



Textkodierung und Textanalyse mit TEI, UE

Einheit 4:
Wiederholung, Einleitung XPath/XSLT,
Publikationsstrategien

Sarah Lang

*Zentrum für Informationsmodellierung -
Austrian Centre for Digital Humanities*

Ablauf

9:00-10:30 / 1,5h: Vorstellung; Wiederholung; Publikationslösungen

10:30-15:00 / 4,5h: Arbeit an den neuen Themen der heutigen Einheit): Sie bekommen jeweils 2h pro Thema (XPath und XSLT) Zeit, in der Sie sich das Video selbstorganisiert ansehen und die Übungen absolvieren. Währenddessen können Sie im Videochat mit Kolleg:innen kommunizieren, telefonieren, zusammenarbeiten - wie Sie möchten. Ich stehe jederzeit für Fragen zur Verfügung. Machen Sie selbstorganisiert Pausen, wie Sie es benötigen.
Abgabe der Übungen über Moodle.

15:00-15:30 / 30min: Sie können mit der Hausübung/Assignment 4 beginnen. Ich stehe für Fragen zu Verfügung. Nutzen Sie diese Möglichkeit!

Wiederholung

Wiederholung: XML

- **Damit Daten automatisiert von Computern verarbeitet werden können, müssen wir sie strukturieren.**
 - Ein Beispiel dafür ist XML, das wir im Kurs erlernen. Es basiert - wie auch das verwandte HTML- auf SGML. (Im Unterschied zu HTML ist XML eben *extensible*, d.h. man kann seine eigenen Elemente erfinden, was bei HTML nicht der Sinn der Sache ist).
 - XHTML ist HTML, das der XML-Struktur folgt.
- XML
 - ist eine **Meta-Syntax**, mit der man andere Texte beschreiben kann.
 - ist ein Subset von SGML, das aber mehr als nur Text beschreiben kann (auch zB Bilder, Musik, RDF).
- Damit die Daten untereinander austauschbar werden (und auch Zeit und Energie bei der Wahl passender Kodierungsschemata gespart werden kann), gibt es **XML-Standards**.
- Ein sehr bekannter ist die **TEI** (*Text Encoding Initiative*), die zum De-Facto-Standard in den Geisteswissenschaften geworden ist.

Wiederholung: XML

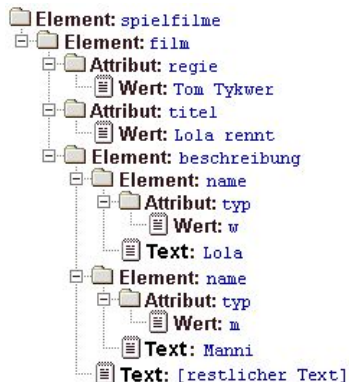
- XML → Datenstandard zur **Speicherung und zum Austausch von Daten**, der **menschen- und maschinenlesbar** ist.
- **Wohlgeformtheit** (Syntax, “Vokabular”), **Validität** (nach Schema, Semantik, “Grammatik”).
- **Speicherung unabhängig vom Ausgabe-/Präsentationsmedium** (*single source Prinzip*/Publishing)
- **XML** liefert Syntax; das **Schema** die zu verwendenden Elemente und wie sie eben zu verwenden sind. Es gibt Schemasprachen zur Beschreibung dieser Schemata, z.B. DTD und XMLSchema; Bsp. in der Einheit: Dublin Core.
- **XML-Standards/TEI**: Man muss das nicht alles auswendig können, sondern schaut Elemente immer nach Bedarf nach, weil sie sehr ausführlich definiert sind. Wie XML generell funktioniert, sollte man wissen (und lernt es auch schnell). Dieses Wissen gilt unabhängig von den Schemata und kann für jedes Schema weiterverwendet werden.

Wiederholung: XML

Wir merken uns:



1. **Russische Puppen** (Verschachtelung, kein *overlapping markup* erlaubt, falls nötig gibt es aber Behelfslösungen dafür)
2. **Doppelkeks** (Starttag-Inhalt-Endtag)
3. **Baumstruktur** (mit nur 1 Wurzelknoten/-element = root).



Schemata

Maintenance Helfer, um Redundanz und Inkonsistenz zu vermeiden, d.h. eigentlich eine Art Hilfstool für das praktische Arbeiten, da wichtig wird, sobald die Anzahl der verwendeten Elemente, Attribute und Attributwerte zu groß wird, als dass man leicht den Überblick verlieren könnte oder z.B. sehr viele Leute am Projekt beteiligt sind, aber ggf. alle mit wenig Stunden (zB Studienassistenten), sodass die Einzelnen nie so tief in der Materie sind, dass sie da noch durchsteigen würden bzw das realistisch möglich ist, dass sie das noch können, ohne Inkonsistenzen einzubauen. Hier helfen ODD und Schemata, weil sie eben nur erlaubte und dem Schema entsprechende Strukturen zulassen.

Was bisher getan wurde...

1. Block

- a. Struwwelpeter-Text kodiert und dabei mind. 1 DC-Element verwendet.
- b. Nach einem TEI-Basics-Intro: 1 Bsp in TEI kodiert - inkl. ein paar Entitäten verwenden; Names, Places, etwas normalisieren oder klassifizieren mit Attributen (z.B. @type).

2. Block

- a. Einführung in die Grundstruktur von TEI Dokumenten (teiHeader etc.)
- b. Überblick über die TEI-Guidelines
- c. Auszeichnung von Metadaten im <teiHeader>
- d. Projektspezifische Anpassung
- e. ROMA
- f. ODD-Übung zum Struwwelpeter

3. Block

- a. Informationen zum Textbegriff nach Sahle
- b. Kodierung der Materialität des Textes

The historian's question of relevance: “So what?”

- Prof. Siegfried Beer



XML vs XSLT

→ XML

Daten gut aufnehmen und in adäquatem Detailgrad / Detailliertheit dokumentieren (z.B. bei komplexer Textkritik / schwieriger Überlieferungssituation mit vielen Textzeugen). → super

Das wird dann aber gerade in den Fällen, für die es sehr gut geeignet ist (komplexe Situationen) sehr unübersichtlich und zerreit quasi das Dokument. Es bleibt zwar durch die Verwendung von XML in der Theorie menschenlesbar, in der Praxis wird es aber so unübersichtlich, dass man zB bei 100 *<choice>*s und Lesarten irgendwann den Überblick über den Haupttext verliert.

→ XSLT

Nach dem *Single Source* Prinzip andere/unterschiedliche, ggf interaktive Repräsentation aus denselben Daten für Ansicht, Präsentation, Visualisierung und Interaktion mit den Daten generieren

Mit XSL-Basisstylesheets zur Digitalen Edition

- **Basisstylesheets** als Weg zu dem, was für viele, die eine “Digitale Edition” von einem wollen (oder eine “Datenbank”), das eigentliche Endziel ist: **Eine möglichst ansehnliche Website bzw Web- oder Buchrepräsentation der Daten, mit guten Funktionen.**
- Die vorhandenen Rundumschlag-Basisstylesheets sind sehr gut und nützlich, aber es passiert schnell mal, dass man andere spezifische Funktionen bräuchte / gern hätte.
- **Customizing:** Man muss XSLT können, aber dafür hat man ein sehr machtvolleres Tool bei der Hand. Es gibt hier ganz unterschiedliche Wege anzusetzen, von ganz easy bis extrem komplex.
- **Vorteile:** Anderer Aspekt der Arbeit mit Stylesheets auf eine Vorzeigerepräsentation hin: Oft merkt man erst beim Visualisieren, dass Informationen in den Daten fehlen / nicht kodiert sind, die man später dann gern visualisieren / anzeigen / einzeln ansprechen möchte.
Weiterer Vorteil: Allgemeine Ansprechbarkeit mithilfe von XSLT (Bsp Dracula: von hand kodieren würde ewig dauern, weil so viele Seiten. Mit ein paar Regex kann ich in Nullkommanix ein schön gesetztes LaTeX-PDF aus den Projekt Gutenberg Daten machen), was schon ziemlich cool / nützlich ist.
- Was nicht in den Daten kodiert ist, kann ich hinterher nur mehr sehr schwer ansprechen. Z.B. alle Personen, unabhängig vom Namen - d.h. Ich muss keine Namensformen oder sowas mehr beachten und brauche nur eine Regel für das ganze Phänomen.

TEI und was nun?

Publikationslösungen

Vielfalt an Möglichkeiten

- Ressourcenliste:
https://wiki.tei-c.org/index.php/Category:Publishing_and_delivery_tools
- teiPublisher
- EVT
- TEICHI
- Versioning Machine
- Boilerplate, Oxgarage, GAMS...

TEI Publisher

- “TEI Publisher: The Instant Publishing Toolbox”
- <https://teipublisher.com/index.html>
- https://wiki.tei-c.org/index.php/TEI_Publisher
- <https://wiki.tei-c.org/index.php/TeiPublisher>
- <http://hisoma.huma-num.fr/exist/apps/tei-publisher/index.html>
- Als “Testimonial”: Florian Chiffolleau (2020), Blogbeitrag: Publication of my digital edition – Working with TEI Publisher.
<https://digitalintellectuals.hypotheses.org/3912>

Edition Visualization Technology (EVT)

- <http://evt.labcd.unipi.it/>
- <https://visualizationtechnology.wordpress.com/>
- Roberto Rosselli Del Turco, Giancarlo Buomprisco, Chiara Di Pietro, Julia Kenny, Raffaele Masotti and Jacopo Pugliese: *Edition Visualization Technology: A Simple Tool to Visualize TEI-based Digital Editions*, Journal of the Text Encoding Initiative 8 (2014/15): Selected Papers from the 2013 TEI Conference “TEI Processing: Workflows and Tools”; <https://doi.org/10.4000/jtei.1077>
- Del Turco, Roberto Rosselli. “Designing an Advanced Software Tool for Digital Scholarly Editions: The Inception and Development of EVT (Edition Visualization Technology).” *Textual Cultures*, vol. 12, no. 2, 2019, pp. 91–111. JSTOR, www.jstor.org/stable/26821538. Accessed 13 Jan. 2021.
- <https://tei-c.org/activities/projects/the-versioning-machine/>

TEI CHI

- “TeiCHI – Bringing TEI Lite to Drupal”
- Sebastian Pape, Christof Schöch and Lutz Wegner: *TEI CHI and the Tools Paradox. Developing a Publishing Framework for Digital Editions*, Journal of the Text Encoding Initiative 2 (2012): Selected Papers from the 2010 TEI Conference; <https://doi.org/10.4000/jtei.432>
- TEI CHI - <http://www.teichi.org/>
- <https://wiki.tei-c.org/index.php/TEI CHI>

Versioning Machine

- <http://v-machine.org/>
- <https://digitalhumanities.duke.edu/tools/versioning-machine>

TEI Boilerplate

- <http://teiboilerplate.org/>
- Demo: <http://dcl.slis.indiana.edu/teibp/content/demo.xml>
- https://wiki.tei-c.org/index.php/TEI_Boilerplate
- Slides: <https://www.i-d-e.de/wp-content/uploads/2014/07/boilerplate.pdf>
- Exercise: <http://tei.it.ox.ac.uk/Talks/2015-07-dhoxss/ex-boilerplate.pdf>
- Poster: <http://journalofdigitalhumanities.org/2-3/tei-boilerplate/>

OxGarage

- REST-Transformationsservice für TEI-Dokumente
- <https://oxgarage.tei-c.org/#>
- Diverse Datenformate nach TEI
- Basiert auf TEI-Basisstylesheets (<https://tei-c.org/Vault/P5/3.6.0/doc/tei-xsl/>)
- Von TEI in diverse Datenformate
 - **Vorteil:** universell anwendbares Tool
 - **Nachteil:** nicht alle TEI-Elemente werden bedacht; man möchte vielleicht dann doch lieber auf die eigenen Bedürfnisse anpassen.

→ Siehe Video-Bsp.

Hausübung / Ass4

Verwenden Sie das bereitgestellte Stylesheet “mini-bootstrap.xsl”, wo bereits Bootstrap eingebunden und ein Grundaufbau vorhanden ist.

1. **Recherchieren** Sie Inspirationen und Möglichkeiten von Bootstrap, z.B.
 - a. <https://www.w3schools.com/howto/>
 - b. <https://www.w3schools.com/bootstrap/>
2. **Ergänzen Sie das Basisstylesheet um mindestens 5 neue Template-Regeln** (davon mind. 3 mit “komplexeren” XPath-Abfragen, d.h. die über *t:p* hinausgehen). Nutzen Sie diese, um für Benutzer:innen sinnvolle Visualisierungen Ihrer TEI-XML-Daten zu erzielen. Entfernen Sie überflüssige Elemente des Bootstrap, also z.B. Textteile, die Sie nicht benötigen. Sie können bereits in der Stunde damit beginnen und Fragen stellen.
3. **Dokumentieren Sie Ihre Entscheidungen** und Erfahrungen in der Umsetzung (½ Seite lesbarer und nachvollziehbarer Fließtext, inkl. Einleitungssatz “Mit dem Ziel, eine TEI-Repräsentation von Struwwelpeter Text Soundso als Webseite darzustellen.... Wahl der Elemente XY, um ABC zu visualisieren, etc.)).