

Die erweiterbare Markierungssprache XML

1. Einleitung

Eine Markierungssprache (markup language) dient dazu, Textdateien mit zusätzlichen Informationen anzureichern. Die verbreitete Markierungssprache HTML (hypertext markup language) erlaubt es z.B., bestimmte Teile einer Textdatei als Überschriften und andere Teile als normale Absätze zu kennzeichnen, z.B. so:

```
1 <H1>1. Über den Umgang mit Drachen</H1>
2
3 <P ALIGN="left">
4   Allgemein unterscheidet man zwischen Drachen <B>erster</B> Ordnung
5   und Drachen <B>zweiter</B> Ordnung.
6 </P>
7
8 <H2>1.1. Drachen erster Ordnung</H2>
9 <P ALIGN="left">
10  Drachen <B>erster</B> Ordnung gibt es nicht. Deshalb sind sie in
11  der Natur sehr schwer zu finden und werden von vielen Liebhabern sehr
12  geschätzt.
13 </P>
14
15 <H2>1.2. Drachen zweiter Ordnung</H2>
16 <P ALIGN="left">
17  Drachen <B>zweiter</B> Ordnung gibt es auch nicht, aber aus ganz
18  anderen Gründen als Drachen erster Ordnung.
19 </P>
```

Die Markierungen (tags) in den spitzen Klammern drücken unter anderem aus, dass der Text "1. Über den Umgang mit Drachen" in Zeile 1 ein Überschrift der Stufe 1 ist (H1 wie header 1) und dass die Zeilen 4 und 5 einen Abschnitt (P wie paragraph) bilden, der linksbündig ausgerichtet werden soll (und nicht rechtsbündig oder zentriert).

Die Markierungssprache HTML hat (neben dem Vorteil ihrer weiten Verbreitung) folgenden Nachteil: Sie erledigt zwei Aufgaben auf einmal, die man häufig besser getrennt erledigen sollte. Mit HTML-Markierungen beschreibt man zum einen, aus welchen Teilen ein Text besteht (z.B. aus Überschriften verschiedener Stufe, normalen Abschnitten etc.). Gleichzeitig legt man damit fest, wie diese Teile z.B. auf einem Bildschirm dargestellt werden sollen, z.B. linksbündig oder fett etc..

HTML-Markierungen haben grosse Ähnlichkeit mit Klammern. Z.B. entspricht die Markierung <H1> einer öffnenden Klammer und </H1> ist die zugehörige schliessende Klammer. Leider ist es nach den Regeln der Sprache HTML erlaubt, bestimmte schliessende Markierungen wegzulassen. Eine syntaktische Prüfung von HTML-Dateien wird dadurch so stark erschwert, dass man in der Praxis häufig auf sie verzichtet. Ergebnis: Viele HTML-Dateien enthalten syntaktische Fehler, die zu falschen Darstellungen führen, aber nicht automatisch lokalisiert werden.

2. Die erweiterbare Markierungssprache XML

Die Sprache XML (extensible markup language) ist deutlich mächtiger als HTML. Sie erlaubt es, die abstrakte Struktur von Texten, die konkreten Texte selbst und die Regeln für ihre Darstellung z.B. auf einem Bildschirm getrennt voneinander zu erstellen und zu verwalten. Davon versprechen sich viele Anwender erhebliche Vorteile.

Um einen mit XML strukturierten Text zu erstellen, kann man z.B. so vorgehen:

1. Man erstellt eine Datentyp-Definition (eine .dtd-Datei). Darin legt man z.B. fest, dass die später


```

145     <Fon>          030-4504-2265 </Fon>
146     <Fon>          0172-3456-7890 </Fon>
147     </Adresse>
148
149     <Adresse>
150         <Nachname> Seidelsdorf </Nachname>
151         <Vorname>  Wilhelmine </Vorname>
152         <Fon>      030-883-7689 </Fon>
153         <Fon>      040-674-9215 </Fon>
154     </Adresse>
155
156     <Adresse>
157         <Nachname> Trossman </Nachname>
158         <Vorname>  Fridolin </Vorname>
159         <Fon>      030-765-1243 </Fon>
160         <Fon>      040-394-5643 </Fon>
161     </Adresse>
162
163 </Unterliste>
164 </Adressen-Liste>

```

Wenn man sich diese Datei AdrList01.xml mit dem IE anzeigen lässt, erscheint ungefähr folgendes auf dem Bildschirm (für den Fall, dass Sie eine schwarz-weiß-Kopie dieses Papiers lesen, steht rechts neben jeder Zeile, in welcher Farbe sie auf dem Bildschirm erscheint: schwarz, rot, blau oder grün):

Dieser Text wird zuerst ausgegeben!

Adressliste von Ottokar Meyerbeer

Persoenliche Bekannte und Freunde

Heinz-Dieter Schmidt 030-4504-2265

Gertrude Mueller 030-883-7689 040-674-9215 0172-3456-7890

Berufliche Kontakte

Kunigunde Aschenhus 030-4504-2265 0172-3456-7890

Wilhelmine Seidelsdorf 030-883-7689 040-674-9215

Fridolin Trossman 030-765-1243 040-394-5643

schwarz

rot

blau

grün

grün

blau

grün

grün

grün

4. Programme zum Bearbeiten von XML-Dateien

Im Prinzip kann man .dtd-Dateien, .xsl-Dateien und .xml-Dateien mit jedem gewöhnlichen ASCII-Editor erstellen und verändern. Allerdings ist es ziemlich mühsam, die vielen Markierungen alle von Hand zu tippen.

Es gibt heute (Ende 2007) zahlreiche XML-Editoren. Alle erleichtern einem (mehr oder weniger) das Erstellen von XML-Dateien. Einige unterstützen auch das Erstellen und Verändern von DTD-, von XSL-, von XSD- (XML-Schema Dateien) und weiteren Dateien.

Umsonst zu haben sind unter andern die XML-Editoren Merlot (siehe <http://www.merlotxml.org/>) und der XML Notepad von Microsoft. Weitere interessante Internetadressen und Hinweise findet man unter der Adresse <http://www.xmlguru.de>.