **theater.xsl**, das auf Instanz-Dokumente des Schemas **theater.xsd** aus der vorigen Aufgabe angewandt werden kann und ein XML-Dokument gemäß folgender DTD liefert:

```
<?xml version="1.0" encoding="UTF-8"?>
<!ELEMENT auftritte (name,spielt*)*>
<!ELEMENT name (#PCDATA)>
<!ELEMENT spielt (#PCDATA)>
```

Erläuterungen zu dieser dtd:

- Das gesuchte Stylesheet soll für jede Schauspielerin alle Aufführungen, in der sie mitwirkt, sammeln.
- Das “name”-Element enthält den Name der Schauspielerin aus dem “schauspielerin” Element aus **theater.xml**
- Das “spielt”-Element besteht aus Text der sich folgendermaßen zusammensetzt:
 - den fixen Text “am ”
 - das Datum der Aufführung aus dem entsprechenden Attribut im Spielplan (siehe **theater.xml**)
 - den fixen Text “in: ”
 - den Namen des Stückes (Subelement von “stueck” in **theater.xml**).
- Beachten Sie dass mittels des “kennz”-Elementes im “besetzung”-Element die entsprechenden “auffuehrung”-Elemente gefunden werden können! Nutzen Sie das Attribut “stueck” in “auffuehrung” um den vollen Namen des Stückes zu ermitteln.

Beispiel: Angewandt auf das XML-Dokument **theater.xml** soll **theater.xsl** daher folgendes Ergebnis liefern (Anmerkung: die Einrückungen dienen nur zur besseren Lesbarkeit und müssen nicht berücksichtigt werden).

```
<?xml version="1.0" encoding="UTF-8"?>
<auftritte>
  <name>Alexandra Henkel</name>
  <spielt>am 2008-11-08 in: Das Haus des Richters</spielt>
  <name>Birgit Minichmayr</name>
  <spielt>am 2008-11-08 in: Der Weibsteufel</spielt>
  <spielt>am 2008-12-19 in: Macbeth</spielt>
  <spielt>am 2008-12-20 in: Macbeth</spielt>
  <spielt>am 2009-01-04 in: Der Weibsteufel</spielt>
  <name>Dorothee Hartinger</name>
  <spielt>am 2008-11-08 in: Das Haus des Richters</spielt>
  <name>Johanna Wokalek</name>
</auftritte>
```

Datei **theater.xsl**:

```
<?xml version="1.0" encoding="UTF-8"?>
<xsl:stylesheet version="1.0" xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
<xsl:output method="xml" indent="yes" encoding="UTF-8" version="1.0" />

<xsl:template match="/">

<!-- Vervollständigen Sie dieses Template -->
```

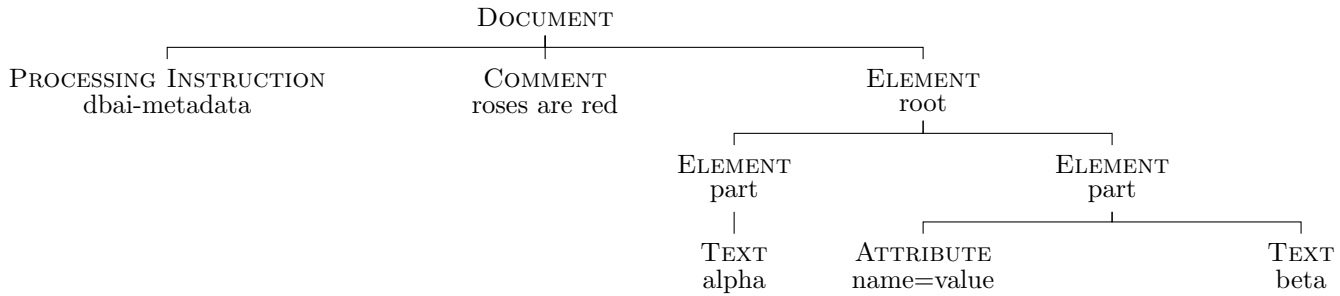
```
</xsl:template>
```

```
</xsl:stylesheet>
```

Aufgabe 4:

(7)

Geben Sie das XML-Dokument an, das dem folgenden DOM-Baum entspricht. Die XML-Deklaration (<?xml...?>) brauchen Sie nicht zu berücksichtigen.

**Aufgabe 5:**

(9)

Vervollständigen Sie die folgende Java Klasse sodass ein SAX Content-Handler herauskommt, der alle verwendeten *Namespaces* auf der Konsole ausgibt. Stellen Sie sicher dass Namespaces nicht doppelt ausgegeben werden. Ein einfacher String-Vergleich reicht aus um die Gleichheit fest zu stellen.

Tipp: Verwenden Sie eine **Set** Datenstruktur um doppelte Namespaces zu filtern! Um Objekte in ein **Set** zu speichern verwenden Sie `namespaceSet.put(...)`, um alle Objekte aus einem **Set** zu extrahieren können Sie die kompakte Java-Schreibweise verwenden: `for(String ns : namespaceSet) { ... }`.

```
class NamespacePrinterHandler extends DefaultHandler {
    private Set<String> namespaceSet = new HashSet<String>();

    }
}
```

Aufgabe 6:

(8)

Schreiben Sie XPath-Anfragen um für XML-Dokumente wie **theater.xml** (siehe Anhang) folgende Informationen zu selektieren:

1. Selektieren Sie das Kürzel (**kennz**) von “Markus Hering”.
2. Zählen Sie wie oft das Stück “Macbeth” aufgeführt wird. Sie dürfen das Kürzel “m” nicht direkt verwenden!
3. Selektieren Sie alle **beschreibung** Elemente die kein **ort** Element enthalten.
4. Selektieren Sie alle **auffuehrung** Elemente in denen der Schauspieler “bm” und “dk” gemeinsam vorkommen.

Aufgabe 7:

(12)

Kreuzen Sie an, ob die folgenden Aussagen wahr oder falsch sind.

1. SAX ist ein “Pull-Parser”. wahr falsch
2. Mit XML-Schema ist es möglich die Struktur von XML-Dokumenten genauer zu beschreiben als mit DTDs. wahr falsch
3. Per Default erlaubt XML-Schema bei der Elementdeklaration dass ein Element beliebig oft, aber auch gar nicht vorkommen darf. wahr falsch
4. Der XPath-Ausdruck `./a[@b=c]` ist die Kurzschreibweise des XPath-Ausdrucks `self::node()/child::a[attribute::b = child::c]` wahr falsch
5. Folgende zwei Ausdrücke sind äquivalent:
`//paragraph[last()][i]`
`//paragraph[i][last()]` wahr falsch
6. Bei XSLT wird pro Knoten maximal ein Template angewendet. wahr falsch
7. Um die Wohlgeformtheit eines XML-Dokuments zu überprüfen wird eine DTD oder ein XML-Schema benötigt. wahr falsch
8. Die **descendant**-Achse von XPath liefert alle Nachfahren des aktuellen Knotens *inklusive* des aktuellen Knoten zurück. wahr falsch

(Pro korrekter Antwort 1.5 Punkte, **pro inkorrektter Antwort -1.5 Punkte**, pro nicht beantworteter Frage 0 Punkte, für die gesamte Aufgabe mindestens 0 Punkte)

Sie können diese Seite abtrennen!

Das folgende XML-Dokument **theater.xml** gilt für Aufgaben 2,3,6 (**Rückseite beachten!**):

```
<?xml version="1.0" encoding="UTF-8"?>
<theater>
  <ensemble>
    <schauspielerin> <kennz>ah</kennz> <name>Alexandra Henkel</name> </schauspielerin>
    <schauspielerin> <kennz>bm</kennz> <name>Birgit Minichmayr</name> </schauspielerin>
    <schauspielerin> <kennz>dh</kennz> <name>Dorothee Hartinger</name> </schauspielerin>
    <schauspieler> <kennz>dk</kennz> <name>Dietmar Koenig</name> </schauspieler>
    <schauspielerin> <kennz>jw</kennz> <name>Johanna Wokalek</name> </schauspielerin>
    <schauspieler> <kennz>mh</kennz> <name>Markus Hering</name> </schauspieler>
    <schauspieler> <kennz>no</kennz> <name>Nicholas Ofczarek</name> </schauspieler>
    <schauspieler> <kennz>ww</kennz> <name>Werner Woelbern</name> </schauspieler>
  </ensemble>
  <repertoire>
    <stueck skz="w">
      <name>Der Weibsteufel</name>
      <autor>Karl Schoenherr</autor>
      <person id="w1" name="Der Mann"/>
      <person id="w2" name="Seine Frau"/>
      <person id="w3" name="Junger Grenzjaeger"/>
      <beschreibung>
        Ein Mann, der von der Hehlerei geschmuggelter Waren lebt, erfahrt, dass ihm der
        oertliche Polizeikommandant nach vielen vergeblichen Versuchen endlich das Handwerk
        legen will, indem er einen jungen Grenzjaeger auf seine attraktive Frau "ansetzt".
        Der Grenzjaeger verspricht sich eine schnelle Befoerrderung, ...
      </beschreibung>
    </stueck>
    <stueck skz="m">
      <name>Macbeth</name>
      <autor>William Shakespeare</autor>
      <person id="m1" name="Macbeth"/>
      <person id="m2" name="Lady Macbeth"/>
      <person id="m3" name="Macduff"/>
      <!-- ..... -->
      <beschreibung>
        Schicksalsgestalten hatten Lord Macbeth und seiner Lady die Krone
        <ort>Schottlands</ort> vorhergesagt. Der Erfuellung dieser Weissagung half das Paar
        gewaltsam nach, ...
      </beschreibung>
    </stueck>
    <stueck skz="h">
      <name>Das Haus des Richters</name>
      <autor>Dimitre Dinev</autor>
      <person id="h1" name="Ada"/>
      <person id="h2" name="Sohn"/>
      <person id="h3" name="Wera"/>
      <!-- ..... -->
      <beschreibung>
        Das Stueck ist eine moderne <ort>mitteleuropaeische</ort> Variante auf den
        urspruenglich <bezug>kretischen Minotauros-Mythos</bezug>, in dem Koenig Minos den
        Kuenstler Daidalos mit dem Bau eines Verwahrungsortes fuer seinen missgebildeten
        und als gefaehrlich geltenden Sohn beauftragt....
      </beschreibung>
    </stueck>
  </repertoire>
```

BITTE WENDEN. Fortsetzung auf Rückseite!

<!-- FORTSETZUNG des Dokuments theater.xml -->

```
<spielplan>
  <auffuehrung datum="2008-11-08" stueck="w">
    <besetzung kennz="ww" spielt="w1"/>
    <besetzung kennz="bm" spielt="w2"/>
    <besetzung kennz="no" spielt="w3"/>
  </auffuehrung>
  <auffuehrung datum="2008-11-08" stueck="h">
    <besetzung kennz="ah" spielt="h1"/>
    <besetzung kennz="dk" spielt="h2"/>
    <besetzung kennz="dh" spielt="h3"/>
  </auffuehrung>
  <auffuehrung datum="2008-12-19" stueck="m">
    <besetzung kennz="dk" spielt="m1"/>
    <besetzung kennz="bm" spielt="m2"/>
    <besetzung kennz="mh" spielt="m3"/>
  </auffuehrung>
  <auffuehrung datum="2008-12-20" stueck="m">
    <besetzung kennz="dk" spielt="m1"/>
    <besetzung kennz="bm" spielt="m2"/>
    <besetzung kennz="mh" spielt="m3"/>
  </auffuehrung>
  <auffuehrung datum="2009-01-04" stueck="w">
    <besetzung kennz="ww" spielt="w1"/>
    <besetzung kennz="bm" spielt="w2"/>
    <besetzung kennz="no" spielt="w3"/>
  </auffuehrung>
</spielplan>
</theater>
```

Gesamtpunkte: 75