

# TEI

Eine XML-Datei kann auf ein Schema zurückgreifen, dass die „Grammatik“ vorgibt, d.h. welche Elemente gibt es und wie bzw. wo dürfen sie verwendet werden. Anhand dieses Schema kann die Software überprüfen, ob das Dokument das Schema korrekt umsetzt (validiert) oder nicht.

Im Fall des Arbeitsvorhabens CAGB wird ein Schema auf Basis der TEI-P5-Richtlinie verwendet, die von der „Text Encoding Initiative“ bereitgestellt wurde. Letztere arbeitet seit 1987 an diesen Richtlinien, die Elemente bereitstellen, mit denen u.a. Manuskripte ausgezeichnet bzw. beschrieben werden können.

Die TEI ist also eine speziell für geisteswissenschaftliche Forschungsprojekte entwickelte Auszeichnungssprache. Anstelle von bestimmten Formatierungen und Kürzeln benutzt man entsprechende Elemente. Beispiele:

In bisheriger Druckausgabe	TEI-basiertes XML
<i>Proffessor</i>	Pr<i>offessor</i>
außerdem] <i>über der Zeile</i>	<add place="über der Zeile">außerdem</add>

Wie im Beispiel zu sehen ist werden die Textpassagen mit Hilfe der TEI semantisch aus7 gezeichnet, d.h. die Elemente tragen schon die Bedeutung der Auszeichnung in sich: so steht <ex> für „editorial expansion“ und <add> für „addition“. Dadurch sind die Auszeichnungen unabhängig von ihrer späteren Formatierung. Beispiel:

ich konnte <hi rend="underline">gänzlich</hi> nicht verstehen

Hier wird lediglich vermerkt, dass im Manuskript das Wort „gänzlich“ unterstrichen war. Ob im Druck oder im Web die entsprechende Stelle nun tatsächlich unterstrichen oder vielleicht eher gesperrt wird, muss hier nicht entschieden werden.

Da die TEI für viele verschiedene Textsorten und Anwendungsfälle gedacht ist, umfasst sie sehr viel mehr Elemente als normalerweise in einem Projekt benötigt werden. Im Teucho – Zentrum für Handschriften- und Textforschung an der Universität Hamburg wurde mit Hilfe der TEI eine Leitlinie erarbeitet, wie die unterschiedlichen Teile einer Handschriftenbeschreibung mit XML strukturiert und ausgezeichnet werden können. [FUSSNOTE ]. Von TELOTA wurde auf dieser Basis ein XML-Schema für die digitale Arbeitsumgebung entwickelt, gegen das die XML-Dokumente zukünftig validiert, d.h. geprüft werden können.

Der Baum eines TEI-kodierten XML-Dokument besteht immer aus zwei Teilen: dem <tei-Header/> und dem <text/>. Während die eigentliche Handschriftenbeschreibung sich im letzteren befindet, werden im teiHeader Metangaben zum XML-Dokument notiert. Grobe Struktur eines TEI-kodierten XML-Dokuments:

Struktur	Struktur in XML-Dokument
Metaangaben zum XML-Dokument	<pre> &lt;TEI xmlns="http://tei-c.org/ns/1.0" xml:lang="de"&gt;   &lt;teiHeader&gt;     &lt;fileDesc&gt;       &lt;titleStmt&gt;         &lt;title&gt;[Dokumenttitel]&lt;/title&gt;       &lt;/titleStmt&gt;       &lt;publicationStmt&gt;         ... [Angaben zum Herausgeber]       &lt;/publicationStmt&gt;     &lt;/fileDesc&gt;     &lt;revisionDesc&gt;       ... [Angaben zur Bearbeitung des XML-Dokuments]     &lt;/revisionDesc&gt;   &lt;/teiHeader&gt; </pre>
Handschriftenbeschreibung	<pre>     &lt;text&gt;       &lt;body&gt;         ... [Eigentliche Handschriftenbeschreibung]       &lt;/body&gt;     &lt;/text&gt;   &lt;/TEI&gt; </pre>