

XML, TEI & and. Abkürz.

Zu verwirrenden Begriffen für Editorinnen und Editoren
im digitalen Zeitalter

{Warnhinweis}

- Alle Punkte, die in geschwungenen Klammern {} stehen, dienen der Täuschung und Verwirrung des Auditoriums. Sie sollen nur dann angehört werden, wenn zumindest eines der folgenden Gegenmittel zugegen ist:
 - Ohrstöpsel (Ohr-O-Pax)
 - Schläfrigkeit (Schlaftabletten sind im Vortrag integriert)
 - Tomaten und Eier (können als Notbremse für den Vortrag dienen)
 - Deutsche Staatsbürgerschaft

Outline {inlined}

- Digitale Edition, was ist das?
- {Datenmodellierung: Klärung eines Begriffs}
- TEI: Guideline of the Text Encoding Initiative
 - XML: Grundlage eines notwendigen Übels.
 - Wichtige Elemente: Sirenen für Editorinnen und Editoren auf der Odyssee
- {Weitere Anwendungen: Volltextsuche, Elektronische Datenverarbeitung, Verlinkungen, Linked Data}

I. Digitale Edition

Geldbeschaffung oder
-verschwendung?

Edition

- Was ist eine Edition?
 - Veröffentlichung von Quellenmaterialien? Lesehilfe? Recherchierhilfe? Sammlung von rektifizierten Texten?
 - Unterschiedliche Arten: Textkritisch? Kommentiert? Leseausgabe? Faksimile-Ausgaben? Regesten-Ausgabe?
 - Für wen wird ediert? Wofür sollte die Edition dienen?
 - {Wer finanziert die Edition? Wer verdient Geld mit ihr? Wessen Ruhm wird vermehrt durch sie?}

Editionspraxis

- Schematische Darstellung, wie man früher eine Edition hergestellt hat:
 - Materialien sammeln und erschließen
 - Editorische Richtlinie erstellen
 - Transkribieren (nach der Richtlinie)
 - Kommentieren (nach der Richtlinie)
 - Editorischen Bericht, Vorwort etc. schreiben
 - Druckfahnen erstellen
 - Fahnen korrigieren und Register erstellen
 - Edition drucken

{„Digitale Edition“}

- Praxis des digitalen Edierens = Praxis des Edierens mittels Computer?
 - Materialien sammeln und erschließen (mit Computer)
 - Editorische Richtlinie erstellen (mit Computer)
 - Transkribieren (nach der Richtlinie) (mit Computer)
 - Kommentieren (nach der Richtlinie) (mit Computer)
 - Editorischen Bericht, Vorwort etc. schreiben (mit Computer)
 - Druckfahnen erstellen (mit Computer)
 - Fahnen korrigieren und Register erstellen (mit Computer)
 - Edition drucken (mit Computer)

Digitale Edition

- Wodurch zeichnen sich die digitalen Editionen aus?
 - Erschließung der Materialien mittels Computer
 - Edieren mit „digitalen Werkzeugen“ und Erstellung digitaler Datenmodelle
 - Trennung zwischen Daten und Metadaten
 - [Präsentation als Webseiten]
 - [Präsentation im digitalen Medium]
 - [Verlinkung der Daten als Hyperlinks]
 - [Verlinkung der Daten durch Semantic Web]

{2. Daten-Modellierung}

„Wie bitte, die Herausgeberinnen modellieren schon seit Jahrhunderten Daten?“

Modelle

- Modelle sind „Abbildung“ bestimmter Objekte nach bestimmten Gesichtspunkten.
- Lego-Feuerwehrauto sind Modelle von echten Feuerwehrautos.
- Tiermodelle für erkrankte Menschen (z.B. LDL-Rezeptor-Knockout-Mäuse für Menschen mit Arteriosklerose)
- Zwei Einschränkungen
 - Modelle bilden nur unter bestimmten Rücksichten das Original nach!
 - Modelle unterliegen auch der Materialursache des „Substrats“.

Editionen als Modelle

- Editionen kann man auch als Modelle des Originals auffassen.
 - Der Text ist dabei meist die Rücksicht, unter der das Modell gebildet wird.
 - Print-Editionen sind Modelle des Originals als Buch. Und so unterliegen sie auch der Materialursache des Buches und können z.B. das Alter des Originals und Überarbeitungen nur schwer abbilden. Dafür wird der Text als Print-Text gut modelliert und die Lesbarkeit wird stark erleichtert.
 - Digitale Editionen sind Modelle des Originals als digitale Daten. Sie unterliegen der Materialursache der digitalen Daten. Das Papier kann überhaupt nicht adäquat abgebildet werden, dafür können Überarbeitungen gut abgebildet werden. Außerdem können sie leicht über das Internet verschickt werden.

3. XML und TEI

w. l. abk.

eXtensible Markup Language

- {Mathematische Grundlage: Graphentheorie und Baumstruktur}
- Aufbau:
 - Datenstruktur: XML-Prologue + {Processing-Instructions + Comments +} | XML-Element
 - XML-Element: Start-Tag + XML-Attributes + (XML-Element | Text) + End-Tag
 - XML-Attribute werden in Starttags geschrieben und haben die Form key="value". Die Reihenfolge der Attribute spielt keine Rolle
 - Beliebig aufeinander folgende Leerzeichen (auch \r \t und \n) zählen zusammen als ein Leerzeichen.
 - {Text-Kodierung: UTF-8 oder UTF-16}

XML Einschränkungen

- Manche Zeichen müssen als „XML-Entity“ geschrieben werden.
- Entity kann man prinzipiell für alle Zeichen verwenden
- Überlappende Elemente sind nicht erlaubt. Hier muss man entweder Elemente verdoppeln oder mit Milestones arbeiten
- Im Prinzip sind alle Daten als XML-Daten formulierbar

Notwendige XML-Entities

,<' = ,<' ,>' = ,>'
,&' = ,&'
,'" = ,'' ,'" = ,"'

Überlappende Elemente

```
<fett>nur fett <kursiv> fett & kursiv</  
fett> nur kursiv</kursiv>
```

Auflösung durchs Verdoppeln

```
<fett>nur fett <kursiv> fett & kursiv</  
kursiv></fett><kursiv> nur kursiv</kursiv>
```

Text Encoding Initiative

- TEI ist eine Richtlinie, wie man geisteswissenschaftliche Daten in XML kodiert.
- TEI verwendet „sprechende Elementennamen“, damit die Datei auch ohne spezielle Programme von Menschen gelesen werden kann.
- Aber v.a. zielt TEI darauf, diese Daten maschinenlesbar zu machen, damit der Computer mit diesen Daten „rechnen“ kann.

<TEI>

- <TEI>
 - <teiHeader>
 - <facsimile>
 - <text>
- Trennung der strukturellen Daten (in <teiHeader>) von den semistrukturellen Daten (in <text>)

<teiHeader>

- <fileDesc>
- <titleStmt>
 - <title>
 - <author>
- <publicationStmt>
- <sourceDesc>
- <encodingDesc>
- <revisionDesc>

<text>

- <front>, <body>, <back>
 - <div>
 - <head>
- <p>, <table> mit <row> & <cell>, <list> mit <item>
 - <hi>
 - <formula>, <graphic>
 - <note>
 - <ref>

Semantische Auszeichnungen

- TEI bietet auch Elemente an, die für semantische Auszeichnungen verwendet werden.
- `<rs>`, `<persName>`, `<placeName>`, `<date>`
- Verwendungsweise: der Verweis wird eingeklammert und mit dem `@key-` oder dem `@ref-`Attribut wird die Person oder die Ortschaft genannt, auf die verwiesen wird.

Textapparat

- `<app>`
 - `<rdg>`
 - `<lem>`
 - `<add>`, ``, `<gap>`

4. Anwendungsmöglichkeiten

Wofür machen wir das überhaupt?

Suchen & Sortieren

- Suche in Metadaten
 - Beispiel: „Manuskripte von N.N.“, „Briefe zwischen ### und ###“
 - Diese Suche werden in entsprechende technische Sprachen (SQL, XPath, XQuery) übersetzt.
- Suche im Text
 - Suche nach Zeichenketten (MS-Word®-Funktion „Suche
 - Suche im Index: Google, Lucene
 - Lemmatisierte Suche: Suche nach Worten inklusiven ihren Flexionen.

Suchen & Sortieren

- Die Such-Funktion ist der augenscheinliche Mehrwert digital verfügbarer Daten.
- N.B.: Die Suche am Computer kann fehlschlagen.
 - Computer findet oft zu viele Daten (falsch positive Treffer, vgl. Anzahl der Google-Treffer)
 - Computer findet das nicht, wonach man suchen will (falsch negative Treffer!)
- Sortieren nach unterschiedlichen Sprachen

{Elektronische Datenverarbeitung}

- Beispiele für EDV in Geisteswissenschaften:
 - Automatische Generierung von Registern
 - Multiple Repräsentation
 - Visualisierung von Daten
 - Statistische Auswertung
- Stylometrie

{Verlinkungen}

- Neben der Suche erkennt man den Mehrwert einer digitalen Edition am Verlinken einer Edition mit anderen Editionen bzw. Webressourcen.
- Zwei Arten:
 - Verweise auf Webseiten (z.B. beim Quellennachweis) oder „Anker-Punkte“ (bei Verweise im Text)
 - Verweise auf Web-Ressourcen mittels Semantic Web Technology (RDF und OWL)

{Semantic Web}

- Haupttechnologien:
 - Resource Description Framework (RDF)
 - Web Ontology Language (OWL)
- Idee
 - RDF: Modellieren der Daten als Objekte und Relationen zwischen Objekten.
 - OWL: Die Objekte werden zusätzlich noch in ihrer Klassen, die Relationen in ihren logischen Beziehungen gefasst.
 - Daraus ergeben sich Abfrage-Möglichkeiten, die über Metadaten-Abfrage hinausgehen.

{Semantic Web}

- Beispiel:
 - *Herr A* schreibt *Frau B* den *Brief M* am *Tag X*.
- Daraus ergeben sich folgende RDF-Tripel
 - M hatAutor A.
 - M hatAdressat B.
 - M istDatiertAuf X.

Und das war's...

{.}