

XML-Auszeichnung mittels TEI und DTABf

Susanne Haaf

3. Münchner Digital Humanities Summerschool
11.-15. September 2017

Problem





Extensible Markup Language

- Standard-Markupsprache für wissenschaftliche (und andere) Anwendungen:

```
<markup type="XML">XML</markup>
```

- Vorteil:
 - XML ist ein freier Standard.
 - Ich kann alles annotieren (auch semantische Informationen), ich muss mich nur an wenige Regeln halten
 - Inhalt und Darstellung des Inhalts sind getrennt
 - Verschiedene, leistungsstarke XML-Technologien
→ **XPath**, **XSLT**, **XQuery**
 - XML-basierte Tools verfügbar, die ich für meine Daten nachnutzen kann, z.B.:



Extensible Markup Language

- Ich kann auch mein eigenes Markup erfinden:

||:Auch das:|| ist eine /Art/ von Mark[^]up.

Nachteil:

- Ich muss es gut dokumentieren, damit ich verstanden werde.
- Ich muss mir selbst Tools (i.e. Programme) bauen, die mein Markup verstehen. Das ist aufwändig.
- Jeder muss sich immer wieder aufs Neue überlegen, was ausgezeichnet werden soll und wie

Problem



- Es gibt noch andere verbreitete XML-Formate, z.B.:

frei

- **DITA**, **Docbook**: für die Dokumentation
- **TCF**, **GrAF**: für die rein linguistische Annotation
- **Dublin Core**, **CMDI**, **METS**, **MODS**: für Metadaten
- **XHTML**: für Webseiten

proprietär

- **DOCX**: für die Textauszeichnung in Word
- **CTE**: für die Textauszeichnung im Classical Text Editor

→ **TEI**: freies Format für die Textauszeichnung in Editionen, Korpora etc.

- TEI
 - Text Encoding Initiative, www.tei-c.org
 - Gegründet 1987
 - Struktur: Board of Directors, Technical Council, Community
- TEI-Guidelines
 - Richtlinien für die XML-Auszeichnung von Texten
 - Annotation semantischer, typographischer, linguistischer, graphischer etc. Phänomene

```
<hi rendition="#bold">Platon</hi>
```

```
<persName>Platon</persName>
```

»The impact of the TEI on digital scholarship has been enormous. «

- »Encoding Schema der Wahl für:
 - Kritische Editionen
 - Wissenschaftliche Referenzwerke
 - Große linguistische Korpora
 - Detaillierte Metadaten zu elektronischen Texten und Kulturerbe-Sammlungen«

Cf. <http://www.tei-c.org/About/history.xml>

Projects, using the TEI



University of Oxford Text Archive



etc. ...

Cf. www.tei-c.org/Activities/Projects/

Die TEI-Guidelines P5-Release

P5: Guidelines for Electronic Text Encoding and Interchange

Version 2.6.0. Last updated on 20th January 2014, revision 12802

[\[English\]](#) [\[Deutsch\]](#) [\[Español\]](#) [\[Italiano\]](#) [\[Français\]](#) [\[日本語\]](#) [\[한국어\]](#) [\[中文\]](#)



Front Matter

[Title](#)

- i. [Releases of the TEI Guidelines](#)
- ii. [Dedication](#)
- iii. [Preface and Acknowledgments](#)
- iv. [About These Guidelines](#)
- v. [A Gentle Introduction to XML](#)
- vi. [Languages and Character Sets](#)

Back Matter

- Appendix A [Model Classes](#)
- Appendix B [Attribute Classes](#)
- Appendix C [Elements](#)
- Appendix D [Attributes](#)
- Appendix E [Datatypes and Other Macros](#)
- Appendix F [Bibliography](#)
- Appendix G [Prefatory Notes](#)
- Appendix H [Colophon](#)

Text Body

- 1 [The TEI Infrastructure](#)
- 2 [The TEI Header](#)
- 3 [Elements Available in All TEI Documents](#)
- 4 [Default Text Structure](#)
- 5 [Non-standard Characters and Glyphs](#)
- 6 [Verse](#)
- 7 [Performance Texts](#)
- 8 [Transcriptions of Speech](#)
- 9 [Dictionaries](#)
- 10 [Manuscript Description](#)
- 11 [Representation of Primary Sources](#)
- 12 [Critical Apparatus](#)
- 13 [Names, Dates, People, and Places](#)
- 14 [Tables, Formulæ, Graphics and Notated Music](#)
- 15 [Language Corpora](#)
- 16 [Linking, Segmentation, and Alignment](#)
- 17 [Simple Analytic Mechanisms](#)
- 18 [Feature Structures](#)
- 19 [Graphs, Networks, and Trees](#)
- 20 [Non-hierarchical Structures](#)
- 21 [Certainty, Precision, and Responsibility](#)
- 22 [Documentation Elements](#)
- 23 [Using the TEI](#)

TEI sourcecode

- [Using the TEI Sourceforge Repository](#)
- [Sourceforge Subversion Repository](#)
- [Bug Reports, Feature Requests, etc.](#)

TEI: Grundlegende Dokumentstruktur I (vereinfacht)

```
1  <?xml version="1.0" encoding="UTF-8"?>↵
2  <TEI>↵
3    ····<teiHeader>[Metadata]</teiHeader>↵
4  ····<text>↵
5    ······<front>↵
6      ········<titlePage>[Title ·Page]</titlePage>↵
7      ········<div type="preface">[Preface]</div>↵
8      ········</front>↵
9    ······<body>↵
10     ········<div n=" [number] ">[Chapter]</div>↵
11     ········</body>↵
12  ······<back>↵
13     ········<div type="postface">[Postface]</div>↵
14     ········<div type="index">[Index]</div>↵
15     ········<div type="appendix">[Appendix]</div>↵
16     ········</back>↵
17  ····</text>↵
18 </TEI>↵
```

TEI: Grundlegende Dokumentstruktur II (vereinfacht)

```
1  <?xml version="1.0" encoding="UTF-8"?>↵
2  <body>↵
3    ····<pb n="[page number]" facs="[number of facsimile]"/>↵
4    ····<fw>[column title]</fw>↵
5  ▾ ····<div n="[number]">↵
6      ······<head>[title]</head>↵
7      ······<p>[prose paragraph]</p>↵
8      ······<p>[prose paragraph] <table>[table]</table></p>↵
9      ······<fw>[signature/catch word]</fw>↵
10     ······<pb n="[page number]" facs="[number of facsimile]"/>↵
11     ······<p>[prose paragraph]</p>↵
12     ····</div>↵
13 ▾ ····<div n="[number]">↵
14     ······<head>[title]</head>↵
15 ▾ ······<lg type="poem">↵
16     ········<l>[poem verse]</l>↵
17     ········<l>[poem verse]</l>↵
18     ······</lg>↵
19     ······<figure>[illustration]</figure>↵
20     ····</div>↵
21 </body>↵
```

Das TEI Schema

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <?xml-model
2 href="http://www.tei-c.org/release/xml/tei/custom/schema/relaxng/
2 tei_all.rng" type="application/xml"
2 schematypens="http://relaxng.org/ns/structure/1.0"?>
3 <body>
4   ... <pb n="[page number of facsimile]"/>
5   ... <fw>[column title]</fw>
6   ... <div n="[number]">
7     ... <head>[title]</head>
8     ... <p>[prose paragraph]</p>
9     ... <p>[prose paragraph] <table>[table]</table></p>
10    ... <fw>[signature/catch word]</fw>
11    ... <pb n="[page number]" facs="[number of facsimile]"/>
12    ... <p>[prose paragraph]</p>
13  ... </div>
14  ... <div n="[number]">[Chapter]</div>
15 </body>
```

Validierungsproblem

Annotationsübung Ia: Tagging in TEI (gestrige Beispielseite)

TEI und Interoperabilität

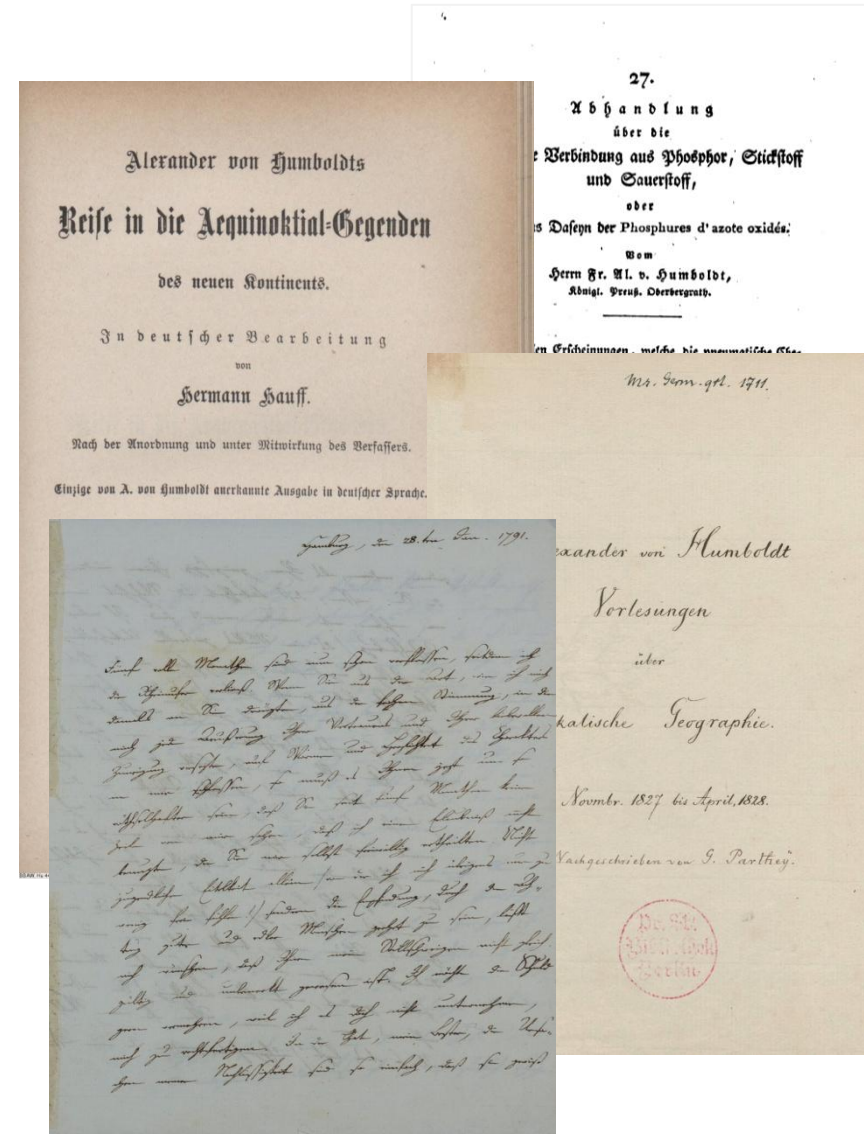
- Wenn jeder nach den TEI/P5-Richtlinien auszeichnet: Erhalten wir interoperable Texte?

Oder:

- Sind TEI-Texte per se interoperabel? (Unsworth 2011)
- Interoperabilität:
 - Möglichkeit, Texte mit denselben Methoden und Tools auszuwerten mit verlässlichen Resultaten

Bsp. Vernetzung von Daten – AvH im DTA

1. Werke AvHs im DTA-Kernkorpus
2. Projekt der HU Berlin: Nachschriften zu Humboldts Kosmos-Vorlesungen
3. DTAE-Projekt: Aufbereitung der unselbständigen Schriften AvHs
4. BBAW-Vorhaben: Alexander von Humboldt auf Reisen



Bsp. Vernetzung von Daten – AvH im DTA

8:	[dta:humboldt_aequinoktial04_1859:163]	Man erzählt rührende Fälle, wo	afrikanische	Sklaven ihr Leben aufs Spiel setzten, um...
9:	[dta:humboldt_aequinoktial04_1859:191]	... Demerary Feuergewehre, und als der Handel mit	amerikanischen	Sklaven (Poitos) in Blüte stand,...
10:	[dta:humboldt_aequinoktial04_1859:253]	... frei zur Welt kommen, nicht einer des	andern	Sklaven werden.“
11:	[dta:humboldt_aequinoktial04_1859:296]	... auf dem Festlande von Spanisch-Amerika die Zahl der	afrikanischen	Sklaven so unbedeutend, daß sie sich zur...
12:	[dta:humboldt_aequinoktial01_1859:203]	... Sklavenhandel abge schafft hat, und dennoch waren die	ersten	Sklaven, die wir aufgestellt haben, auf...
13:	[dta:humboldt_aequinoktial01_1859:203]	Die zum Verkauf	ausgesetzten	Sklaven waren junge Leute von fünfzehn bis zwanzig...
14:	[dta:humboldt_aequinoktial01_1859:204]	Der Handel mit	afrikanischen	Sklaven, den die spanischen Gesetze niemals begünstigt...
15:	[dta:humboldt_aequinoktial01_1859:308]	... behelfen sich bei ihrer mühsamen Arbeit mit einem	einzigen	Sklaven; sie brechen mit eigener Hand den...
16:	[dta:humboldt_grenzen_1851:133]	...; allein er würde mir immer ein Haufe	ernährter	Sklaven, nicht eine Vereinigung freier, nur...
17:	[dta:humboldt_aequinoktial02_1859:75]	... fein Wohlstand gehoben hatte, kaufte er zwei	schwarze	Sklaven, und dies ward fein Verderben:
18:	[dta:humboldt_aequinoktial02_1859:186]	... wir drei entprungene Neger einbringen, vor kurzem	gekaufte	Sklaven.
19:	[dta:humboldt_aequinoktial02_1859:207]	... und Indigo erzeugen kann, und daß die	unglücklichen	Sklaven Bauern, Pächter und Grundbesitzer werden können...
20:	[dta:humboldt_aequinoktial02_1859:229]	In dieses Bad werfen sich die	unglücklichen	Sklaven, wenn sie gegen Sonnenuntergang, mit...
21:	[dta:humboldt_aequinoktial02_1859:266]	... mit der Förderung und dem Schmelzen des Erzes	befähigten	Sklaven beträgt nur 60 bis 70.
22:	[dta:humboldt_reizmittel_1797:19]	... man, einem alten Herkommen gemäß, die	toten	Sklaven, ehe man sie ins Meer senkt...
23:	[dta:humboldt_verhaeltnis_1826:2]	... dem Festlande vom spanischen Amerika die Zahl der	afrikanischen	Sklaven verhältnißmäßig zur Sklavenbevölkerung von Brasilien oder vom...
24:	[dta:humboldt_verhaeltnis_1826:2]	... pflanzt und die fast keine dem afrikanischen Gebiet	entriffene	Sklaven hat.

DDC: "\$p=ADJA Sklave" #has[author, /Humboldt/]

TEI und Interoperabilität

- TEI bietet unterschiedliche Markup-Optionen für ähnliche Phänomene
z.B. Eigennamen

TEI

```
<rs type="propNounPersName">Sokrates</rs>,  
<name type="person">Platon</name> und  
<persName>Aristoteles</persName> waren Philosophen.
```

- Warum?
 - TEI als ein “set of building blocks for creating a schema suitable for a particular project”
 - Philosophie, Projekten Angebote, nicht Vorschriften für das Tagging zu machen

Cf. <http://www.tei-c.org/Guidelines/Customization/odds.xml>

- Format für die Auszeichnung von historischen Texten in Korpora
- Echte Teilmenge von TEI-P5
- Datengrundlage: DTA-Korpora
 - 4.247 Texte des 17.-19. Jh.
 - unterschiedliche Textsorten und Disziplinen
 - breite Datenbasis
- Leitgedanken:
 - Identisches Tagging für semantisch ähnliche Phänomene
 - Homogene Textstrukturierung des gesamten Korpus
 - Möglichst hohe Abdeckung von Phänomenen in geschriebenen historischen Texten
 - Vermeidung von Ambiguitäten (Interoperabilität gewährleisten)
 - Dokumentation von Entscheidungen

TEI

```
<rs type="propNounPersName">Sokrates</rs>,  
<name type="person">Platon</name> und  
<persName>Aristoteles</persName> waren Philosophen.
```

DTABf

```
<persName>Sokrates</persName>,  
<persName>Platon</persName> und  
<persName>Aristoteles</persName> waren Philosophen.
```

Schema zur Validierung der XML-Instanzen

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-model href="http://www.deutschestextarchiv.de/basisformat.rng"
type="application/xml" schematypens="http://relaxng.org/ns/structure/1.0"?>
<?xml-model href="http://www.deutschestextarchiv.de/basisformat.sch"
type="application/xml" schematypens="http://purl.oclc.org/dsdl/schematron"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0">
<teiHeader>
<fileDesc>
<titleStmt>
<title type="main">Von wahrem Christenthumb</title>
```

```
mut/ Gedult/ Gottesfurcht. Sihe das<lb/>
heilige gebet des <hi rendition="#cap">Herr</hi>n das Vater<lb/>
vnfer an. Daffelbe gehet wider dich vnd<lb/>
fu&#x0364;r dich. Sol Gottes Name allein<lb/>
```

```
mut/ Gedult/ Gottesfurcht. Sihe das<lb/>
heilige gebet des <hi rendition="#k">Herr</hi>n das Vater<lb/>
vnfer an. Daffelbe gehet wider dich vnd<lb/>
fu&#x0364;r dich. Sol Gottes Name allein<lb/>
```

Annotationsübung Ib:

Tagging in DTABf (gestrige Beispielseite)

DTABf: Was wird ausgezeichnet

- **Formale (Layout-) Textauszeichnung**
(Spaltensatz, Abbildungen und Trennlinien, Formeln, Tabellen, Listen, Seiten-/Zeilenumbrüche, ...)
- **Inhaltliche (Logische, Konzeptuelle) Textauszeichnung**
(Titelblätter, Kapitel, Gedichte, Sprechakte im Drama, Erläuterungen/Metatext, Zitate, ...)
- **Auszeichnung von Metadaten**
(Aufbewahrungsort der Quelle, Signatur, Autor, Erscheinungsjahr, Textsorte, Sprache, Schriftart, ...)
- **Richtlinien zur Transkription**
(Nähe zur Quelle, Umgang mit Ligaturen, Sonderzeichen, Interpunktion, ...)

Über das Tagset

Tagset: Echtes Subset der TEI

TEI: **552** Elemente + Attribute

DTABf: **129** Elemente + Attribute und Werte
 83 im Textbereich (<text>)
 (davon 7 nur in Manuskripten, 76 in Drucken)
 46 im Headerbereich (<teiHeader>)

Über das Tagset

- Festlegungen im DTABf nicht nur für Elemente, sondern auch für Attribute und deren Werte
 - *Element*: übergeordnete Semantik
 - *Attribute und Werte*: nähere Spezifikation der Elemente
- **Attribute**: Beispiel Anmerkungen
 - Element:
`<note>`
 - Attribute:
 - „lokale“ Attribute (= TEI P5): `@place, @type, @n`
 - globale Attribute (= TEI P5): `@xml:id, @sameAs, @corresp, @prev, @next, @rendition`
 - nicht gültig (\leftrightarrow TEI P5): `@cert, @resp, @subtype, @evaluate, @facs, @ana, @targetLang, @target, ...`

DTABf: Tagset

- **Werte:** drei Möglichkeiten
 - Wert-Auswahl je Attribut festgelegt (je Element oder Klasse)

Beispiel `<note>`:

`@place="left|right|end|foot"`

`@type="editorial"`

- Attribut festgelegt, Datenstruktur des Wertes festgelegt

Beispiel `<lg>`:

`@n="[data.count]"`

- keine Festlegung beim Wert (wenn unabdingbar)

Beispiel `<note>` :

`@n` (für das Fußnotenzeichen)

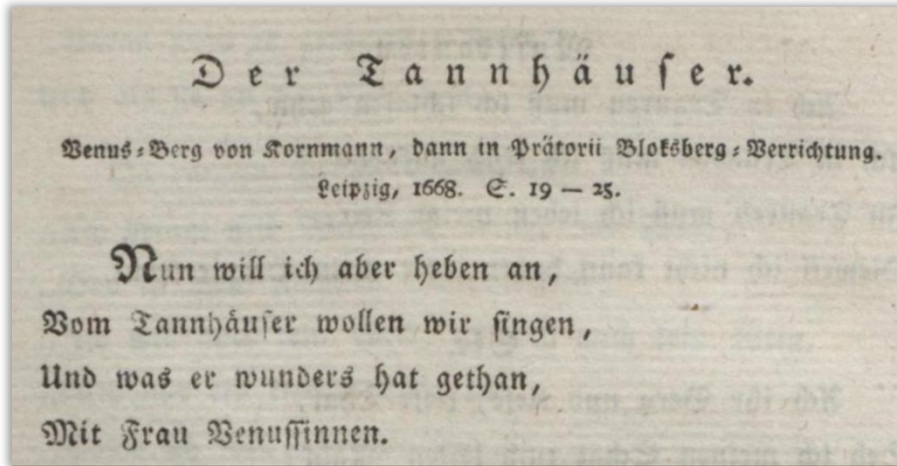
Tagging nach dem DTABf



... haben in dero
Welt-<lb/>
<note place="left">
zu Mu^enchen
</note>beru^ehnten
Residenz ...

Sandrart, Joachim von: L'Academia Todesca. Bd. 1,1. Nürnberg, 1675.
http://www.deutschestextarchiv.de/sandrart_academie0101_1675

Tagging nach dem DTABf



Nun will ich aber heben an,
Vom Tannhäuser wollen wir singen,
Und was er wonders hat gethan,
Mit Frau Venussinnen.

```
<lg type="poem">  
  <lg n="1">  
    <l>Nun will ich aber heben an,</l><lb/>  
    <l>Vom Tannhäuser wollen wir singen,</l><lb/>  
    <l>Und was er wonders hat gethan,</l><lb/>  
    <l>Mit Frau Venussinnen.</l>  
  </lg><lb/>  
</lg><lb/>
```

Arnim/Brentano: Des Knaben Wunderhorn. Bd. 1. Heidelberg, 1806, S. 86,
http://www.deutschestextarchiv.de/arnim_wunderhorn01_1806/95

DTABf-Metadaten

```
<teiHeader>
  <fileDesc>
    <titleStmt>
      <title type="main">Des Knaben Wunderhorn</title>
      <title type="sub">Alte deutsche Lieder</title>
      <title type="volume" n="1"></title>
      <author>
        <persName ref="http://d-nb.info/gnd/118504177">
          <surname>Arnim</surname>
          <forename>Achim von</forename>
        </persName>
      </author>
      <author>
        <persName ref="http://d-nb.info/gnd/118515055">
          <surname>Brentano</surname>
          <forename>Clemens</forename>
        </persName>
      </author>
      [...]
    <availability>
      <licence target="http://creativecommons.org/licenses/by-sa/4.0/">
        <p>Distributed under the Creative Commons
          Unported License.</p>
      </licence>
    </availability>
    [...]
    <publicationStmt>
      <publisher><name>Mohr u: Zimmer</name>
      <pubPlace>Heidelberg</pubPlace>
      <pubPlace>Frankfurt</pubPlace>
      <date type="publication">1806</date>
    </publicationStmt>
    [...]
```



Arnim, Achim von; Brentano, Clemens:
Des Knaben Wunderhorn. Bd. I.
Heidelberg, 1806.

BIBLIOGRAPHISCHE ANGABEN

URN:	urn:nbn:de:kobv:b4-20090519157
Titel:	Des Knaben Wunderhorn
Untertitel:	Alte deutsche Lieder
Autor/innen:	Achim von Arnim (GND, Wikipedia, ADB/NDB) Clemens Brentano (GND, Wikipedia, ADB/NDB)
Erscheinungsjahr:	1806
Verlag:	Mohr u: Zimmer
Ort:	Heidelberg; Frankfurt
Band:	1 von 3
Auflage:	1. Auflage
Bibliothek:	Staatsbibliothek zu Berlin – Preußischer Kulturbesitz
Signatur:	SBB-PK, 19 ZZ 2004-1 R

Dokumentation des Formats

D

T

A

Bf

EINFÜHRUNG METADATEN TRANSKRIPTION FORMAL INHALTLICH SPEZIAL ÜBERSICHTEN IMPRESSUM

DTA-BASISFORMAT

Suche

Das von CLARIN-D und der DFG empfohlene TEI-Format

Einführung

Zielsetzung des DTA-Basisformats, zugehörige Publikationen, Nutzung und erste Schritte

Metadaten

Erfassung (Transkription) und Auszeichnung von Metadaten

Formal

Auszeichnung von formalen Strukturen in Volltexten (Besonderheiten in Typographie und Layout)

Inhaltlich

Auszeichnung von inhaltlichen Strukturen (logischen, konzeptuellen, thematischen) und von Volltexten

DTA Bf

Suche

Home / Formal / Erscheinungsbild Text / Absätze

Absätze

Die Absatzstruktur wird aus der Vorlage übernommen. Die jeweiligen Absätze werden mittels `<p>`-Elementen gekennzeichnet.

Beispiel: Absatz

„Das wird ja ordentlich interessant,“ bemerkte Max Berner und stand auf, „da könnte ich am Ende noch hier für Fenia gegen irgend einen sibirischen Drachen zu Felde ziehen?“

Aber der Onkel teilte die heitere Stimmung nicht; seine Miene blieb so feierlich und besorgt wie zuvor.

```
<p>„Das wird ja ordentlich interessant,“ bemerkte Max Berner und stand auf, „da könnte ich am Ende noch hier für Fenia gegen irgend einen sibirischen Drachen zu Felde ziehen?“</p><p>Aber der Onkel teilte die heitere Stimmung nicht; seine Miene blieb so feierlich und besorgt wie zuvor.</p>
```

Quelle: Andreas-Salome, Lou: Fenitschka. Eine Ausschweifung. Stuttgart, 1898. [Faksimile 48]

Einrückungen der ersten Zeile eines Absatzes bleiben unberücksichtigt. Größere Abstände zwischen Absätzen werden nur dann ausgezeichnet, wenn diesbezüglich innerhalb eines Buches eine Varianz festzustellen ist. In diesem Fall wird in der betreffenden Zeile mittels `<space dim="vertical"/>` auf den signifikanten Abstand hingewiesen.

www.deutschestextarchiv.de/doku/basisformat

Dokumentation: Volltextsuche

D

T

A

Bf

EINFÜHRUNGMETADATENTRANSKRPTIONFORMALINHALTLICHSPEZIALÜBERSICHTENIMPRESSUM

Klammerung

Ergebnisse für: **Klammerung**

- Klammerungen in Tabellen
 - Zellen von Tabellen können in der Druckvorlage durch Klammerungen verbunden werden innerhalb des <cell>-Elements ausgedrückt. Den Hinweis auf die Klammerung erhält dabei der verbindende Tabelleneintrag, der in der Regel über mehrere Zeilen oder Spalten hinwegreicht.
- Klammerungen in Listen
 - Klammerungen von Textstücken treten häufig innerhalb von <item>-Element umschlossen...
- Listenähnliche Klammerungen zusammengefasst
 - Klammerungen können in Textstrukturen auf einer höheren Ebene Struktur auf. Sie werden daher innerhalb...
- Horizontale Klammerungen
 - Einen Spezialfall bilden Stammbäume, die in einer Zeile dargestellt werden (z.B. <cell>). Dabei erhält das <cell>...
- Grundstruktur der Kodierung von Klammerungen
 - Es kann vorkommen, dass Textausschnitte auf einer höheren Ebene angezeigt wird. In diesem Fall...
- Klammerungen

Home / Formal / Erscheinungsbild Text / Klammerungen / Klammerungen in Tabellen

Klammerungen in Tabellen

Zellen von Tabellen können in der Druckvorlage durch **Klammerung** miteinander verbunden und einem gemeinsamen Tabelleneintrag zugeordnet sein. Solcherlei **Klammerungen** werden innerhalb des <cell>-Elements ausgedrückt. Den Hinweis auf die **Klammerung** erhält dabei der verbindende Tabelleneintrag, der in der Regel über mehrere Zeilen oder Spalten hinwegreicht.

Vertikale Klammerung in einer Tabelle:

```
<table>
  <row>
    <cell>[Tabellenwert 1]</cell>
    <cell rows="2" rendition="#leftBraced">[verbindender Tabelleneintrag zu 1 und 2]</cell>
  </row>
  <row>
    <cell>[Tabellenwert 2]</cell>
  </row>
</table>
```

Horizontale Klammerung in einer Tabelle:

```
<table>
```

Dokumentation: Systematischer Einstieg



EINFÜHRUNG

Zielsetzung

Schema und Dokumentation

Publikationen

How to

Introduction (English)

METADATEN

TRANSKRIPTION

Grundstruktur jedes TEI-Dokuments

Das Wurzelement jedes TEI-Dokuments im DTA bildet das `<TEI>`-Element, dessen `@xmlns`-Attribut den TEI-Namensraum (namespace) spezifizieren, der für das jeweilige Dokument gilt:

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">...</TEI>
```

Jedes TEI-Dokument im DTA besteht aus einem Header, der Metadaten zum publizierten Text umfasst, und einem Textbereich, der alle Transkriptionen mit den zugehörigen Annotationen enthält. Dieser Volltext umfasst dabei nicht allein den eigentlichen Buchtext, sondern alle Textbestandteile, so auch Titelseite und Vorwort sowie in der Regel die Register, Beigaben und Anhänge.

Die folgende Struktur wird in jedem TEI-Dokument eingehalten:

```
<teiHeader>[Metadaten]</teiHeader>
<text>
  <front>[Elemente vor Beginn des Buchtextes]</front>
  <body>[Textkörper]</body>
  <back>[Elemente nach Abschluss des Buchtextes]</back>
</text>
```


Element	Beschreibung	Attribute/Anmerkungen	Funktionale Kategorie	Level
ab Doku: TEI	Anonymer Block/Container	<i>stattdessen</i> <code>div</code> , <code>floatingText</code> oder <code>p</code> benutzen	textStructure	4 (unzulässig)
abbr Doku: TEI, DTABf	Abkürzung		phraseStructure, editorial	3 (fakultativ)
actor Doku: TEI, DTABf	Name eines Schauspielers innerhalb einer Darstellerliste		drama, phraseStructure	2 (empfohlen)

corr Doku: TEI, DTABf	(bei Korrekturen:) korrigierte Form	@resp	Manuskripte: Person, die verantwortlich ist für die Ergänzung ["[pointer]"]	editorial	3 (fakultativ)
		@type	Ursprung der Korrektur ["addenda" , "corrigenda" , "editorial"]		
date Doku: TEI, DTABf	Datumsangabe	@when	Datum in ISO-8601-Form	phraseStructure	3 (fakultativ)

DTABf-Komponenten

1. Dokumentation:

<http://www.deutschestextarchiv.de/doku/basisformat>

2. ODD: <http://www.deutschestextarchiv.de/basisformat.odd>

3. Schema (RNG):

<http://www.deutschestextarchiv.de/basisformat.rng>

*Welche Elemente & Attribute der TEI sollen Teil des DTABf sein?
Welche Werte werden festgelegt?*

4. Schematron-Regeln: [~/basisformat.sch](#)

*Welche besonderen Regeln sollen für den Elementinhalt gelten?
Welche Beziehungen zwischen den Elementen sollen festgelegt werden? ...*

Des Autoris Gedancken in seiner letzten Krankheit.

Wenn Blut und Lüste schäumen/
So stärke meinen Geist/
Daß er sich auch im Träumen
Aus Satans Neze reiße.

Hilff für mein Bestes sorgen/
Verändere meinen Sinn/
Und mache/ daß ich morgen
Ein neu Geschöpfe bin.

Ich seh' das Licht verschwinden/
Die trübe Nacht bricht ein/
Ach Herr/ laß meine Sünden
Auch mit verschwunden seyn/
Streich' sie aus deinem Buche/
Das mich zum Schuldner macht/
Und rette mich vom Fluche/
Der mir war zugebracht.

Wenn heut mein Ziel der Jahre
Mein letzter Abend ist/
Wolan! wann ich nur fahre
Wo deine Wohnung ist.
(Wo du/ mein Vater/ bist.)
Doch soll ich länger leben/
So laß den festen Schluß
Mir stets für Augen schweben/
Daß ich einst scheiden muß!

A

Das

www.deutschestextarchiv.de/canitz_gedichte_1700/14

1. Textanalyse → Gedicht!

2. Recherche in der DTABf-Doku:

Home / Inhaltlich / Textkörper / Gedichte und gebundene Sprache / Gedichte in Gedichtbänden

Gedichte in Gedichtbänden

In Gedichtbänden können innerhalb von Gedichten neben dem Gedichttitel und den einzelnen Strophen zusätzliche Textbestandteile vorkommen, z.B. Datumszeilen, Sprecher oder zusätzliche Angaben zur Entstehungsgeschichte des Gedichts. In Gedichtbänden mit solcherlei Gedichten ebenso wie in literarischen Sammlungen verschiedener Textsorten wird die Grundstrukturierung mittels `<div>`-Elementen vorgenommen. Ein Gedicht wird dabei durch ein `<div>`-Element umschlossen; der Gedichttitel steht im `<head>` des umschließenden `<div>`-Elements. Datumszeilen, Sprecher, zusätzliche Angaben in einfacher Prosa werden wie gewohnt ausgezeichnet. Die Gedichtauszeichnung mittels `<lg>`s beginnt auf Strophenebene.

Auszeichnung von Gedichten in Gedichtbänden (beispielhaft):

```
<div type="poem">
  <head>[Titel]</head><!-- sofern vorhanden -->
  <dateline>[Datumsangabe]</dateline>
  <lg type="poem">
    <lg n="[Strophennummer]"><!-- sofern kein einstrophiges Gedicht -->
      <1>[Vers]</1>
      <1>[Vers]</1>
    </lg>
  </lg>
</div>
```

3. Annotation mithilfe des Schemas

Validierung der XML-Instanzen im oXygen

121 <lb/>
122 <l
123 <l> lg (line group) #x017F;chwinden/</l>
124 <lb> l contains one or more
125 <l> lb verse lines l>
126 <lb> functioning as a
127 <l> formal unit, e.g. a >
128 <lb> sterss-refrein
129 <l>
130 <lb/>
131 <l>Streichſchwinden/</l>
132 <lb/>
133 <l>Das mich
134 <lb/>
135 <l>Und rette
136 <lb/>
137 <l>Der mir w
138
139 <lb/>
140 <lg n="3">
122 <lg type
123 <l>Ich type characterizes the x017F;chwinden/</l>
124 <lb/> corresp element in some
125 <l>Die an sense, using any
126 <lb/> next convenient
127 <l>Ach prev classification
128 <lb/> rend scheme or typology.
129 <l>Auch rend
130 <lb/>
131 <l>Streichſchwinden/</l>
132 <lb/>
133 <l>Das mich
134 <lb/>
135 <l>Und rett
136 <lb/>
137 <l>Der mir
138
139 <lb/>
140 <lg n="3">
122 <lg type=""
123 <l>Ich ſchwinden/</l>
124 <lb/>
125 <l>Die ͤbe Nacht bricht ein/</l>
126 <lb/>
127 <l>Ach Herr/ laß meine Suͤnden/</l>
128 <lb/>
129 <l>Auch mit verſchwunden ſeyn/</l>
130 <lb/>
131 <l>Streichſchwinden/</l>
132 <lb/>
133 <l>Das mich zum Schuldner macht/</l>
134 <lb/>
135 <l>Und rette mich vom Fluche/</l>

Annotationsübung II: Tagging von Lyrik/Drama nach DTABf