

Ein erster Eindruck

XML, eXtensible Markup Language, bezeichnet ein Standard zur Auszeichnung von Daten in Textform. Während diese Möglichkeit der Textauszeichnung bei weitem nicht von allen Computernutzern beherrscht wird, ist die auch für XML typische Klammerschreibweise wahrscheinlich zumindest aus HTML-Dokumenten bekannt. Bei beiden Kodierungsstandards handelt es sich um Auszeichnungssprachen (Markup Languages), die dazu dienen, einen Text im Hinblick auf Format, Struktur und Merkmale zu beschreiben.



Ein einfaches HTML-Dokument könnte beispielsweise folgendermaßen aussehen:

```
<!DOCTYPE html>
<html>
  <head>
    <title> Ein Satz </title>
  </head>
  <body>
    <h1>Ein Satz</h1>
    <p>Ich lerne eine Auszeichnungssprache</p>
  </body>
</html>
```

Die HTML-Struktur kodiert einen Text nach Regeln, die es Internetbrowsern möglich machen, Inhalte in einem bestimmten Format zu präsentieren. So markieren die Tags `<h1>` und `</h1>` Anfang und Ende einer Überschrift, `<p>` und `</p>` Anfang und Ende eines Paragraphen, und `<body>` und `</body>` den Anfang und Ende der Webseite. Eine Überschrift, die mit `<h1>` ausgezeichnet wurde, stellt eine übergeordnete Überschrift dar und ist in der Regel größer als eine in `<h2>`, welche wiederum einer `<h3>` übergeordnet ist, etc.. Browser und Texteditoren nutzen einen solchen Code, um puren Text in eine Form zu bringen und ihm Funktionen (z.B. Links) zuzuweisen.

Die XML-Kodierung befolgt ähnliche Regeln wie HTML. Auch hier wird etwas mit `<...>` begonnen und mit `</...>` beendet. Damit ein HTML-Dokument zu einem XML-Dokument wird, reicht es in unserem Beispiel bereits, die erste Zeile zu verändern:

```
<?xml version="1.0" encoding="ISO-8859-1" standalone="yes" ?>
<html>
  <head>
    <title> Ein Satz </title>
  </head>
  <body>
    <h1>Ein Satz</h1>
    <p>Ich lerne eine Auszeichnungssprache</p>
  </body>
</html>
```

Im Unterschied zu HTML weist XML den unterschiedlichen Tags jedoch keine „Bedeutung“ zu. Tags wie `<h1>` bedeuten hier eben nicht, dass der nachfolgende Text das Format einer Überschrift annimmt, sondern zeichnen ihn lediglich als `<h1>` aus, ohne dass ein Standard zur Interpretation von `<h1>` vorliegt. `<h1>` und `</h1>` ließen sich dementsprechend zu `<h2>` und `</h2>` oder `<p>` und `</p>` u.s.w. ändern, ohne dass sich das Format des Textes verändert. Ebenso wenig bedeutet `<html>`, dass es sich bei dem nachfolgenden Text um eine HTML-Kodierung handelt. Es ließen sich alternativ (unter Einhaltung bestimmter Regeln) auch beliebige Tagnamen wie `<ueberschrift>`, `<wichtig>` oder `<textdatenromanistik>` kreieren. Der Satz „Ich lerne eine Auszeichnungssprache“ kann in XML genauso gut auch folgendermaßen ausgezeichnet werden:

```
<?xml version="1.0" encoding="ISO-8859-1" standalone="yes" ?>
<beispiel>
  <titel> Ein Satz </titel>
  <satz>
    <subjekt>
      <pronomen> Ich </pronomen>
    </subjekt>
    <prädikat>
      <verb> lerne </verb>
    </prädikat>
    <objekt>
      <artikel>eine</artikel>
      <nomen>Auszeichnungssprache</nomen>
    </objekt>
  </satz>
</beispiel>
```

Da die Tagnamen frei wählbar sind, spielt die inhaltliche "Korrektheit" der Tags keine Rolle. Eine Abänderung von `<nomen> Auszeichnungssprache </nomen>` zu `<verb> Auszeichnungssprache </verb>` hätte für die Wohlgeformtheit des XML-Dokuments keine Bedeutung.

