

## **Kleines XML-Glossar**

aus: Helmut Vonhoegen Einstieg in XML, Galileo Press

### ***API – Application Programming Interface***

Öffentlich verfügbare Programmierschnittstelle, die bei der Anwendungsentwicklung genutzt werden kann. DOM ist zum Beispiel eine solche Schnittstelle, die Anwendungen unterschiedlicher Sprachen und Plattformen einen gezielten Zugriff auf Teile eines XML-Dokuments bereitstellt.

### ***Attribut***

Eine Informationseinheit, die einem Element in einem XML-Dokument beigefügt werden kann. Attribute erscheinen im Start-Tag hinter dem Elementnamen in Form von Name-Wert-Paaren, getrennt durch ein Gleichheitszeichen, wobei der Wert in Anführungszeichen gesetzt wird. Mehrere, jeweils einmalige Attribute können getrennt durch Leerraum aufeinander folgen. Alle Attributwerte sind Zeichenketten.

### ***B2B – Business-to-Business***

Bezeichnung für die computergestützte Kommunikation zwischen Unternehmen.

### ***B2C – Business-to-Consumer***

Bezeichnung für die computergestützte Kommunikation zwischen Unternehmen und ihren Kunden, etwa beim E-Commerce über das Web.

### ***Callback***

Programmiertechnik, bei der ein Prozess einen anderen auslöst. Der zweite Prozess ruft später den ersten als Ergebnis eines bestimmten Werts, einer Aktion oder eines Ereignisses wieder auf.

### ***CDATA – Character Data***

Ein Block von Zeichen in einem XML-Dokument, der nicht geparkt werden soll. CDATA-Bereiche werden in die beiden Trennzeichenfolgen `<!CDATA[` und  `]]>` eingepackt. Sie werden verwendet, um beispielsweise Texte in das Dokument mit aufzunehmen, die reservierte Zeichen wie `<`, `>` oder `&` enthalten.

### ***Container-Element***

Ein Element in einem XML-Dokument, das selbst wiederum andere Elemente oder Elementgruppen oder Mischungen aus Elementen und einfachen Zeichendaten enthält.

### **CSS – Cascading Style Sheets**

Stylesheets enthalten Formatierungsregeln für XML- oder HTML-Elemente. Durch die Zuordnung eines Stylesheets kann in einfacher Form die Präsentation der Inhalte eines XML-Dokuments gesteuert werden. Dabei können mehrere Regeln überlagert werden.

### **Dateninsel – data island**

Ein XML-Dokument, das in eine HTML-Seite eingefügt wurde. Dazu wird das Tag `<XML>` verwendet. Auf die Informationen dieser Dateninseln können Skripte angewendet werden. Die andere Möglichkeit ist die der Datenbindung, bei der HTML-Elemente gezielt mit XML-Elementen verknüpft werden.

### **Datentypen**

XML Schema erlaubt die Verwendung einer umfangreichen Hierarchie von Datentypen, mit denen festgelegt werden kann, welche Inhalte die Elemente eines XML-Dokuments und welche Werte die verwendeten Attribute annehmen dürfen. Dabei werden zahlreiche vorgegebene Datentypen angeboten, aus denen der Anwender eigene Datentypen ableiten kann.

### **Dokument**

Das Dokument ist in XML die Informationseinheit, die alle untergeordneten Informationseinheiten in sich einschließt. Dabei kann ein XML-Dokument physikalisch aus mehreren Dateien zusammengesetzt sein.

### **Dokumentelement**

Das Wurzelement eines XML-Dokuments, das alle anderen Elemente enthält. Muss als erstes Element im Dokument erscheinen. Das Dokumentelement steht für das Dokument insgesamt.

### **Dokument-Entität**

Startpunkt für einen XML-Parser. Diese Entität hat keinen Namen und kann auch nicht referenziert werden. In der Dokument-Entität können die XML-Deklaration und eine Dokumenttyp-Deklaration vorkommen.

### **Dokumentinstanz**

Ein XML-Dokument ist eine Dokumentinstanz eines Dokumentmodells, das mit Hilfe einer DTD oder von XML Schema definiert ist, wenn es diesem Inhaltsmodell entspricht, also ein gültiges, konkretes Beispiel des abstrakten Modells ist.

### ***Dokumentmodell***

Inhaltsmodell für ein XML-Dokument, das in Form einer DTD oder eines XML Schemas die Struktur, d.h. das Vokabular und die Grammatik einer bestimmten Klasse von konkreten Dokumenten festlegt.

### ***Dokumenttyp-Deklaration***

Festlegung der Struktur, der ein XML-Dokument zu entsprechen hat, um als gültiges Dokument von einem validierenden Parser eingestuft zu werden. Die Syntax ist `<!DOCTYPE Inhalt>`. Statt interner Dokumenttyp-Definitionen kann die Deklaration auch auf externe Dateien verweisen, die DTDs enthalten. Die Deklaration muss der XML-Deklaration folgen und vor dem ersten Element erscheinen.

### ***DOM – Document Object Model***

DOM definiert eine abstrakte Programmierschnittstelle, die unabhängig von einer bestimmten Plattform oder Sprache ist, um Programmen und Skripten einen wahlfreien Zugriff auf den Inhalt und die Struktur eines XML-Dokuments zu ermöglichen. Dabei wird jeweils ein Baum von Objekten im Speicher aufgebaut, die bestimmte Methoden des Zugriffs und der Veränderung bereitstellen.

### ***DTD – Dokumenttyp-Definition***

Definition der erlaubten Elemente und Attribute und ihrer Zusammensetzung für eine Klasse von Dokumenten. Die DTD legt damit die in der Dokumentinstanz zu verwendenden Tags fest und bestimmt, in welcher Reihenfolge sie auftreten und wie sie verschachtelt werden dürfen. Außerdem werden die bei einem Element erlaubten Attribute und die Art der Werte festgelegt. Eine genaue Datentypbestimmung ist allerdings erst durch XML Schema möglich geworden.

Liegt eine DTD vor, kann ein validierender Parser ein konkretes Dokument an der zugeordneten DTD prüfen. Die Verwendung von DTDs ist aber nicht vorgeschrieben. Fehlt eine DTD (oder ein Schema), wird ein XML-Dokument von einem Parser nur auf Wohlgeformtheit geprüft.

### ***Element***

Grundlegende Informationseinheit in einem XML-Dokument. Elemente bestehen aus einem den Typ des Elements benennenden Start-Tag, dem Elementinhalt und dem End-Tag. Das Start-Tag kann dabei auch eine unbestimmte Zahl von eindeutigen Attributen enthalten. Die gültigen Elemente können durch eine DTD oder ein XML Schema vorgeschrieben werden. Der Elementinhalt kann direkt aus Zeichendaten bestehen oder

wiederum aus untergeordneten Elementen oder aus einer Mischung aus Zeichendaten und Elementen.

### ***Entität – Entity***

Ein XML-Konstrukt, das als Platzhalter für unterschiedliche Inhalte verwendet werden kann. Der XML-Parser ersetzt bei der Verarbeitung des Dokuments die Platzhalter durch die zugeordneten Inhalte, das Dokument wird dann mit den geparsten Entitäten weiterverarbeitet. Eine Entität kann ein einzelnes Zeichen oder auch ganze Dateien beinhalten. Für jede Entität muss ein eindeutiger Name verwendet werden. Der Bezug auf eine Entität wird durch eine entsprechende Referenz hergestellt, ähnlich wie bei einer Makrodefinition für Textergänzung. In einer DTD können Entitäten deklariert werden.

### ***Entitätsreferenz – entity reference***

Mit einer Entitätsreferenz kann der Inhalt einer Entität in ein Dokument einbezogen werden. Die Referenz beginnt mit dem Zeichen &, dann folgt der Name der Entität und ein abschließendes Semikolon. Wenn der Prozessor die betreffende Zeile verarbeitet, wird der Verweis durch den darin angegebenen Inhalt ersetzt.

### ***Facette***

Ein bestimmter Aspekt des Werteraums eines Datentyps. Unterschieden wird dabei zwischen fundamentalen und einschränkenden Facetten. Dass ein Wert numerisch ist, gilt zum Beispiel als fundamentaler Aspekt, dass er Werte zwischen 1 und 100 annehmen darf, ist eine Einschränkung dieses Aspekts.

### ***Fragmentbezeichner***

Erweiterung eines URL, die einen direkten Zugriff auf ein benanntes Element auf einer Webseite erlaubt.

### ***Komplexer Datentyp – complex data type***

Ein Element in einem XML Schema, das andere Elemente oder Attribute enthält. Komplexe Datentypen können anonym oder mit einem Namen definiert werden. Benannte komplexe Datentypen können als Basistypen für abgeleitete Datentypen verwendet werden.

### ***Knoten***

Die Objekte, aus denen DOM seine Baumstruktur aufbaut, um die logische Struktur eines XML-Dokuments zu präsentieren, werden Knoten genannt. XPath verwendet eine ähnliche Baumrepräsentation des XML-Dokuments.

**Kontextknoten – reference node**

Der Kontextknoten ist bei einer XPath-Lokalisierung innerhalb eines Knotenbaums der Knoten, von dem aus die Operation gestartet wird.

**HTML – Hypertext Markup Language**

Markup-Sprache, die hauptsächlich für die Erstellung von Webseiten verwendet wird. Die aktuelle Version HTML 4.0 ist durch drei DTDs als SMGL-Abkömmling definiert.

**HTTP – Hypertext Transport Protocol**

Das Protokoll, das für den Datenaustausch im World Wide Web verwendet wird. Es legt fest, auf welche Weise ein Client – ein Browser – Daten von einem Webserver anfordert und wie der Webserver auf diese Anfrage antwortet und die geforderten Webseiten zur Verfügung stellt. Das Protokoll realisiert ein einfaches Nachrichtenmodell, das aus einer Klasse von Anforderungsnachrichten und einer Klasse von Antwortnachrichten besteht. Eine Anforderung ist zum Beispiel die Eingabe einer URL-Adresse, die um Parameter erweitert werden kann.

**Infoset – XML information set**

Die im Oktober 2001 vom W3C verabschiedete Empfehlung ist ein Satz von Definitionen, die festlegen, wie in den Empfehlungen der XML-Familie auf die verschiedenen Informationseinheiten in konsistenter Weise Bezug genommen werden soll, die in einem XML-Dokument vorkommen. Der Grundgedanke ist, dass jedes wohlgeformte XML-Dokument, das zusätzlich noch die Einschränkungen in Bezug auf die Verwendung von Namensräumen erfüllt, ein Infoset enthält, womit eine hierarchische Struktur von Informationseinheiten gemeint ist, die unterschiedlicher Art sein können. Die Empfehlung unterscheidet insgesamt elf Informationstypen, die jeweils ihre speziellen Eigenschaften haben. Das Infoset wird teilweise im Knotenbaum des DOM-Modells abgebildet, aber nicht vollständig. Ähnliches gilt für das Datenmodell von XPath.

**IDL – Interface Definition Language**

Eine sprachneutrale Sprache, in der die Schnittstellen des DOM in der Empfehlung des W3C beschrieben sind. Der Empfehlung sind zwei "Übersetzungen" der IDL-Definitionen in Form so genannter Bindungen für Java und für ECMAScript beigefügt.

### ***Inhaltsmodell – Content-Model***

Beschreibung der Inhalte eines Elements, entweder innerhalb einer DTD oder in XML Schema. Gibt an, welche Elemente und Daten in welcher Anordnung in einem Element auftauchen dürfen.

### ***Lokalisierungsausdruck***

Ein Ausdruck, der in XPath benutzt wird, um bestimmte Knoten im Knotenbaum auszuwählen. Lokalisierungsausdrücke sind absolute Ausdrücke, wenn sie mit einem Schrägstrich beginnen, der für den Wurzelknoten steht, ansonsten sind es relative Ausdrücke, die vom aktuellen Kontextknoten ausgehen. Während XPath immer nur der Navigation innerhalb eines XML-Dokuments dient, benutzt Xpointer ähnliche Ausdrücke, um Teilmengen in externen XML-Dokumenten zu identifizieren. Dabei werden Lokalisierungsausdrücke innerhalb von Fragmentsbezeichnern aufgebaut.

### ***Markup***

Die Teile eines XML-Dokuments, die zur Auszeichnung verwendet werden, insbesondere die Start- und End-Tags, die dafür genutzt werden, die einzelnen Elemente voneinander zu trennen und sie zugleich zu benennen. Auch Entitätsreferenzen, Kommentare, die Trennzeichen für CDATA-Abschnitte, Dokumenttyp-Deklarationen und Verarbeitungsanweisungen werden zum Markup gezählt und damit von den Zeichendaten unterschieden, die den eigentlichen Inhalt des Dokuments ausmachen.

### ***Metadaten***

Daten die nicht direkt zum Inhalt eines Dokuments gehören, sondern Informationen über das Dokument zur Verfügung stellen, wie Name des Autors, Erstellungsdatum etc.

### ***Mixed content***

Ein spezieller Elementtyp, der es erlaubt, den Inhalt eines Elements aus Zeichendaten und untergeordneten Elementen zu mischen. DTDs können in diesem Fall aber weder die Reihenfolge noch die Häufigkeit des Vorkommens einzelner Elemente festlegen, während bei einem XML Schema entsprechende Einschränkungen möglich sind.

### ***Namensraum – namespace***

Ein XML-Namensraum ist eine Sammlung von Namen, die durch eine URI-Referenz identifiziert wird. Durch die Zuordnung von Element- oder Attributnamen zu Namensräumen ist es möglich, Informationseinheiten

eindeutig zu identifizieren und Namenskonflikte zu vermeiden. Auf diese Weise können gleichlautende Namen mit unterschiedlicher Bedeutung in verschiedenen Vokabularen verwendet werden.

Um Namen, die zu einem Namensraum gehören, einfacher verwenden zu können, kann der URI-Referenz eine Abkürzung zugeordnet werden, die dann als Namensraumpräfix vor dem lokalen Namen eingefügt wird, getrennt durch einen Doppelpunkt. Diese Zuordnung geschieht durch eine Namensraumdeklaration mit der Syntax `<xmlns:namensraumpraefix=uri>`. Solche Deklarationen können auch innerhalb einer Element-Deklaration verwendet werden. Alle Kindelemente dieses Elements erben dann den Namensraum. Die URI-Referenz muss nur eindeutig sein, der Parser sucht anders als bei einem Verweis auf eine DTD oder ein XML Schema keine dort abgelegte Namensliste.

#### ***NCName***

Ein "non-colonized name", also ein XML-Name, der keinen Doppelpunkt enthält. Ein gültiger NCName beginnt entweder mit einem Buchstaben oder einem Unterstrich. Die folgenden Zeichen können eine beliebige Kombination von Buchstaben, Zahlen, Akzenten, diakritischen Zeichen, Punkten, Bindestrichen und Unterstrichen sein.

#### ***Notation***

Notationen werden eingesetzt, um Daten in einem Nicht-XML-Format in ein XML-Dokument einzubinden, etwa eine Bilddatei. Die Notation identifiziert über ihren Namen das Format einer Entität, die vom Parser ignoriert werden soll, also einer "unparsed entity", das Format von Elementen, die ein Notationsattribut enthalten, oder eine Anwendung, an die sich eine Verarbeitungsanweisung richtet.

#### ***Parameter-Entität***

Parameter-Entitäten können nur in einer DTD verwendet werden, und zwar, um Textabschnitte zu benennen, die in der DTD wiederverwendet werden sollen, etwa Teile eines Inhaltsmodells oder eine Liste von Datentypen. Sie werden mit folgender Syntax deklariert: `<!ENTITY % name "Ersetzungstext">`. Die Deklaration muss jedem Verweis auf die Entität vorausgehen, der in der Form `%name` erfolgt.

#### ***Parsed entity***

Eine Entität des XML-Dokuments, die vom XML-Prozessor auf Entitätsreferenzen geprüft werden soll. Der Parser löst alle gefundenen Verweise rekursiv auf, bis für alle Entitäten die entsprechenden

Ersatztexte eingebunden worden sind. Geparsete Entitäten können nur Zeichendaten oder XML-Markup enthalten.

### **Parser**

Ein Programm, das XML-Dokumente liest, auf Wohlgeformtheit prüft und, wenn es sich um einen validierenden Parser handelt, auch auf Gültigkeit. Dabei werden Entity-Referenzen rekursiv aufgelöst. Das Ergebnis des Parsens wird in der Regel an eine weiterverarbeitende Anwendung geleitet. Ist das Dokument nicht wohlgeformt, handelt es sich um kein XML-Dokument und die Verarbeitung wird abgebrochen.

### **PCDATA – Parsed Character Data**

Akronym für "parsed character data". Datentyp für Elemente, die nur Zeichendaten, aber keine Unterelemente enthalten. Dagegen sind Entitäten erlaubt, die vom Parser aufgelöst werden.

### **Processing instruction – Verarbeitungsanweisung**

Verarbeitungsanweisung für eine Anwendung, die ein XML-Dokument verarbeitet. Auf diese Weise können Informationen vom Parser an eine Anwendung weitergereicht werden. Solche Anweisungen können an beliebigen Stellen des Dokuments auftreten. Die Syntax ist `<? Name der Zielanwendung Text der Anweisung ?>`.

### **Prolog**

Ein optionaler, einleitender Teil eines XML-Dokuments, der eine XML-Deklaration, Stylesheet-Deklarationen, Dokumenttyp-Deklarationen, sonstige Processing Instructions und Kommentare enthalten kann.

### **QName – qualifizierter Name**

Ein qualifizierter Name ist ein Name, dem ein Namensraumpräfix vorgestellt ist. Der Namensteil hinter dem Doppelpunkt wird auch als lokaler Name bezeichnet.

### **Renderer**

Prozessor, der ein Dokument in einem bestimmten Format ausgibt.

### **SAX – Simple API for XML**

Eine XML-Programmierschnittstelle, die es ermöglicht, ein XML-Dokument sequenziell abzuarbeiten. Immer wenn dabei ein bestimmtes Ereignis abtritt, etwa das Auftauchen eines Start-Tags, werden bestimmte Methoden aufgerufen, die der Anwendung Informationen bereitstellen, auf die sie dann reagieren kann. Insofern wird von einem ereignisgesteuerten Verfahren gesprochen. Im Unterschied zu DOM wird

kein Objektbaum im Speicher aufgebaut, sodass selbst bei großen XML-Dokumenten nur wenige Speicherressourcen benötigt werden. SAX wurde von David Megginson und der XML-DEV-Community 1998 zunächst für Java entwickelt. SAX 2 ist seit Mai 2000 auf dem Markt und unterstützt inzwischen mehrere Sprachen.

### **Serialisieren**

Die Abbildung eines geordneten, hierarchischen Objekt- oder Knotenbaums in einer sequenziell aufgebauten Datei wird als Serialisierung bezeichnet. Die Ordnung des Baums, der einem XML-Dokument entspricht, ist dabei nach der Regel der Tiefensuche sortiert, d.h., beim ersten Kindknoten werden zunächst die Kindeskindknoten abgelaufen etc., ehe zum nächsten Geschwisterknoten fortgeschritten wird. Die Abbildung der Datei als Knotenbaum wird umgekehrt als Deserialisierung bezeichnet.

### **SGML – Standard Generalized Markup Language**

Internationaler Standard für plattformunabhängige Auszeichnungssprachen. SGML ist eine Metasprache, von der konkrete Auszeichnungssprachen wie HTML abgeleitet worden sind. XML ist eine vereinfachte Untermenge von SGML.

### **Simple Data Type**

Ein Element vom Typ `<simpleType>` enthält nur Text. Auch Attribute werden als Simple Data Types betrachtet, weil ihre Werte nur Text enthalten können.

### **SOAP – Simple Object Access Protocol**

Ein XML-Nachrichtenformat, mit dem Anwendungen über das Web in einer Weise kommunizieren können, die unabhängig von bestimmten Plattformen, Objektmodellen und Programmiersprachen ist. Die Nachrichten werden dabei in ein Umschlagelement verpackt und können auch über Firewalls hinweg Anwendungen direkt verbinden.

### **Stylesheet**

Eine Zusammenstellung von Formatierungsregeln, die angeben, wie Elemente im jeweiligen Zielmedium dargestellt werden sollen. Die Optionen des zunächst gebräuchlichen CSS-Standards werden durch XSL und XSLT wesentlich erweitert.

### **Tag**

Ein von spitzen Klammern umschlossener Elementname, der zur Auszeichnung eines Dokuments verwendet wird. In XML ist dabei die

Wahl der Elementnamen prinzipiell frei, während HTML nur mit vorgegebenen Elementnamen arbeitet. Durch DTDs oder XML Schemas kann die Zahl der gültigen Tags eingeschränkt werden.

### ***Template-Regel***

XSL- oder XSLT-Stylesheets arbeiten mit Template-Regeln, um aus den Knoten des Quelldokuments die Knoten des Zieldokuments zu erzeugen. Diese Template-Regeln, die ähnlich wie Funktionen eingesetzt werden können, bestehen aus zwei Teilen. Der erste Teil beinhaltet meist einen Mustervergleich mit Hilfe des `match`-Attributs, der benutzt wird, um im Quelldokument die Knoten zu identifizieren, die umgewandelt oder formatiert werden sollen. Der zweite Teil ist ein Template, das mit Hilfe einzelner Kindelemente die vorgesehenen Umwandlungen festlegt. Alternativ dazu können auch benannte Templates eingesetzt werden.

### ***Token***

Token sind Zeichen oder Zeichenfolgen, die als nicht zerlegbare Grundbausteine einer Sprache eingesetzt werden. Zeichen wie `< oder >` oder Schlüsselwörter, Element- und Attributnamen sind Token in XML. Auch die Werte bestimmter Attribute (ID, IDREF, ENTITY, NMTOKEN) werden als Token oder als Liste von Token behandelt (IDREFS, ENTITIES, NMTOKENS).

### ***UDDI – Universal Description, Discovery and Integration***

Das von IBM, Microsoft und Ariba Ende 2000 in Gang gesetzte UDDI-Projekt ist ein weltweites Unternehmensnetz, in dem Firmen Webdienste sowohl registrieren lassen können, um sie anderen verfügbar zu machen, als auch selbst suchen können, um die entsprechenden Dienstleistungen zu nutzen.

### ***Unicode***

Zeichensatzstandard, in dem versucht wird, möglichst alle Zeichen, die in den Sprachen der Welt vorkommen, zu erfassen. Informationen über das Unicode Consortium, das den Standard pflegt, sind über [www.unicode.org](http://www.unicode.org) zu bekommen.

### ***URI – Uniform Resource Identifier***

Oberbegriff für Zeichenketten, die zur eindeutigen Identifizierung von Ressourcen im Web benutzt werden. Am Anfang steht ein URI-Schema-Name, der meist das Zugriffsprotokoll nennt – `http`, `ftp`, `mailto` etc. –, dann folgt getrennt durch einen Doppelpunkt der Pfad zu der Ressource. In der Praxis werden heute fast ausschließlich ULRs als URI verwendet.

Die allgemeine Syntax von URIs ist in **IETF RFC 2396** definiert, **www.ietf.org/rfc/rfc2396.txt** enthält alle Details dazu.

Der Nachteil einer URL – Uniform Resource Locator – ist allerdings, dass sie ungültig werden, wenn sich der Speicherort der Ressource ändert. Deshalb gibt es seit einigen Jahren Versuche, eine weltweit eindeutige Benennung von Webressourcen in Form von URNs – Uniform Resource Name –, also mit Hilfe von Namensräumen, zu ermöglichen. Zum Stand dieser Bemühungen gibt es vom W3C eine Art Klarstellung vom 21.9.2001 unter [www.w3.org/TR/2001/NOTE-uri-clarification-20010921/](http://www.w3.org/TR/2001/NOTE-uri-clarification-20010921/).

#### ***URL – Uniform Resource Locator***

Adresse, die den genauen Ort einer Ressource im Internet angibt. Die Syntax ist durch die Internet Proposed Standards RFC 1738 und 1808 für absolute und relative Adressen definiert. Ein URL besteht aus einem Schema (http, https, news, ftp, mailto, nntp, telnet, ldap) und einem schemaspezifischen Teil, der aus folgenden Elementen zusammengesetzt ist:

```
"//" [benutzer [":" kennwort ] "@" ] host [ ":" port ] "/"  
url_path
```

wobei die eingeklammerten Elemente optional sind.

#### ***Unparsed entity***

Ein beliebiger Block von Nicht-XML-Daten in einem XML-Dokument, etwa binäre Daten für ein Bild oder auch Textdaten, die der Parser nicht auswerten soll. Um solche Datenblöcke in das Dokument einzubeziehen, ist eine entsprechende Notation notwendig, die das Datenformat oder den Typ der entsprechenden Ressource angibt. Der Parser muss die Informationen über die ungeparsten Entitäten an eine Anwendung weiterreichen.

#### ***Validierung***

Wenn einem XML-Dokument eine DTD oder ein XML Schema zugeordnet ist, kann ein XML-Parser eine Validierung durchführen, die prüft, ob das Dokument dem darin festgelegten Vokabular und der Grammatik und den geforderten Datentypen entspricht und die in der XML-Spezifikation festgelegten Regeln einhält. Ist das der Fall, handelt es sich um ein gültiges XML-Dokument. Wird ein Fehler gefunden, muss er an die jeweilige XML-Anwendung gemeldet werden.

### **Vererbung**

Generell die Weitergabe der Eigenschaften und Methoden einer Klasse von Objekten an davon abgeleitete Objekte. Entsprechend erben bei XML Schema abgeleitete Datentypen die Eigenschaften des jeweiligen Basistyps.

### **Wohlgeformtes XML**

Ein Dokument, das die syntaktischen Regeln der XML-Spezifikation erfüllt. Das Dokument muss ein Dokumentelement enthalten und alle anderen Elemente müssen darin eingeschlossen sein, wobei korrekt geschachtelt werden muss. Für jedes Start-Tag muss ein passendes End-Tag existieren. Eine DTD oder ein XML Schema sind nicht erforderlich.

### **W3C – Worldwide Web Consortium**

International anerkannte Institution, die seit 1994 mit der Schaffung von Standards für Webtechnologien befasst ist. Arbeitsentwürfe für neue Standards oder Erweiterungen bestehender Standards werden zunächst in Form von Working Drafts zur Diskussion gestellt. Die verabschiedeten Standards werden als Empfehlung – recommendation – bezeichnet.

### **WSDL – Web Services Description Language**

Ein spezielles XML-Vokabular, mit dem für einen Webdienst beschrieben wird, wie SOAP-Nachrichten aussehen müssen, um den Dienst nutzen zu können.

### **Wurzelement**

Das Basiselement eines XML-Dokuments, das alle anderen Elemente in sich einschließt, auch Dokumentelement genannt.

### **Wurzelknoten**

Basisknoten des Knotenbaums, von dem alle anderen Knoten abstammen.

### **XHTML**

Ein Neuformulierung von HTML 4 in Form einer XML-Anwendung. XHTML 1.0 ist seit dem 26. Januar 2000 als Standard vom W3C definiert. Dabei wird die Absicht verfolgt, eine Familie verschiedener Module und Dokumenttypen von XHTML bereitzustellen, die eine Nutzung für ganz unterschiedliche Gerätetypen ermöglicht.

### **XML – Extensible Markup Language**

Eine Untermenge von SGML, die für Webanwendungen optimiert und vereinfacht wurde. XML bietet eine einheitliche Methode für die Beschreibung und den Austausch strukturierter Daten, unabhängig von

einer bestimmten Plattform oder Programmiersprache. Dabei wird der Inhalt völlig von der Form der Darstellung getrennt, die aber jederzeit über Stylesheets zugeordnet werden kann.

#### ***XML-Anwendung***

Ein konkretes Vokabular für einen Sachbereich, das mit Hilfe der Metasprache XML formuliert worden ist. Beispiele sind etwa MathML, eine Sprache für mathematische Formeln oder SVG, eine Beschreibung von Vektorgrafiken.

#### ***XLink – XML Linking Language***

Ein XML-Vokabular für Elemente, die in XML-Dokumente eingefügt werden können, um Links zu Ressourcen aufzubauen und zu beschreiben. Über URL-basierte Hyperlinks und Anker hinaus, wie sie von HTML bekannt sind, können Verknüpfungen zu beliebigen Positionen in einem Dokument oder auch Verknüpfungen in mehrere Richtungen hergestellt werden. Seit Juni 2001 ist XLink als Empfehlung vom W3C verabschiedet.

#### ***XML-Deklaration***

Die erste Zeile eines XML-Dokuments kann und sollte in der Regel eine XML-Deklaration als Verarbeitungsanweisung enthalten. Sie macht das Dokument für einen Prozessor sofort als XML-Dokument erkennbar. Neben der Angabe der XML-Version kann hier der zu verwendende Zeichensatz explizit deklariert und außerdem mit dem Attribut `standalone` angegeben werden, ob das Dokument als abgeschlossene Einheit verwendet werden kann oder ob externe Dateien geladen werden müssen, etwa eine externe DTD. Ein typisches Beispiel ist:

```
<?xml version="1.0" encoding="ISO-8859-1" standalone="yes"
?>
```

#### ***XML-Dokument***

Ein Dokument, das im Sinne der XML-Empfehlung wohlgeformt ist. Ein XML-Dokument kann zugleich ein gültiges oder auch ein ungültiges Dokument sein, das einer DTD oder einem XML Schema entspricht oder nicht. Die logische Struktur eines XML-Dokuments ist zusammengesetzt aus Informationseinheiten wie Elementen, Kommentaren und Verarbeitungsanweisungen. Physikalisch ist das Dokument zusammengesetzt aus Entitäten.

### ***XML-Prozessor***

Generischer Begriff für eine Software, die XML-Dokumente lesen und auf ihren Inhalt und ihre Struktur zugreifen kann. Der XML-Prozessor arbeitet in der Regel im Dienst einer Anwendung, etwa eines Webbrowsers wie dem Internet Explorer.

### ***XPath – XML Path Language***

Eine Sprache, die es ermöglicht, Teile eines XML-Dokuments zu adressieren. XPath verwendet dafür eine Adressierungssyntax, die auf den Knotenbaum Bezug nimmt, der aus der Struktur eines Dokuments aufgebaut werden kann. Die Lokalisierungsausdrücke, die ähnlich wie Dateipfade aufgebaut sind, werden sowohl von XSLT als auch von XPointer benutzt.

### ***XPointer – XML Pointer Language***

Im September 2001 hat das W3C für XPointer eine Candidate Recommendation veröffentlicht. Die Sprache kann verwendet werden, um auf der Basis von XPath-Ausdrücken über entsprechend erweiterte URI-Verweise Fragmente in Webressourcen zu identifizieren, etwa einzelne Elemente oder auch Teile einer Zeichenkette, die über Positionsangaben angesteuert werden können.

### ***XML Schema***

Eine formale Festlegung der Struktur einer bestimmten Klasse von Dokumenten. XML Schema legt fest, welche Elemente und Attribute in einer gültigen Dokumentinstanz erlaubt sind und in welcher Anordnung die Elemente auftreten dürfen. Gleichzeitig können, anders als bei DTDs, denen sie ansonsten funktionell entsprechen, sehr präzise Definitionen der Datentypen gegeben werden, die den Inhalt der Elemente und die Werte von Attributen einschränken. Im Unterschied zu DTDs ist ein XML Schema zudem selbst ein XML-Dokument, kann also auch von XML-Prozessoren entsprechend verarbeitet werden.

### ***XML Schema Definition Language***

Die XML-Sprache zur Definition von XML Schemas, die seit Mai 2001 als W3C-Standard vorliegt. Die Sprache besteht aus zahlreichen Elementen, die die Konstruktion von Datenmodellen und die Definition und Ableitung von Datentypen erlauben, wobei eine große Anzahl von Datentypen bereits vorgegeben wird.

### ***XSL – Extensible Stylesheet Language***

Seit Oktober 2001 vom W3C als Standard fixierte Sprache für die Formulierung von Stylesheets. XSL besteht aus zwei Teilen, einer Sprache für die Steuerung von Transformationen, die ein XML-Quelldokument in ein anderes XML- oder auch Nicht-XML-Dokument umwandeln kann, etwa eine HTML-Datei, und einem Vokabular für die Spezifizierung von Formaten für die Präsentation der Quelldaten. XSL arbeitet dabei mit Formatierungsobjekten, die dem Gestalter jeweils einen Satz von Eigenschaften anbieten, die er benutzen kann, um die Gestaltung des Ergebnisdokuments im Detail festzulegen.

Anders als bei CSS kann die Formatierung mit Verfahren wie Datensortierung, Filterung oder auch mit Berechnungen verknüpft werden, wobei das Quelldokument jeweils unverändert bleibt, da alle Ausgaben in einem separaten "result tree" zusammengestellt werden.

### ***XSLT – XSL Transformations***

Die schon 1999 verabschiedete Empfehlung für XSLT erlaubt auf der Basis von XPath-Ausdrücken, die für die Auswahl von Elementen in einem XML-Dokument verwendet werden, aus einem Quelldokument ein umgewandeltes Zieldokument zu erzeugen, das selbst wiederum ein XML-Dokument sein kann, aber nicht muss. Die vorzunehmenden Veränderungen werden in einem wohlgeformten XML-Dokument in Form von Templates zusammengestellt. XSLT kann sowohl als Teil von XSL als auch unabhängig von XSL verwendet werden, etwa um ein XML-Vokabular in ein anderes zu übersetzen.

### ***Zeichendaten – character data***

Alle Zeichenketten in einem XML-Dokument, die nicht zum Markup gehören, werden als Zeichendaten – charakter data – verstanden, etwa der Inhalt eines Elements und der Wert eines Attributs. Zeichendaten werden im DOM-Baum als Textobjekte gespeichert.