

XML und XML Schema

Standardisierte Darstellung strukturierter Information

Vorher: proprietäres Datenformat

„Warum 4be brüten EVö gel5 EierF ?“ für „Warum *bebrüten* **Vögel Eier**?“

Problem

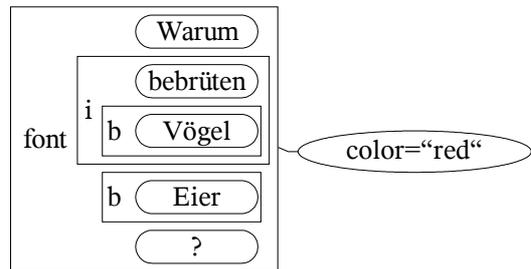
- Strukturierte Information soll nicht nur Computer-lesbar, sondern auch Menschen-lesbar dargestellt werden.
- Strukturierte Information soll möglichst deklarativ, nicht prozedural dargestellt werden.
- Abwärtskompatibilität: Die Erweiterung des Datenformats bestehende Anwendungen nicht stören.
- Aufwärtskompatibilität: Das Datenformat soll überhaupt erweiterbar sein.

Nachher: Markup-Datenformat

„Warum <i>bebrüten Vögel </i> Eier?“

eXtensible Markup Language

- Standard-Sprache für Markup-Dokumente.
- Haupt-Baustein der Sprache ist der Knoten
 - Text-Knoten bestehen aus einer Zeichenkette
 - Element-Knoten
 - bestehen aus einer Liste von Kind-Knoten
 - haben eine Menge von Attributen
 - Jedes Attribut besteht aus einer Zuordnung von einem Attribut-Namen zu einem Attribut-Wert
- Da Element-Knoten wieder Element-Knoten enthalten können, lassen sich Elemente beliebig verschachteln.
 - Somit lassen sich Informationen, die lokal zu einem Element sind, als Kind innerhalb des Elements platzieren.
- XML-Dokumente bestehen ausschließlich aus Text (Der Zeichensatz ist effektiv Unicode.)
 - Dies lädt dazu ein, Elemente und Attribute sinnvoll und damit für Menschen lesbar zu benennen.
- Haben Elemente und Attribute mit dem selben Namen immer die selbe Bedeutung, so belegen die Elemente das, was nicht fest vorgegeben ist, mit einer Bedeutung.
 - Wenn also „<i>...</i>“ bedeutet, dass der eingeschlossene Text kursiv dargestellt werden soll, dann heißt das, dass in „Warum <i>bebrüten Vögel </i> Eier?“ die Zeichenkette „bebrüten Vögel“ kursiv dargestellt werden soll.
- Damit lässt sich ein XML-Dokument auffassen wie ein ausgefülltes Formular. XML erlaubt also eine formalisierte Beschreibung von Sachverhalten.
- Man kann also versuchen, mit XML beliebige Objekte beliebiger Klassen zu beschreiben. In der Praxis schafft man das nur, wenn die Klassen (und ihre Zuordnung zu XML) genau spezifiziert sind.



XML Namensräume

Angenommen, es gibt eine Beziehung zwischen den Namen von Elementen und Attributen und ihrer Bedeutung, dann ist noch lange nicht klar, ob diese Beziehung eindeutig ist.

Aus diesem Grund schreibt man im XML-Dokument vor jeden Element- und Attribut-Namen noch ein Prefix und ordnet diesem Prefix eine Bedeutung zu. Diese Bedeutung ist zwar für die XML verarbeitende Software in der Regel opaque (also nicht erschließbar), jedoch reicht dies aus, um eindeutig zu sagen, welche von verschiedenen möglichen Bedeutungen gemeint ist.

Beispiel

```
<matrix xmlns:mathml='http://www.w3.org/1998/Math/MathML'>
  <mathml:matrixrow>
    <mathml:cn>1.5</mathml:cn>
  </mathml:matrixrow>
</matrix>
```

XML Schema

Haben Elemente und Attribute bestimmte Bedeutungen, dann heißt das so gut wie immer, dass nicht alle theoretisch möglichen Kombinationen von Elementen, Attributen und Text bezüglich dieser Bedeutungen sinnvoll sind. Deswegen lassen sich verbal Einschränkungen formulieren, denen ein XML-Dokument genügen muss, um entsprechend einer Bedeutung überhaupt als potentiell sinnvoll anerkannt zu werden.

Diese verbalen Einschränkungen lassen sich teilweise formalisieren. Eine relativ fortgeschrittene formale Sprache für solche Einschränkungen ist XML Schema. Sie erlaubt

- die Einschränkung von Text (Text-Knoten oder Attribut-Wert) anhand von
 - regulären Ausdrücken
 - Aufzählungen aller möglichen Zeichenketten
 - Intervallen
 - Vereinigungen von Sprachen
 - Listen von Wörtern einer Sprache
- die Einschränkung der Struktur von Knoten (Element oder Text-Knoten) anhand
 - der Häufigkeit von Vorkommen bestimmter „Strukturen von Kind-Knoten“ (Klassen)
 - der Reihenfolge von Vorkommen bestimmter „Strukturen von Kind-Knoten“ (Klassen)
 - der Auswahlpflicht zwischen bestimmten „Strukturen von Kind-Knoten“ (Klassen)
- die Vererbung
 - durch Erweiterung:
 - durch Anhängung von bestimmten Strukturen von Kind-Knoten an bestehende Klassen
 - durch Restriktion
 - durch Neuformulierung der Klasse, aber mit schärferen Nebenbedingungen bezüglich einzelner Elemente oder Attribute als bei der Oberklasse
 - Restriktion über Invarianten („Ein Quadrat ist ein Rechteck mit gleicher Kantenlänge.“) ist nicht möglich.

Fazit

- XML Schema ist nur für die Beschreibung von Restriktionen von Datentypen ausgelegt, Restriktionen für beliebige Zusammenhänge lassen sich nicht formulieren.
- Damit reicht XML Schema als Grundlage für automatisierte Schlussfolgerungen nicht aus. XML-Schema-konforme Dokumente sind aber sinnvoll für den automatisierten und standardisierten Datenaustausch.