

Grundlagen Des Web

Unterschiede geläufiger Standards für Zeichensätze im Web

- ASCII (8 Bit) : US-amerikanische Sprache, somit keine Sonderzeichen, es werden 128 Zeichenplätze verwendet.
- ISO 8859-1, „Latin 1“ (8 Bit) : West/Nordeuropäische Sprache, somit wird die komplette Kapazität verwendet.
- UNICODE (32 Bit) : Alle Kulturen (auch ältere)
- UTF-8 (8 Bit) : variable 1 bis 4 Byte Kodierung

Die einzelnen Zeichensätze unterscheiden sich u.a. durch ihre Codierung, Größe und den Indizes.

Zeichen		A	ü	π
Zeichensatz	ASCII	65	n/a	n/a
	ISO 8859-1	65	252	n/a
	Unicode	65	252	960
Zeichenkodierung	ASCII	41	n/a	n/a
	ISO 8859-1	41	FC	n/a
	UTF-8	41	C3 BC	CF 80

Betrachten wir allein die Buchstaben „A“, „ü“ und „π“. Das „A“ ist in allen Zeichensätzen vertreten und wird aufgrund der Kompatibilität von ISO 8859-1 und UTF-8 zu ASCII auch durch dieselbe Bit-Sequenz kodiert. Dort gibt es keine Probleme. Das „ü“ ist nun im ASCII-

Zeichensatz nicht mehr enthalten und kann somit auch nicht kodiert werden. Hier ist auch schon zu sehen, dass obwohl das Zeichen in den Zeichensätzen ISO 8859-1 und Unicode an derselben Position steht, es doch unterschiedlich kodiert wird. Das letzte Zeichen, das „π“, ist hingegen nur noch im Unicode-Zeichensatz enthalten und kann daher auch nur durch UTF-8 kodiert werden. Also ist es wichtig, die verwendete Zeichenkodierung immer anzugeben. Im HTTP ist dafür das Header-Feld „Content-Type“ vorgesehen, das neben dem Inhaltstyp (etwa „text/html“ für ein HTML-Dokument) auch einen Parameter für die Zeichenkodierung (verwirrenderweise als „charset“ bezeichnet) besitzt.

Das Konzept von Auszeichnungssprachen (Markup Languages)

- Eine Auszeichnungssprache ist eine maschinenlesbare Sprache für die Gliederung und Formatierung von Texten und anderen Daten. z.B.: HTML, XHTML, SGML)
- Eine Auszeichnungssprache ist ein System für die Annotation von Texten, sodass Annotation und Text syntaktisch voneinander unterscheidbar sind. Mit Annotationen sind Mittel zur Strukturierung von Problemquellen gemeint, die dafür sorgen, dass die damit verbundenen Dateien automatisch erzeugt werden. Die dazu notwendigen Informationen aus dem Quelltext werden Metadaten genannt. Hierbei gibt es drei verschiedene Auszeichnungssprachen:

-> Präsentation-Markup : Codes oder Tastenkombinationen, mit denen die Darstellung der Zeichen bestimmt werden kann (fett, kursiv, ...) > **WYSIWYG** (What You See Is What You Get)

-> Prozedurale Markup: Die Einbindung von zusammengefassten Folgen von Anweisungen o. Deklarationen (-> Makros) in einem Text, sodass die Texte z.B. auf eine vorgegebene Weise formatiert werden. „Wie soll ein Text aussehen bzw. Verarbeitet werden“

-> Deskriptiver Markup: Eigenschaften eines Dokuments oder Teile davon werden in einem separaten Dokument beschrieben, die die Darstellung, Struktur und Bedeutung von ebendiesem festlegt, wie es im Kopfteil (Head Element) einer Quelldatei der Fall ist. „Welche Bedeutung hat ein Text“

Marshalling

Marshalling beschreibt den Prozess bei dem strukturierte Daten und primitive Datentypen in eine externe zum Umwandeln von strukturierten Datenrepräsentation gebracht werden.

Client und Server kommunizieren mittels eines strikt definierten Übertragungsformats, daher müssen konkrete Anfragen, die der Server erzeugt, vom jeweils systemspezifischen Format in dieses Übertragungsformat überführt werden.

Unmarshalling beschreibt den Vorgang des Überführen des gemeinsamen des Übertragungsformats in das jeweils systemspezifische Format.

Rolle und Bedeutung von Schemasprachen für den Datenaustausch im Web

Schemasprachen sind Sprachen, mit denen Schemata verfasst werden können. Beispiele solcher Sprachen sind DTD (Document Type Definition), XML-Schema oder RelaxNG. Mithilfe der mit den Schemasprachen definierten Schemata können Instanzdokumente verfasst, und bezüglich ihrer Validität geprüft werden. Beim Datenaustausch über ein Netzwerk können die Inhalte von einem Instanzdokument dann mittels Schema Prozessor geprüft werden. Schemasprachen erlauben es verschiedenste für Nutzungskontexte Auszeichnungsregeln zu verfassen, die die Rohdaten beschreiben. Damit kann beispielsweise eine Website generiert werden (XSLT) oder Abfragen auf den Datenbestand erfolgen (XPath, XQuery).

Unterschied zwischen "Wohlgeformtheit" und "Gültigkeit" von XML Dokumenten

Der Unterschied liegt darin, dass ein XML Dokument als wohlgeformt bezeichnet wird, wenn es einige grundlegende syntaktische Regeln befolgt. So darf es nur ein Root-Element enthalten, nicht leere Elemente müssen einen öffnenden und einen schließenden Tag haben, alle im Dokument verwendeten Zeichen müssen im angegebenen Zeichensatz erhalten sein.

Was ist ein "Element" in einem XML Dokument?

XML (Extensible Markup Language) ist ein Schema zur Validieren von Inhalten und kann Instanzen auf Richtigkeit in der Reihenfolge überprüfen. Es ist eine Metasprache für hierarchisch strukturierte Daten in Form von Textdateien.

Namespaces

Namespaces werden benutzt um bei Verwendung mehrerer XML Sprachen in einem Dokument Mehrdeutigkeit zu vermeiden. Es wird somit sichergestellt, dass die verwendeten Vokabulare eindeutig identifiziert werden können.

Erklären Sie den folgenden regulären Ausdruck **recipe (title,date,ingredient*,preparation,comment?, nutrition,related*)!**

Ein Rezept hat einen <Titel> bzw. Namen
Ein <Datum>, beliebig viele <Zutaten> (Name, Anzahl, Nährwert)
<Vorbereitung> die man treffen muss.
<Kommentarfeld> mit Text, <Nährstoffe> (Kalorien, Fett,...)
Related* (verwandte Rezepte)

Der Stern bedeutet mehrfaches auflisten.

```
<driveassistant> <address> <start> <street>W Mahoning
St</street><city>Punxsutawney</city><zipcode>PA
15767</zipcode><country>US</country></start><end> <street>1600 Pennsylvania Ave
NW</street><city>Washington</city><zipcode>DC
20500</zipcode><country>US</country></end></address><total> <totaltime>
<hour>4</hour><minute>33</minute></totaltime><distance>
<number>246.68</number><standard>miles</standard></distance></total><route>
<point>
<number>1</number><maneuver>Start out
going</maneuver><direction>NORTHWEST</direction><address>PA-36 / N MAIN ST</
address><distance>
<number>< 0.1</number><standard>miles</standard></distance></point> <point>
<number>2</number><maneuver>Turn LEFT</maneuver><direction>onto LINCOLN
AVE</direction><address>PA-36 / N MAIN ST</address><distance>
<number>0.2</number><standard>miles</standard></distance></point> ...<point>
<number>36</number><maneuver>End</maneuver><direction>at</
direction><address>1600 Pennsylvania
Ave Nw, Washington, DC US</address><distance>
<number>0</number><standard>miles</standard></distance></point> </route></
driveassistant>
```

Schreiben Sie ein gültiges Instanzdokument für das Schema movies.xsd, das jedes deklarierte Element mindestens einmal verwendet!

```
<movies> <movie> <title>Harry Potter and the Philosopher's Stone</title><length>159</
length><year>2001</year><cast>
<person> <name>Daniel Radcliffe</name><birth> 23. July
1989</birth><id>2366325</id></person></cast></movie></movies>
```