



Praktikum zur Veranstaltung XML-Technologie: **Übung 01**

Erstellung eines ersten
XML-Dokuments,
Umgang mit dem Editor / XML-Modus
Umgang mit XML-Zeichenreferenzen, ext.
Entities & DTDs, XML-Deklaration



Organisatorisches



- Arbeitsverzeichnis:
`~/lv/xmltech/01/`
- Dateinamen:
`01-unicode.xml` # **abzugeben**
`01-unicode.dtd` # nur verwenden
- Werkzeuge:
`emacs, xemacs` # X-Emacs mit Syntax highl.
`nsgmls` # implizit via Emacs
`firefox` # für Anzeigetests



Die Aufgabe



- A: Codieren Sie die Tabelle aus der Unicode-Vorübung als XML-Datei!

– Erzeugen Sie eine XML-Datei wie folgt:

```
<Codetabelle>
  <Eintrag>
    <Zeichen>...</Zeichen>
    <Beschreibung>...</Beschreibung>
    <Unicode>U+...</Unicode>
    <ISO-Code TabNr="1">0x...</ISO-Code>
  </Eintrag>
  <!-- Weitere Eintraege ... -->

</Codetabelle>
```



Zeichenreferenzen in XML



- Hinweise:
 - Die ... sind Platzhalter und geeignet zu füllen.
 - Umschreiben Sie Ihre Zeichen im Element „Beschreibung“ (mit 7-bit ASCII-Text, ohne Sonderzeichen!)
 - Verwenden Sie im Element „Zeichen“ eine XML-Zeichenreferenz zur Codierung des Unicodes:

- Syntax für **Zeichenreferenzen**:

```
&#x...;          /* Hex-Ziffern ergaenzen */
```

- Beispiel:

```
&#xFC;          /* ü */
```



Hinweise zu Teil A



- Tipp: Schablone anlegen:
 - Legen Sie einen „leeren“ Eintrag an
 - Erstellen Sie neue Einträge durch
 - Kopieren des leeren Eintrags
 - Ausfüllen der Kopien
- „ISO-Code“ ist optional:
 - Einträge für Unicode-Zeichen ohne (gefundenen) ISO-Code legen Sie ganz ohne das Element „ISO-Code“ an.
- Syntaxprüfung:
 - Lesen Sie die Datei mittels **firefox**. Der eingebaute XML-Prozessor übernimmt die Syntaxprüfung.
 - Werden Syntaxfehler gemeldet? Ggf.: Korrigieren
 - Sind alle Sonderzeichen korrekt dargestellt?



Aufgaben, mit Arbeitsanweisungen



- B: Einbau einer externen DTD
 - Legen Sie **01-unicode.dtd** an durch Kopieren aus folgender Quelle:
</local10/werntges/lv/xmltech/01/01-unicode.dtd>
 - Fügen Sie nun per Editor folgende Zeile am Anfang von 01-unicode.xml ein:

```
<!DOCTYPE Codetabelle SYSTEM  
  "01-unicode.dtd" [  
]>
```
 - Sichern und schließen Sie die XML-Datei, laden Sie die Datei neu mit **Emacs**. Nun sollte die DTD beachtet werden!
 - Prüfen Sie die Datei mit **firefox** auf Syntaxfehler.



Aufgaben, mit Arbeitsanweisungen



- C: Der XML-Modus des Emacs
 - Fügen Sie per Editor einen neuen Tabelleneintrag ein
 - Gewünschtes Zeichen: Kleines griechische „delta“. Ermitteln Sie zunächst Unicode und ISO-Code/TabNr dazu.
 - Positionieren Sie nun den Cursor an die richtige Stelle.
 - Verwenden Sie dann Menüpunkt **Markup/Insert Element**
 - Wählen Sie das gewünschte Element „Eintrag“ aus.
 - Ergänzen Sie den eingefügten Text um Nutzdaten.
 - Ergänzen Sie analog das Element „ISO-Code“ im neuen Eintrag.
 - Validieren Sie nun Ihre Datei über **SGML/Validate!**

Es erscheint unten ein Fenster mit Warnungen und evtl. Fehlermeldungen. Ignorieren Sie die „üblichen“ Warnungen (vgl. mündliche Erläuterungen), beseitigen Sie gemeldete Fehler.

Tipp: Das untere Fenster lässt sich ausblenden mit CTRL-X 0 (Null), wenn der Cursor in diesem Fenster steht (aktives Fenster).



Aufgaben, mit Arbeitsanweisungen



- D: Einbau der XML-Deklaration; Zeichensätze.
 - Fügen Sie folgende Zeile am Anfang von 01-unicode.xml ein:

```
<?xml version="1.0"?>
```
 - Validieren Sie die Datei mit Emacs und firefox.
 - Geben Sie nun passende Umlaute per Tastatur „direkt“ in Element „Beschreibung“ zu den Umlauten ä und Ä ein.
 - Validieren Sie erneut. Ergebnis?
 - Ergänzen Sie nun die *encoding declaration*:

```
<?xml version="1.0" encoding="UTF-8"?>
```
 - Validieren Sie erneut. Begründen Sie Ihre Beobachtung!



Aufgaben, mit Arbeitsanweisungen



- D: Einbau der XML-Deklaration; Zeichensätze.
 - Ermitteln Sie nun einen geeigneten Zeichensatz und tragen Sie diesen ein:

```
<?xml version="1.0"
  encoding="(Ihr-Zeichensatzname)"?>
```
 - Validieren Sie erneut und passen Sie ggf. den Zeichensatznamen an, bis die Validierung gelingt und firefox die Zeichen korrekt anzeigt.
 - Hinweis: Nutzen Sie das Vorlesungsmaterial bzw. die XML-Spezifikation zur Ermittlung eines geeigneten Zeichensatznamens.



Aufgaben, mit Arbeitsanweisungen



- E: Einbau eines internen *general entity*.
 - Erzeugen Sie zu Beginn des internen Subsets der DTD folgende Zeilen:

```
<!ENTITY ae "a-Umlaut">
<!ENTITY Ae "A-Umlaut">
```
 - Wenden Sie die *entity*-Referenzen im Element „Beschreibung“ von „ä“ und „Ä“ an, etwa so:

```
<Beschreibung>ae, &ae;</Beschreibung>
```
 - Validieren Sie mit Emacs/nsgmls. Ok?
 - Prüfen Sie nun mit firefox. Ergebnis?



Aufgaben, mit Arbeitsanweisungen



- F: Nutzung vordefinierter externer *parameter entities*.
 - Tragen Sie in den internen Subset der DTD folgende 2 Zeilen ein:

```
<!ENTITY % iso-grk1 SYSTEM
"/usr/share/sgml/entities/xml-iso-entities-
8879.1986/ISOgrk1.ent">
%iso-grk1;
```
 - Schauen Sie in die referenzierte Datei hinein mit einem Tool Ihrer Wahl und suchen Sie sich zwei *entities* aus, etwa „dgr“ (aber nicht gerade dieses!). Laden Sie dann die XML-Datei neu.
 - Verwenden Sie Ihre beiden *entities* an geeigneten Stellen, d.h. setzen Sie die entsprechenden *entity*-Referenzen mittels Emacs-Menüpunkt „Markup/Insert Entity“ und validieren Sie mit Emacs/nsgmls.
 - Prüfen Sie schließlich mit firefox. Funktioniert's ebenfalls?



Anmerkungen



- Die Verwendung von *entities* ist weit verbreitet und sollte bis zur sicheren Beherrschung geübt werden.
- Manche XML-Prozessoren halten sich nicht immer an den Standard oder sind schlicht noch fehlerhaft. Im Zweifelsfall sollte man daher mit mehreren Werkzeugen prüfen.
- Die ISO-*Entities* sind ein weit verbreiteter Quasistandard.
 - Wir werden noch eine bessere Methode zu ihrer Adressierung kennen lernen als eine direkte lokale Pfadangabe:
PUBLIC entities und *catalogs!*



- Datei vervollständigen!
 - Sollten Sie noch nicht alle Unicode-Einträge gefunden haben, holen Sie dies im Rahmen des freien Übens nach.
- Datei aufbewahren!
 - Die Datei ist Grundlage weiterer Übungen
 - Sie wird noch weiterentwickelt
 - Ihre endgültige Version ist „abzugeben“, daher:
- Datei nicht von anderen kopieren oder zum Kopieren freigeben