

Übungsblatt 4 WT:III

Bis zum 13.06.2019, 23:59, sind Lösungen zu folgenden Aufgaben abzugeben: 1, 2, 4, 5a,b, 6, 7a-d.

Hinweis: Testen Sie Ihre XPath's mittels xpathtester.com.

Aufgabe 1 : XPath (1+1+1+1 Punkte)

Gegeben sei folgende XML-Datei.

```
<?xml version="1.0" encoding="UTF-8"?>
<XMLBooks>
  <book>
    <title>XML in a Nutshell</title>
    <authors>
      <author>Elliotte Rusty Harrold</author>
      <author>W. Scott Means</author>
    </authors>
    <publisher>O'Reilly</publisher>
    <address>Bejing (China), Cambridge (UK), et al.</address>
  </book>
  <book>
    <title>The XML Handbook</title>
    <authors>
      <author>Charles F. Goldfarb</author>
      <author>Paul Prescod</author>
    </authors>
    <publisher>Prentice Hall</publisher>
    <address>Upper Addle River (USA)</address>
  </book>
  <book>
    <title>XSLT Quickly</title>
    <authors>
      <author>Bob DuCharme</author>
    </authors>
    <publisher>Manning</publisher>
    <address>Greenwich (USA)</address>
  </book>
</XMLBooks>
```

- (a) Geben Sie den XPath-Ausdruck an, der alle Bücher die in Großbritannien (UK) erschienen sind, lokalisiert. (Sie können hierfür die Funktion `contains(.,.)` verwenden.)
- (b) Geben Sie einen XPath-Ausdruck an, der die Titel aller Bücher lokalisiert, die durch mehr als einen Autor geschrieben wurden. (Sie können hierfür die Funktion `count(.)` verwenden.)
- (c) Welche(n) Knoten lokalisiert folgender Ausdruck?
`/XMLBooks/book[(position() div 2) = 1]`
- (d) Welche(n) Knoten lokalisiert folgender Ausdruck?
`//book[authors/author[.='W. Scott Means']]/title`

Aufgabe 2 : XSLT: Built-in Templates (1+1+1 Punkte)

Geben Sie für jede der folgenden Template-Regeln an (1) auf welche Knoten eines Dokumentbaumes die Template-Regel angewandt werden kann und (2) was für diese Knoten getan werden soll.

- (a)

```
<xsl:template match="*/">
  <xsl:apply-templates/>
</xsl:template>
```
- (b)

```
<xsl:template match="text()|@*">
  <xsl:value-of select="."/>
</xsl:template>
```
- (c)

```
<xsl:template match="processing-instruction()|comment()"/>
```

Aufgabe 3 : XSLT: Peanuts

In dieser Datei `peanuts.xml` aus dem Repository dieses Beleges sind einige Namen und weitere Informationen enthalten. Stellen Sie die Informationen aus der Datei als HTML-Dokument dar. Das HTML-Dokument soll die Überschrift „Darsteller der Peanuts“ enthalten, sowie eine dreispaltige Tabelle mit den Spaltennamen „Vorname“, „Name“ und „Einführungsjahr“. Erstellen Sie ein eigenes XSLT-Stylesheet für jede Unteraufgabe.

- (a) Geben Sie ein XSLT-Stylesheet an, dass die Tabelle zusätzlich nach Nachnamen sortiert.
- (b) Geben Sie ein XSLT-Stylesheet an, in der die Tabelle nur Darsteller enthält, die nach dem Jahr 1960 eingeführt wurden.

Aufgabe 4 : Programmiermodul: XSL-Transformation P (3 Punkte)

Schreiben Sie eine XSL-Transformation um einen XML-Feed in eine HTML-Oberfläche zu überführen.

- Referenzieren Sie Ihre XSLT-Datei in der `feed.xml` der letzten Übung.
- Die Transformation soll die `feed.xml` in Ihr HTML aus dem Programmiermodul GUI oder in das der entsprechenden Musterlösung (Kursseite: in `feed-reader-gui-final.zip`) überführen.
- Entnehmen Sie so viele Informationen wie möglich der `feed.xml`, zum Beispiel auch die Seitenbeschreibung aus `<title>` und `<subtitle>`.
- Achten Sie auf eine detailgetreue Umsetzung, zum Beispiel bei der Ausgabe des Datums.
- Es soll immer der komplette Feed angezeigt werden.

Hinweis: Nicht alle Browser laden XSL-Transformationen von lokalen XML-Dateien. Nutzen Sie im Zweifel Firefox um Ihre Abgabe zu testen.

Aufgabe 5 : Wikipedia-XML-Dump (1+2+0 Punkte)

Laden Sie sich die Datei [ndswiki-20190501-pages-articles-multistream.xml.bz2](https://www.mediawiki.org/xml/export-0.10.xsd) von Wikipedia herunter und entpacken Sie sie. Das XML-Instanzdokument (Schema: <https://www.mediawiki.org/xml/export-0.10.xsd>) enthält alle Plattdeutschen Wikipedia-Artikel.

Schreiben Sie für jede der folgenden Fragen ein SAX-Programm. Geben Sie in ihrer Abgabe jeweils die Antwort an und fügen Sie den Code bei, der die Antwort liefert.

- (a) Wie viele `page`-Elemente sind im Wikipedia-Dump enthalten?

- (b) Welche Revisions-ID hat der Artikel mit Titel „Erfurt“ im Wikipedia-Dump?
- (c) Wie viele verschiedene angemeldete Autoren hat der Wikipedia-Dump?
Hinweis: Verwenden Sie die Klasse `java.util.HashSet`.

Aufgabe 6 : DOM/SAX (1+1 Punkte)

- (a) Erklären Sie den Unterschied der in der Vorlesung vorgestellten Parser-Prinzipien DOM und SAX und XML Data Binding (z. B. JAXB)
- (b) Nennen Sie je einen Vorteil und einen Nachteil der Parser-Prinzipien.

Aufgabe 7 : Programmiermodul: Java-API P (1+2+1+3+0+0 Punkte)

Schreiben Sie eine Java-Klasse, um einen XML-Feed mittels JAXB als Objekt zu repräsentieren.

- (a) Generieren Sie Java Klassen zu Ihrem XML-Schema (falls `feed.xml` und `entry.xml` dazu valide waren; alternativ die Musterlösung `atom.xsd` aus dem Repo dieses Belegs) mit Hilfe von `xjc`, welches Teil der Binärdistribution ist dies sie [hier](#) herunterladen können.

Beispiel Unix: `jaxb-ri/bin/xjc.sh -p jaxb atom.xsd`

(Für Windows entsprechend `xjc.bat` verwenden.

Wichtig: Kompilieren Sie nach der Generierung immer die erzeugte `package-info.java`!

Beispiel: `javac -classpath "jaxb-ri/mod/*" jaxb/package-info.java`

Binden Sie beim Kompilieren ihres Projektes stets die JAXB Referenz-Implementierung mittels `-classpath "jaxb-ri/mod/*"` ein!

- (b) Erstellen Sie in der `FeedModel.java` aus dem Repository einen Konstruktor, der das XML-Instanzdokument lädt.
 - Unmarshallen Sie `feed.xml` und setzen Sie den Feed als Member von `FeedModel`.
 - Validieren Sie den Feed. Werfen Sie bei Invalidität eine `RuntimeException`.¹
- (c) Erweitern Sie die `FeedModel.java` um eine Methode, die einen Eintrag per ID sucht und zurückgibt. Wurde kein Eintrag mit der ID gefunden, soll `null` zurückgegeben werden.
- (d) Erweitern Sie die `FeedModel.java` um eine Methode, die einen Eintrag hinzufügt.
 - Parameter der Methode sollen die Strings `title`, `url`, `summary` und `author` sein.
 - Generieren Sie die ID aus dem `hashCode()` der URL.
 - Nutzen Sie die `getXMLGregorianCalendarNow`-Methode für den Zeitpunkt der letzten Änderung.
 - Validieren Sie den Eintrag vor dem Hinzufügen. Werfen Sie bei Invalidität eine `RuntimeException`. Überprüfen Sie dabei mittels der Methode aus Aufgabe (c) ob die ID bereits verwendet wurde.
 - Fügen Sie den Eintrag an erster Stelle hinzu.
 - Aktualisieren Sie den Wert `updated` des Feeds nach der Validierung.

¹Für Java-Profis: Das ist nicht die richtige Exception für diesen Fall, aber sie erlaubt uns hier im Sinne der Übung auf die Einführung von `throws`-Auszeichnungen zu verzichten.

- Aktualisieren Sie zuletzt das XML-Instanzdokument mittels Marshalling.
Nicht erschrecken: Eine etwaige Stylesheet-Anweisung aus Aufgabe 4 wird dabei mit dem Wert der Konstante `XSL_SHEET` überschrieben.
 - Die Methode soll die ID des neuen Eintrags zurückgeben.
- (e) Erweitern Sie die `FeedModel.java` um eine Methode, die den Feed auf einen gegebenen `java.io.PrintWriter` schreibt.²
- (f) Erweitern Sie die `FeedModel.java` um eine Methode, die einen gegebenen Eintrag auf einen gegebenen `java.io.PrintWriter` schreibt.

Hinweise:

Testen Sie mit einer `main`-Methode das Lesen und Ändern des Feeds. Testen Sie auch invalide XML-Instanzdokumente und Einträge (mit Nullwerten).

Hinweise zur Abgabe

- Erstellen Sie *eine* PDF-Datei, in der Texte und Grafiken zu den Aufgaben enthalten sind.
- Quellcode (Java, JavaScript, PHP, HTML, CSS, XML, XSL, XSD, etc.) innerhalb des PDF-Dokuments wird nicht korrigiert. Quellcode jeder Aufgabe soll, wie im Tutorium vorgestellt, über GitLab verwaltet werden.
- Referenzieren Sie Ihre Quellcode-Dateien in dem PDF-Dokument, so dass sie einer Aufgabe eindeutig zuzuordnen sind. Verlinken sie sie entsprechendem Repositories.
- Abzugeben ist *eine* PDF-Datei die wie folgt benannt ist:
`<Nachname>-<MatrikelNr>-webtec-blatt<Übungsblattnummer>.pdf`.
- Gruppenabgaben mit bis zu drei Personen sind erlaubt; pro Gruppe genügt dann eine Abgabe, in der alle Gruppenmitglieder mit Namen, Matrikel, und E-Mail-Adresse verzeichnet sind
- Legen sie die PDF-Datei in das GitLab Repository in dem auch der Code ihrer Gruppe verwaltet wird.

²Diese Methode wird im nächsten Beleg vom Server genutzt werden um den Feed in eine HTTP Response zu schreiben.