

## XML

*Raunig, Elisabeth; elisabeth.raunig@uni-graz.at*

Die *eXtensible Markup Language*, kurz XML, ist ein einfaches, hierarchisch aufgebautes Textformat, das 1998 vom World Wide Web-Konsortium entwickelt wurde. Es wurde vor allem zum Aufzeigen von Strukturen in Texten oder Datenbeständen, zum Austausch und zur Speicherung von Daten im Web entwickelt. Es ist sowohl menschenlesbar als auch maschinenlesbar, da für die Beschreibung die gleichen Zeichen verwendet werden wie die Zeichen der Beschriebenen.

XML fügt in einen Text oder Datenbestand Auszeichnungen ein, indem beispielsweise ein Wort oder eine Wortkombination getagged wird. Beim Tagging wird das Wort von einem Auszeichnungselement oder 'Tag' eingeschlossen und damit beschrieben. Dieses Markup steht immer zwischen den Spitzklammern `lt;` und `gt;`, der Fließtext oder die Daten stehen außerhalb der Klammern. In folgendem Beispiel ist deutlich erkennbar, dass das Markup den Text innerhalb der Klammern genauer beschreibt und ihm eine Bedeutung gibt: `<name>Franz</name>`.

Die Kombination aus Anfangstag, Endtag und Text bildet ein Element. Jedes Element muss mit dem Anfangstag (`<name>`) geöffnet und mit dem Endtag (`</name>`) geschlossen werden, letzteres wird immer durch den, dem Endtag vorangestellten, Schrägstrich gekennzeichnet. Jedes Element kann weiter über Attribute und Attributwerte verfügen. Bei obigem Beispiel könnte das sein: `<name typ="vorname">Franz</name>`. Dabei ist 'typ' das Attribut, das mit seinem Wert 'vorname' angibt, dass es sich bei dem getaggten Namen nicht um einen Nachnamen, sondern um einen Vornamen handelt.

XML-Notation unterscheidet auch zwischen Groß- und Kleinschreibung: `<name>` hat als Tag eine andere Bedeutung als `<Name>`. Weitere Regeln, denen XML unterliegt, sind: leere Tags können von der Schreibung `<name></name>` zu der gleichwertigen Schreibung `<name/>` verkürzt werden; Tags können innerhalb der spitzen Klammern Buchstaben, Zahlen und Striche enthalten, Leerzeichen jedoch sind nur für die Trennung von einem oder mehreren Attributen erlaubt.

XML folgt einer hierarchischen Struktur. Das bedeutet, dass ein Element andere Elemente beinhalten kann, sich Elemente jedoch nicht überlappen können.

Richtig ist die Auszeichnungsvariante: `<name><vorname>Franz</vorname></name>`, falsch wäre: `<name><vorname>Franz</name></vorname>`. Zusätzlich müssen alle Elemente in einem Wurzelement enthalten sein, um eine Baumstruktur zu bilden. Dieses Wurzelement bildet ein Elternelement für alle darin enthal-

tenen Elemente, die Kindelemente oder *child-elements* genannt werden. Unter diesen Kindelementen kann es Geschwisterelemente geben, damit sind Elemente gemeint, die auf derselben Ebene liegen und nicht ineinander verschachtelt sind. Jedes Element unter dem Wurzelement kann weitere Kindelemente einschließen. Wenn alle diese Regeln eingehalten werden, wird von einem wohlgeformten Dokument gesprochen.

Damit bietet XML unendlich viele Möglichkeiten, um Daten hierarchisch zu modellieren und zu annotieren. Daher gibt es unterschiedliche XML-Dialekte, um Standards für den Datenaustausch zu schaffen, wie z. B. XHTML, MathML, SVG oder TEI.

### **Literatur:**

- XML. URL: <https://www.w3.org/XML/>
- Vogeler, Georg; Sahle, Patrick: XML. In: Digital Humanities. Eine Einführung. Stuttgart: 2017, S. 128–148.

### **Verweise:**

Markup, TEI, Modellierung, Textmodellierung, Annotation, Annotationsstandards, Einführung XML/TEI, Office-Formate

### **Themen:**

Einführung, Archivierung, Digitale Editionswissenschaft

### **Lexika**

- Edlex: Editionslexikon
- Lexicon of Scholarly Editing

### **Zitiervorschlag:**

Raunig, Elisabeth. 2021. XML. In: KONDE Weißbuch. Hrsg. v. Helmut W. Klug unter Mitarbeit von Selina Galka und Elisabeth Steiner im HRSM Projekt "Kompetenznetzwerk Digitale Edition". URL: <https://gams.uni-graz.at/o:konde.215>