

Einführung: Was ist XML/TEI?

Lisa Rieger; lrieger@edu.aau.at

XML (*eXtensible Markup Language*) ist ein Standard, der vom World-Wide-Web-Konsortium (W3C) als Vereinfachung der älteren SGML (*Standard Generalized Markup Language*, ISO 8879) zur strukturierten Darstellung von Dokumenten und Daten, v. a. auch für den Gebrauch im Zusammenhang mit dem Internet, entwickelt wurde. Auf der W3C-Homepage (W3C o. J.) finden sich sowohl einführende Erläuterungen und Tutorials als auch sämtliche Dokumente, in denen die Standards definiert werden:

- *Extensible Markup Language* (XML) 1.0 (Fifth Edition) (W3C Recommendation 2008)
- *Namespaces in XML 1.0* (Third Edition) (W3C Recommendation 2009a)
- *XML Inclusions* (XInclude) Version 1.0 (Second Edition) (W3C Recommendation 2006)
- *XML Information Set* (Second Edition) (W3C Recommendation 2004)
- *xml:id* Version 1.0 (W3C Recommendation 2005)
- *XML Fragment Interchange* (W3C Recommendation 2001)
- *XML Base* (Second Edition) (W3C Recommendation 2009b)
- *Associating Style Sheets with XML documents* 1.0 (Second Edition) (W3C Recommendation 2010)

Die Hauptbausteine eines XML-Dokuments sind Elemente, die aus einem öffnenden Tag (`<lt;...gt;`) und einem schließenden Tag (`<lt;/...gt;`) bestehen. Der Text zwischen den Tags gilt dann als Teil des Elements und unterliegt seinen Bestimmungen. Elemente ohne Text zwischen öffnendem und schließendem Tag sind leere Elemente (z. B. `<empty/>`), die dem Text Zusatzinformationen hinzufügen. Attribute können ein Element genauer bestimmen. Sie werden im Starttag der Elementbezeichnung nachgestellt und ihr Wert wird nach einem Gleichheitszeichen durch Anführungszeichen begrenzt (`<lt;Substantiv Kasus="Nominativ"gt;Edition</Substantivgt;`). (Eckstein 2000, S. 6 ff.) Zudem können XML-Dokumente noch Deklarationen, Kommentare, Zeichenreferenzen und Verarbeitungsanweisungen enthalten. Damit ein XML-Dokument gültig ist, muss es sowohl hinsichtlich seiner physikalischen Struktur wohlgeformt als auch, vorgegeben durch den Dokumenttyp, valide sein (vgl. Schema). (Doss 2000, S. 43–46)

Die TEI (TEI Consortium a) ist ein Konsortium, das auf Basis eines erweiterbaren XML-Schemas eine Markup-Sprache – ebenfalls TEI genannt – zur

Auszeichnung der Eigenschaften von Dokumenten im Bereich der Geistes- und Sozialwissenschaften entwickelt und mit einer Open-Source-Lizenz veröffentlicht hat. Der Schwerpunkt liegt dabei auf der Beschreibung von Primärquellen für Forschung und Analysen. Wie bei XML werden Leitfäden (aktuellste Version: P5 *Guidelines*) und andere wichtige Dokumente (TEI Consortium b) direkt auf der Homepage zur Verfügung gestellt. TEI erweist sich gerade für literaturwissenschaftliche Arbeiten wie die Edition von Texten als besonders geeignet, da es sowohl die Repräsentation der Zeichensequenz und Textstruktur als auch das Annotieren editionswissenschaftlicher und dokumentenbezogener Metadaten ermöglicht und zudem langfristig und plattformübergreifend nutzbar ist. (Schöch 2016, S. 335)

Literatur:

- Bader, Winfried: Was ist die Text Encoding Initiative (TEI)? Was ist die Text Encoding Initiative (TEI)? In: Computergestützte Text-Edition. Beihefte zu editio 12: 1999, S. 9-20.
- Box, Don; Skonnard, Aaron; Lam, John: Essential XML. XML für Softwareentwicklung. Deutsche Übersetzung von Mathias Born und Michael Tamm Essential XML. München: 2001.
- Doss, George: XML-Tipps. Deutsche Übersetzung von Adreas Fieback XML-Tipps. Frankfurt: 2000.
- Eckstein, Robert: XML kurz amp; gut. Übersetzung von Nick Klever XML kurz amp; gut. Köln (u. a.): 2000.
- Jele, Harald: XML-TEI-kodierte Texte editieren: 2013. URL: <https://ubdocs.aau.at/open/voll/fodok/ub/AC11224010.pdf>.
- Jones, Christopher A.; Drake Jr., Fred L.: Python amp; XML. Deutsche Übersetzung von Dinu C. Gherman Python amp; XML. Köln (u. a.): 2002.
- Klettke, Maïke; Meyer, Holger: XML amp; Datenbanken. Konzepte, Sprachen und Systeme XML amp; Datenbanken. Heidelberg: 2003.
- Schöch, Christof: Ein digitales Textformat für die Literaturwissenschaften. Die Richtlinien der Text Encoding Initiative und ihr Nutzen für Textedition und Textanalyse Ein digitales Textformat für die Literaturwissenschaften. In: Romanische Studien: 2016, S. 325–364.
- TEI: Text Encoding Initiative. URL: <https://tei-c.org/>
- TEI: Guidelines. URL: <https://tei-c.org/guidelines/>
- XML Technology. URL: <https://www.w3.org/standards/xml/>

- Associating Style Sheets with XML documents 1.0 (Second Edition). W3C Recommendation 28 October 2010 Associating Style Sheets with XML documents 1.0 (Second Edition). URL: <https://www.w3.org/TR/xml-styleSheet/>
- Namespaces in XML 1.0 (Third Edition). W3C Recommendation 8 December 2009 Namespaces in XML 1.0 (Third Edition). URL: <https://www.w3.org/TR/xml-names/>
- XML Base (Second Edition). W3C Recommendations 28 January 2009 XML Base (Second Edition). URL: <https://www.w3.org/TR/2009/REC-xmlbase-20090128/>
- Extensible Markup Language (XML) 1.0 (Fifth Edition). W3C Recommendation 26 November 2008 Extensible Markup Language (XML) 1.0 (Fifth Edition). URL: <https://www.w3.org/TR/2008/REC-xml-20081126/>
- XML Inclusions (XInclude) Version 1.0 (Second Edition). W3C Recommendation 15 November 2006 XML Inclusions (XInclude) Version 1.0 (Second Edition). URL: <https://www.w3.org/TR/2006/REC-xinclude-20061115/>
- xml:id Version 1.0. W3C Recommendation 9 September 2005 xml:id Version 1.0. URL: <https://www.w3.org/TR/xml-id/>
- XML Information Set (Second Edition). W3C Recommendation 4 February 2004 XML Information Set (Second Edition). URL: <https://www.w3.org/TR/2004/REC-xml-infoSet-20040204/>
- XML Fragment Interchange. W3C Candidate Recommendation 12 February 2001 XML Fragment Interchange. URL: <https://www.w3.org/TR/xml-fragment>

Software:

CWRC-Writer, Oxygen, Oxygen-TEI-Facsimile-Plugin, TEI Boilerplate, TEI Comparator, TEI Publisher, Text-TEI-collate, TEITOK

Verweise:

Digitale Edition, Digitale Nachhaltigkeit, Lemmatisierung, Open Access, Annotation, Textmodellierung, Metadaten, Annotationsstandards, Markup, TEI, Part of Speech Tagging, Tagger, TagSets, WebLicht, xTokenizer, Alternative Textkodierung

Themen:

Einführung, Metadaten, Digitale Editionswissenschaft

Zitervorschlag:

Rieger, Lisa. 2021. Einführung: Was ist XML/TEI?. 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.79>