

Winter School 2024

Bergische Universität Wuppertal

18.03 – 22.03.2024

Texttranskription mit TEI

Jana Klinger

Foto: Nadine Sutor



BERGISCHE
UNIVERSITÄT
WUPPERTAL



Dokument
Text
Edition
Graduiertenkolleg 2196

IZ
ED

Interdisziplinäres
Zentrum für
Editions- und
Dokumentwissenschaft



Worum geht es?

1. Wie sind TEI Dokumente grundsätzlich aufgebaut?
2. Welche Möglichkeiten bietet die Arbeit mit TEI?
3. Was sind häufig genutzte Elemente der TEI?
4. Wie kann man mit den TEI Guidelines arbeiten?

1. Wie sind TEI Dokumente grundsätzlich aufgebaut?

Die Grundstruktur

```
<?xml version="1.0" encoding="UTF-8"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <!-- Metadaten zum Dokument -->
  </teiHeader>
  <text>
    <!-- Der eigentliche Inhalt -->
  </text>
</TEI>
```

1. Wie sind TEI Dokumente grundsätzlich aufgebaut?

Der <teiHeader>

```
<TEI xmlns="http://www.tei-c.org/ns/1.0" >
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>Beispieldatei Winter School 2024 </title>
      </titleStmt>
      <publicationStmt>
        <p>Unveröffentlicht, nur für Demonstrationszwecke. </p>
      </publicationStmt>
      <sourceDesc>
        <p>Born digital</p>
      </sourceDesc>
    </fileDesc>
  </teiHeader>
  <text>
    <!-- ... -->
  </text>
</TEI>
```

... lässt sich unendlich erweitern und verkomplizieren.

1. Wie sind TEI Dokumente grundsätzlich aufgebaut?

Der <teiHeader>

- <fileDesc> **Dateibeschreibung**
 - <titleStmt> **Angaben zum Titel**
 - <publicationStmt> **Angaben zur Veröffentlichung**
 - <sourceDesc> **Beschreibung der Quellen**
- <encodingDesc> **Beschreibung der Kodierung**
- <profileDesc> **Beschreibung des Textprofils**
- <revisionDesc> **Beschreibung der Dateihistorie**

1. Wie sind TEI Dokumente grundsätzlich aufgebaut?

Der <text>-Teil

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <!-- ... -->
  </teiHeader>
  <text>
    <front>
      <!-- front matter of copy text, if any, goes here -->
    </front>
    <body>
      <!-- body of copy text goes here -->
    </body>
    <back>
      <!-- back matter of copy text, if any, goes here -->
    </back>
  </text>
</TEI>
```

2. Welche Möglichkeiten bietet die Arbeit mit TEI?

- Quellenbeschreibung
 - Metadaten
- Textbeschreibung
 - Strukturell, topographisch, typologisch
- Textannotation
 - Editionsmethoden

Siehe Roeder, Torsten: Textauszeichnung mit den TEI-Kernelementen. Folien zu IDE Winter School 2020, Folie 2, https://www.i-d-e.de/wp-content/uploads/2020/02/05_TEI_Core.pdf

3. Was sind häufig genutzte Elemente der TEI?

- Kapitel 3 der Guidelines ist für Anfänger:innen ein guter Startpunkt
- Dort sind die Kernelemente vermerkt, die in allen TEI-Dokumenten genutzt werden können.
- Von dort aus können je nach Projektzielen oder Forschungsfragen Elemente und Attribute aus weiteren Kapiteln ergänzt werden.

3. Was sind häufig genutzte Elemente der TEI?

Einbinden von Facsimiles

- `<facsimile>` enthält Verweise auf Bilddateien
- wird zwischen `<teiHeader>` und `<text>`-Element positioniert

```
<TEI>
  <teiHeader>...</teiHeader>
  <facsimile>
    <graphic url="page01.jpg" />
    <graphic url="page02.jpg" />
    <graphic url="page03.jpg" />
    ...
  </facsimile>
  <text>...</text>
</TEI>
```

3. Was sind häufig genutzte Elemente der TEI?

<div>-Elemente

- <front>, <body> und <back> lassen sich mit <div> weiter unterteilen
- <div> *division* → analytische Zugangsweise

```
<body>
  <div type="chapter" n="1">
    ...
  </div>
  <div type="chapter" n="2">
    ...
  </div>
  ...
</body>
```

3. Was sind häufig genutzte Elemente der TEI?

Das <p>-Element

- <p> steht für *Paragraph*
- Wird genutzt, um Absätze als Einheit zu kennzeichnen

```
<text>
  <body>
    <p>Dies ist ein Satz. Ein paar Sätze
      gefolgt von einem Zeilenumbruch bilden
      zusammen einen Absatz.</p>
  </body>
</text>
```

- Mögliche Alternativen: <s>, <seg>, <ab>

3. Was sind häufig genutzte Elemente der TEI?

Milestones

- Bei Umbrüchen wird der Beginn einer neuen Seite/Spalte/Zeile markiert.
- Die Markierung erfolgt mit einem leeren Element
 - Verhindert, dass die XML-Baumstruktur durcheinandergerät.

- `<pb/>` *page beginning*
- `<cb/>` *column beginning*
- `<lb/>` *line beginning*
- `<milestone/>` generische Option

...

```
<page>
```

...

```
<p>Dies ist ein Satz. Ein paar  
Sätze gefolgt von  
</page><page>einem  
Zeilenumbruch bilden zusammen  
einen Absatz.
```

```
</p>
```

...

3. Was sind häufig genutzte Elemente der TEI?

Hervorhebungen

- `<hi>` steht ganz allgemein für *highlighted* ("irgendwie anders")
 - `<emph>` *emphasized*
 - Beides wird gern kombiniert mit
 - `@rend` *rendition* (beschreibt die Hervorhebung)
 - z.B. *italic*, *bold*, *superscript* etc.
- `<head>` Überschrift
 - Muss das erste Kind-Element sein

```
<div type="chapter" n="1">  
  <head>Überschrift Kapitel 1</head>  
  <p>Dies ist der erste Satz eines Kapitels. Ein paar Sätze  
    gefolgt von einem Zeilenumbruch bilden zusammen einen  
    Absatz.</p>  
</div>
```

Siehe Roeder, Torsten:
Textauszeichnung mit den
TEI-Kernelementen. Folien zu IDE
Winter School 2020, Folie 8,
https://www.i-d-e.de/wp-content/uploads/2020/02/05_TEI_Core.pdf

3. Was sind häufig genutzte Elemente der TEI?

Siehe Bleier, Roman: TEI-Einführung
2. Folien zu IDE Summer School
2020, Lübeck/Online Folie 13-15.

Texteingriffe

- Durch die Autor:innen
 - `<add>` (Hinzufügung) → @place, z.B. above, below
 - `` (Tilgung) → @rend, z.B. strikethrough
 - `<subst>` (Ersetzung)
- Durch die Editor:innen
 - `<choice>` (Alternative)
 - `<sic>` und `<corr>` (Korrekturen, seltsame Schreibung)
 - `<orig>` und `<reg>` (Normalisierung)
 - `<abbr>` und `<expan>` (Abkürzung)
 - `<gap>` (Auslassung)
 - `<unclear>` (unklare Stelle)
 - `<note>` (Anmerkung)

```
<choice>  
  <abbr>Rd</abbr>  
  <expan>Road</expan>  
</choice>
```

3. Was sind häufig genutzte Elemente der TEI?

Namen

- `<rs>` *referencing string*
- `<name>` Eigenname
 - Speziellere Elemente in Kapitel 13, „Names, Dates, People, and Places“ → dazu morgen mehr!

“My dear `<name>`Mr. Bennet`</name>`,” said `<rs>`his lady`</rs>` to him one day, “have you heard that `<name>`Netherfield Park`</name>` is let at last?”

3. Was sind häufig genutzte Elemente der TEI?

Siehe Bleier, Roman: TEI-Einführung
2. Folien zu IDE Summer School
2020, Lübeck/Online, Folie 16.

Weiteres

- Listen: `<list>` als Container, `<item>` für Einträge
- Verweise: `<ptr>` und `<ref>` mit `@target`
- Zahlen und Maßangaben: `<num>`, `<measure>`
- `<l>` für Verse, `<lg>` für Versgruppen
- Bibliographie:
 - Bibliographischer Eintrag: `<bibl>`, `<title>`, `<author>`, `<series>` ...
 - Liste von Einträgen: `<listBibl>`

4. Wie kann man mit den TEI Guidelines arbeiten?

- <https://tei-c.org/release/doc/tei-p5-doc/en/html/index.html>

```
<q>My dear <name>Mr. Bennet</name>, </q> said  
<rs>his lady</rs> to him one day, <q>have you  
heard that  
<name>Netherfield Park</name> is let at  
last?</q>
```

Vielen Dank für Ihre Aufmerksamkeit!



BERGISCHE
UNIVERSITÄT
WUPPERTAL



BERGISCHE
UNIVERSITÄT
WUPPERTAL