



Die **Erweiterbare Auszeichnungssprache** (englisch *Extensible Markup Language*), abgekürzt **XML**, ist eine Auszeichnungssprache zur Darstellung hierarchisch strukturierter Daten im Format einer Textdatei, die sowohl von Menschen als auch von Maschinen lesbar ist.

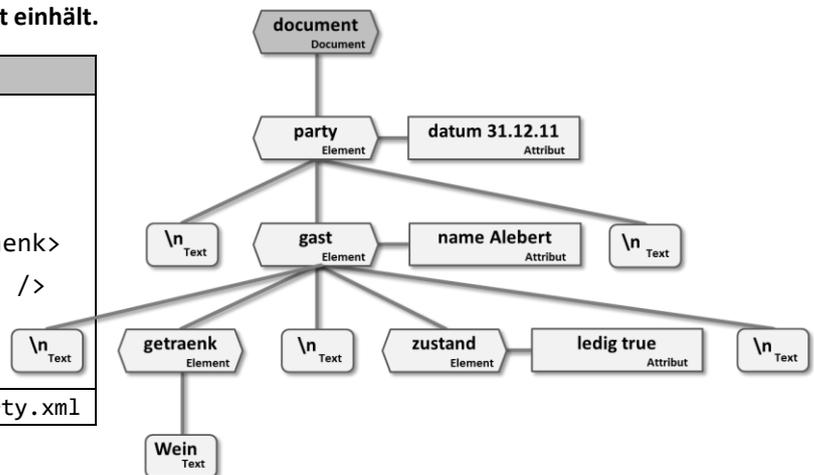
Quelle: wikipedia 2017 „XML“

Ein XML-Dokument heißt „wohlgeformt“ (oder englisch *well-formed*), wenn es alle XML-Regeln einhält. Beispielhaft seien hier folgende genannt:

- Das Dokument besitzt genau ein Wurzelement. Als Wurzelement wird dabei das jeweils äußerste Element bezeichnet, z.B. <html> in XHTML.
- Alle Elemente mit Inhalt besitzen einen Beginn- und einen End-Auszeichner (-Tag) (z.B. <eintrag>Eintrag 1</eintrag>). Elemente ohne Inhalt können auch in sich geschlossen sein, wenn sie aus nur einem Auszeichner bestehen, der mit /> abschließt (z.B. <eintrag />).
- Die Beginn- und End-Auszeichner sind ebenentreu-paarig verschachtelt. Das bedeutet, dass alle Elemente geschlossen werden müssen, bevor die End-Auszeichner des entsprechenden Elternelements oder die Beginn-Auszeichner eines Geschwisterelements erscheinen.
- Ein Element darf nicht mehrere Attribute mit demselben Namen besitzen.
- Attributeigenschaften müssen in Anführungszeichen stehen.
- Die Beginn- und End-Auszeichner beachten die Groß- und Kleinschreibung (z.B. <eintrag></Eintrag> ist nicht gültig).

Der Standard definiert ein XML-Dokument als gültig (oder englisch *valid*), wenn es wohlgeformt ist, den Verweis auf eine Grammatik (z.B. einer Dokumenttypdefinition oder eines XML-Schemas) enthält und das durch die Grammatik beschriebene Format einhält.

Beispiel einer XML-Datei	
①	<?xml version="1.0" ?>
②	<party datum="31.12.11">
③	<gast name="Albert">
④	<getraenk>Wein</getraenk>
⑤	<zustand ledig="true" />
⑥	</gast>
⑦	</party>
	party.xml



Die wichtigsten Knotentypen in DOM sind:

Ein **Dokumentknoten** stellt die gesamte Baumstruktur dar;

Ein **Elementknoten** entspricht exakt einem Element in XML

Ein **Attributknoten** entspricht exakt einem Attribut in XML

Ein **Textknoten** stellt den textuellen Inhalt eines Elements oder Attributs (dessen Value) dar

Ein **Kommentarknoten** repräsentiert den Inhalt eines XML-Kommentars (den Inhalt zwischen '<!--' und '-->').

Die Java-APIs für XML

DOM (Document Object Model): Liest das gesamte XML-Dokument in eine interne Struktur ein. Standard-DOM unabhängig von der Programmiersprache

SAX (Simple API for XML Parsing) SAX basiert auf einem Ereignismodell, die XML-Datei wird wie ein Datenstrom gelesen, und für erkannte Elemente wird ein Ereignis ausgelöst. Nachteil: wahlfreier Zugriff auf ein einzelnes Element nicht ohne Zwischenspeicherung möglich ist.

