

XML-Datenbanken Wintersemester 2008/2009 I. Aufgabenblatt

Hinweis: Zur Diskussion der Übungsaufgaben in der Übung sind vorweg formlose Präsentationsfolien zu erstellen. Diese müssen im PDF-Format per Email an kupfer@ifis.cs.tu-bs.de mit dem Betreff "XMLDatenbanken" sowie der Gruppennummer bis spätestens Sonntag vor der Übung abgegeben werden. Achten Sie bitte stets darauf Ihren Namen auf dem Titelblatt zu vermerken.

Aufgabe 1: Beschreiben Sie informell den Aufbau eines XML-Dokuments und welche Sprachmittel dafür zur Verfügung stehen. Benutzen Sie dazu Text, Stichworte oder Grafiken, aber keine Produktionsregeln.

Aufgabe 2: Implementieren Sie ein Java Programm, das eine beliebige Textdatei einliest und versucht, die verwendete Zeichenkodierung (character encoding) zu bestimmen. Es sollen 3 verschiedene Zeichenkodierungen unterschieden werden. Zur Implementierung können Sie beispielsweise die Klasse `java.io.FileInputStream` verwenden, die mit der Methode `read()` einzelne Bytes der Eingabedatei zurückliefert. Unter http://wwwalt.ifis.cs.tu-bs.de/html_d/wise2008/beispiel.zip kann als Hilfe eine Textdatei in 5 verschiedenen Kodierungen abgerufen werden.

- Finden Sie Unterschiede, z.B. indem Sie die Dateien Byte für Byte betrachten.
- Geben Sie an, welche Kodierungen Ihr Programm unterscheidet und welche Heuristiken Sie zur Erkennung verwenden.
- Gibt es Fälle in denen Ihr Programm die Kodierung falsch erkennt? Können Programme allgemein die korrekte Zeichenkodierung selbst erkennen?

Aufgabe 3: XHTML ist das Ergebnis der Anwendung des XML-Formats auf die bekannte Beschreibungssprache für Webseiten HTML. Diskutieren sie die Unterschiede anhand des folgenden HTML Beispiels. Der Quellcode funktioniert in Webbrowsern, obwohl er Fehler enthält.

```
<HTML>
  <Body>
    <frame noresize>
      <center>Weihnachtungswünsche!</center>
      <br>
      <UL>
        <Li>Bücher
        <Li>XML 文書の作成
      </UL>
    </frame>
  </Body>
</HTML>
```

- Welche Änderungen werden nötig um aus dem Beispiel ein XHTML Dokument zu erstellen?
- Skizzieren sie die Vor- und Nachteile der beiden Versionen.
- Mit der Umstellung auf XHTML ändert sich ebenso die Fehlertoleranz. Welche Auswirkungen hat dies?