Einstieg in DH-Methoden
Textauszeichnung in TEI/XML

Abb.: Susanne Schork

Agnes Brauer, Universitätsbibliothek Johann Christian Senckenberg
a.brauer@ub.uni-frankfurt.de

Das bin ich

- Verantwortlich für das Sachgebiet *Digitalisierungsprojekte und Digital Humanities* an der UB Frankfurt
- in die Digital Humanities im Studium „hineingewachsen“ durch studentische Mitarbeit an Editionsprojekten
- Streiterin für offene und nachhaltige digitale Formate

Digital Humanities – an der UB Frankfurt

UNIVERSITÄTS BIBLIOTHEK FRANKFURT AM MAIN

Eine Plattform für digitale Editionen wäre toll!

Könnten Sie uns bitte helfen, einen Editionstext nach TEI/XML zu migrieren?

Könnten Sie auch mal was zur Text- und Datenvisualisierung machen?

Ich brauche eine Python-Einführung für Geisteswissenschaftler*innen.

Abbildung Sprechblasen: https://www.gimp-handbuch.de/gimp-pfade/gimp-sprechblase-erstellen-1

Was wir im Bereich der DH machen

Praxislabor Digital Humanities (seit 2018)

- Regelmäßige Workshops zum Kennenlernen von Methoden und Werkzeugen der Digital Humanities für Studierende und Mitarbeiter/-innen der Goethe-Universität

- Bisherige Themen (in Auswahl):
  - X-Technologien (XML/TEI, XPath, XSL)
  - Python
  - Textvisualisierung
  - Korpusanalyse
  - Georefenzierung
  - Wikidata
  - Git
  - Python
  - R
  - …

https://www.ub.uni-frankfurt.de/digitalhumanities/workshops.html

Was wir im Bereich der DH machen

Digitale Editionen

- Unterstützung bei Editionsvorhaben in Fragen der einzusetzenden Technologien und der Wahl geeigneter Plattformen sowie Präsentationsformen

- Dies umfasst die:
  - Beratung bei der Planung eines Forschungsantrags
  - Auswahl geeigneter Tools und Werkzeuge
  - Schulung der Projektmitarbeiter*innen in dem Auszeichnungsstandard XML/TEI
  - Einführung in das Arbeiten mit dem XML-Editor Oxygen
  - Einführung in die Grundlagen der automatischen Texterkennung

Was wir im Bereich der DH machen

Individuelle Schulungs- und Beratungstätigkeit (on demand) zu Tools und Methoden der Digital Humanities für Wissenschaftler*innen der Goethe-Universität

Beispiele aus dem Portfolio:

- Textauszeichnung in TEI/XML
- X-Technologien (XPath, XSLT)
- Datenbereinigung mit Openrefine
- Texterkennung mit Transkribus
- Textanalyse-/Textvisualisierungstools

Daten - Formate und Strukturen

UNIVERSITÄTS BIBLIOTHEK FRANKFURT AM MAIN

Setzen Sie auf Plain text-Formate!

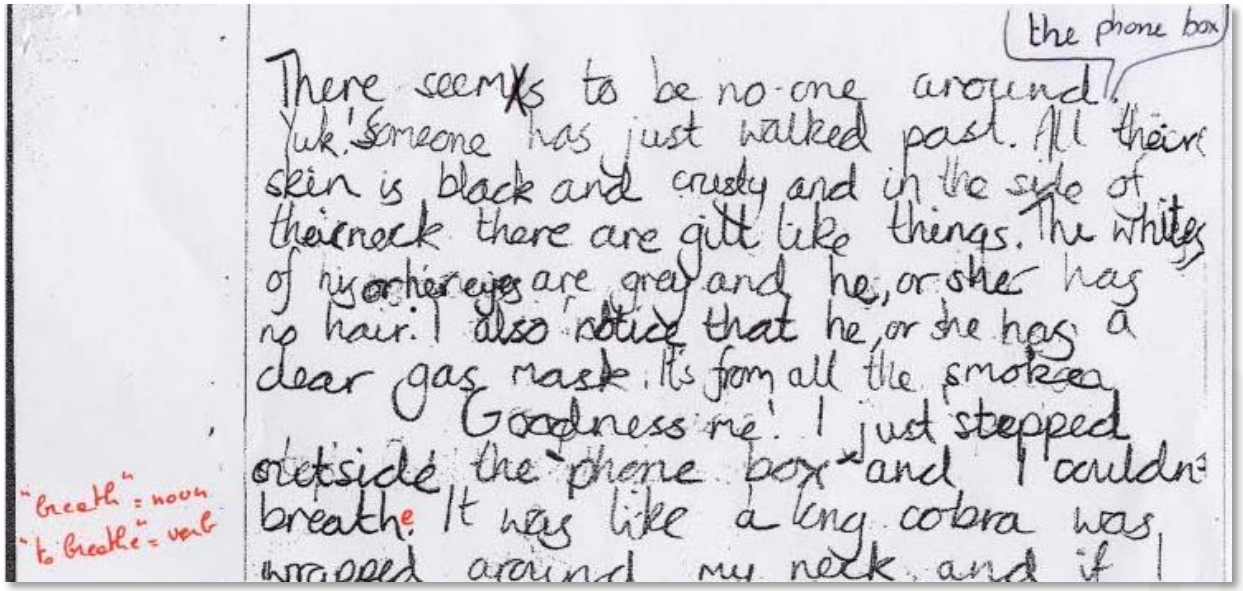
Beispiele für Plain text-Formate:

Vorteile:

- maschinenlesbar
- plattformunabhängig
- langfristig verfügbar
- nachnutzbar
- unabhängig von proprietärer Software
- flexibler Datenaustausch

Ikons:
https://commons.wikimedia.org/wiki/File:Text-tsv-text.svg
https://commons.wikimedia.org/wiki/File:Text-csv-text.svg
https://commons.wikimedia.org/wiki/File:Text-txt.svg

Da gibt’s doch aber noch was Besseres!




```
<p>There seem<del rend="crossout" hand="#HR">e</del>s to be no-one around <add place="above" hand="#HR">the phone box</add>. Yuk! <subst hand="#HR" resp="#TBE">
  <del rend="writeover">s</del>
  <add>S</add>
</subst>omeone has just walked past. All their skin is black and crusty and in the side of their neck are <unclear reason="illegible" resp="#TBE">gitt</unclear> like
things. The whites of his or her eyes are grey and he, or she has no hair. I also notice that he, or she has a clear gas mask. It's from all the smoke.</p>
<p>Goodness me! I just stepped outside the phone box and I couldn'<unclear reason="cutoff while xeroxing" resp="#TBE">t</unclear> breath<add hand="#teacher">e</add>.
<add hand="#teacher" place="margin">"breath" = noun <lb/> "to breathe" = verb</add> It was like a <choice>
  <sic>kng</sic>
  <corr>king</corr>
</choice> cobra was wrapped around my neck and if I was to get a little bit of air in it was horrible, like when you stop at red man beside the traffic lights and a
```

Abb.: https://teibyexample.org/tutorials/TBED06v00.htm

Concepts

paragraph

footnote

heading

XML

Syntax

```
<element>
    <element attribute="value">
        content
    </element>
</element>
```

TEI

Language:
vocabulary and grammar

```
<p>

<note type="foot">

<head>
```

Abb.: https://wwp.northeastern.edu/outreach/seminars/_utils/gfx/concept_syntax_language.png

Textauszeichnung / Annotation macht implizites Wissen über den Text durch Metadaten explizit!

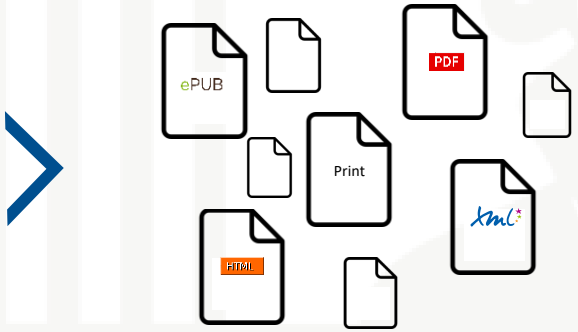
- Die Auszeichnungssprache TEI stellt eine spezielle, für bestimmte Zwecke definierte Markup-Sprache dar.

- Sie beruht auf der **universellen Markup-Sprache XML**

- XML liegt das Prinzip der Trennung des Informationsgehalts eines Dokuments von dessen äußerer Form zugrunde (generic markup)

- Mit der Annotation in TEI/XML werden Aussagen über die logische oder semantische **Bedeutung** der annotierten Textstelle getroffen

- Vorteile:
  - vielfältige Verwendbarkeit der Daten, da Unabhängigkeit vom Ausgabeformat
  - Menschen- und Maschinenlesbarkeit langfristig gewährleistet:



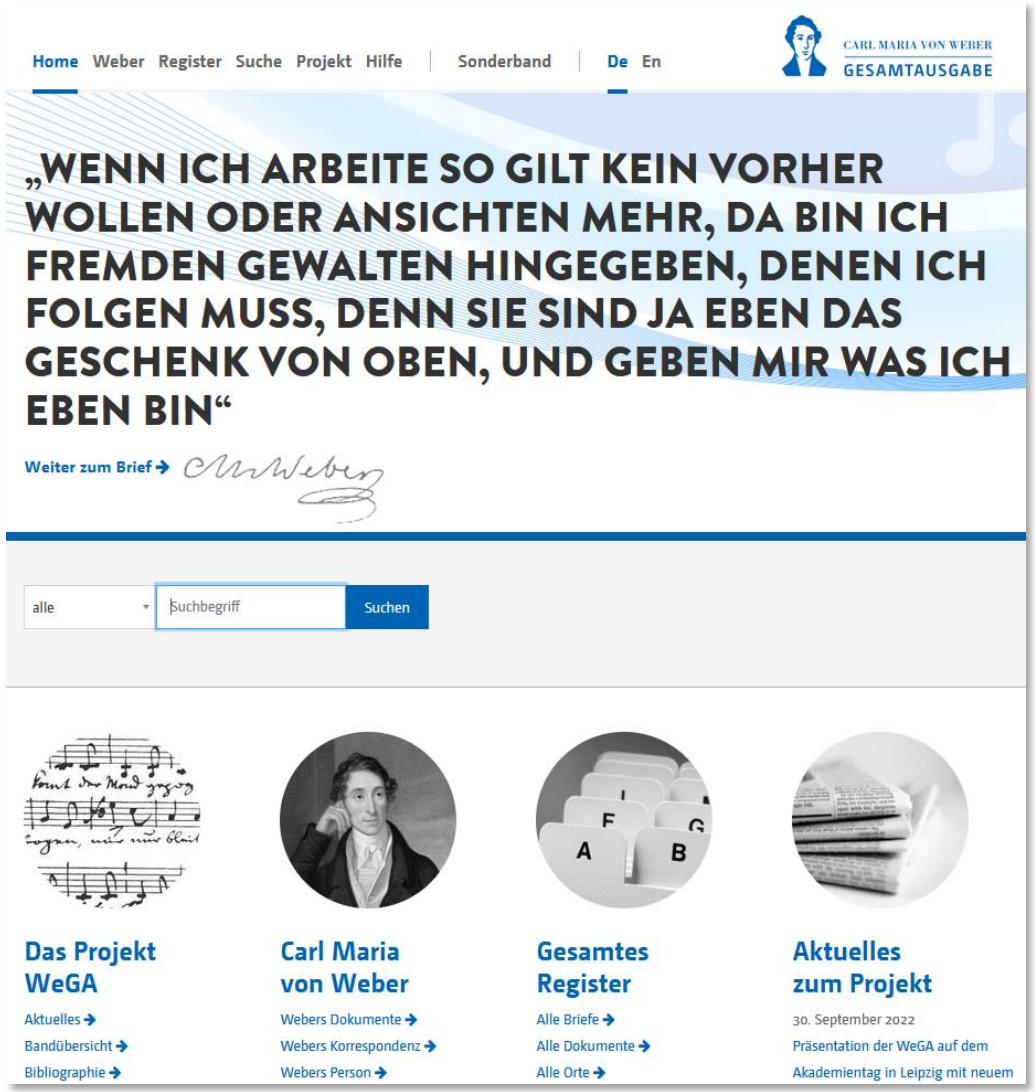
TEI Schema

Module name	Formal public identifier	Where defined
analysis	Analysis and Interpretation	17 Simple Analytic Mechanisms
certainty	Certainty and Uncertainty	21 Certainty, Precision, and Responsibility
core	Common Core	3 Elements Available in All TEI Documents
corpus	Metadata for Language Corpora	15 Language Corpora
dictionaries	Print Dictionaries	9 Dictionaries
drama	Performance Texts	7 Performance Texts
figures	Tables, Formulae, Figures	14 Tables, Formulæ, and Graphics
gaiji	Character and Glyph Documentation	5 Representation of Non-standard Characters and Glyphs
header	Common Metadata	2 The TEI Header
iso-fs	Feature Structures	18 Feature Structures
linking	Linking, Segmentation, and Alignment	16 Linking, Segmentation, and Alignment
msdescription	Manuscript Description	10 Manuscript Description
namesdates	Names, Dates, People, and Places	13 Names, Dates, People, and Places
nets	Graphs, Networks, and Trees	19 Graphs, Networks, and Trees
spoken	Transcribed Speech	8 Transcriptions of Speech
tagdocs	Documentation Elements	22 Documentation Elements
tei	TEI Infrastructure	1 The TEI Infrastructure
textcrit	Text Criticism	12 Critical Apparatus
textstructure	Default Text Structure	4 Default Text Structure
transcr	Transcription of Primary Sources	11 Representation of Primary Sources
verse	Verse	6 Verse

http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ST.html#STMA

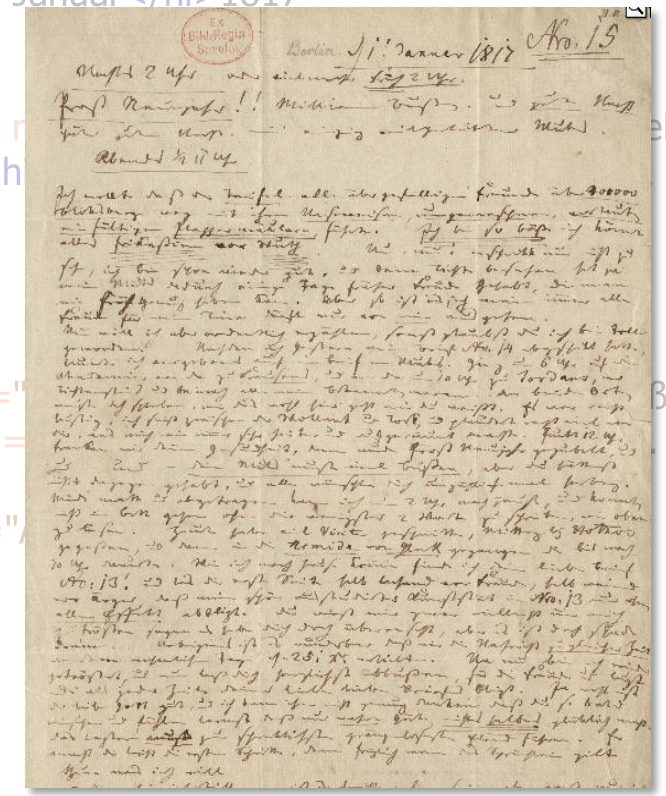
Das TEI Schema definiert die Struktur und Semantik der zu annotierenden Dokumente

Praxisbeispiel *Digitale Edition*



https://weber-gesamtausgabe.de/de/Index

```
<div>
<opener>
  <dateline>
    <hi rend="latintype">Nro</hi>: 15<lb/>d 1
    <hi rend="superscript">t</hi>
    <hi rend="latintype">Januar</hi>1817
  </dateline>
  <dateline rend="left">
    <hi rend="underline" r                                    ehr <hi
rend="underline" n="1"><h
Uhr</hi>.
  </dateline>
</opener>
<p n="1">
  <hi rend="underline" n="                                    ßen.
und<hi rend="underline" n=                                    . mein
einzig vielgeliebter
  <rs type="person" key="
</p></div>
```



Franz Anton von Weber an Fran ×
https://weber-gesamtausgabe.de/de/A002074/Korrespondenz/A040167.html#bs-tab-transcription
bibintern UB @ubffm OTRS Nachlass Schopenhau... VL VL-qa Pixelrechner IVLImageinfo Pica Recherche RMUGitLab BigBlueButton RD-Webclient UB Chat RC Senckenberg Confluence Lerncampus SDD Weitere Lesezeichen

FRANZ ANTON VON WEBER AN FRANZ ANTON HOFFMEISTER (HOFFMEISTER & KÜHNEL) IN LEIPZIG CARLSRUHE IN OBERSCHLESIEN, MONTAG, 26. MAI 1806

Karlsruh in Oberschleßien
d 26t *May* 1806.

Geliebtester alter Herzensfreundt! und Or[dens]br[uder]!!!

Ich hoffe noch Anspruch machen zu dürfen auf unsre alte vieljährige Bekanntschaft ud Freundtschaft, und wenn Sie mich also noch dafür anerkennen, wie ich hoffe, so zeigen dieselben mir es dadurch, das sie mir, wo möglich, mit umgehendem Postwagen, die *Partitur* des Vater unsers vom seel. KapelMeister *Naumann*, versteht sich gegen Baar Zahlung anhero zu schiken belieben wolten, es solte zum Beschluß *Benefice* meines Karls als *Opern Director* in *Bressl* seyn.

Kann ich Ihnen in *Bressl* oder sonsten in hiesigen Gegenden Gefälligkeit erzeigen, so befehlen Sie Geliebter Br: dem jenigen, so sich Achtungsvoll nennt.

Dero
g[an]z ergeb[en]ster D[iene]r und o[rdens]br[uder]
Freyherr *von Weber*.
Kammerherr

Einstellungen

Zeige Markierungen im Text

- Lesarten Δ
- Textkonstitution ‡
- Herausgeberergänzungen [ ]
- Einzelstellenerläuterung *
- Themenkommentare T

Dokument

- Text
- Apparat
- Faksimile
- XML-Vorschau
- Download

Kontext

Die in der Beilage Nr. 72 Alle hervorheben Groß-/Kleinschreibung Akzente Ganze Wörter 1 von 1 Übereinstimmung

UNIVERSITÄTS BIBLIOTHEK FRANKFURT AM MAIN

- ✓ Build once, use many-Ansatz durch TEI/XML als Basis
- ✓ Dokumentation
- ✓ Offene Exportformate
- ✓ Offene Lizenzen
- ✓ API
- ✓ Linked Data

<trailer>Vielen Dank für Ihre Aufmerksamkeit!</trailer>