

Im Netz der Möglichkeiten - Wechselwirkungen in der Entwicklung von Theorie, Methode und Tools in den Digital Humanities am Beispiel der TEI

Schaßan, Torsten

schassan@hab.de

Herzog August Bibliothek Wolfenbüttel, Deutschland

In der Beschreibung der DHd-Konferenz 2018 „Kritik der digitalen Vernunft“ wird die These formuliert, dass die Digital Humanities „häufig als digital transformierte Bearbeitung von Fragestellungen aus den verschiedenen beteiligten Fächern beschrieben“ werden und „in weiten Teilen eine daten-, algorithmen- und werkzeuggetriebene Wissenschaft sei[en], die von ihren unmittelbaren Möglichkeiten und ihren Praktiken dominiert“ werden, welche den Prinzipien kritischer Wissenschaftlichkeit möglicherweise nicht genüge. Am Beispiel der Entwicklung der Guidelines sowie der Analyse einzelner konkreter Anwendungsfälle der Text Encoding Initiative (TEI; Synonym auch für die Richtlinien der Initiative, den „Guidelines for (Electronic) Text Encoding and Interchange“) soll untersucht werden, welche Wechselwirkungen es hierbei zwischen Theorie, Methode und Tool-Entwicklung gegeben hat.

Der Gegenstand

Der Beobachtungsgegenstand dieser Untersuchung sind die Guidelines der TEI selbst, die daraus resultierenden Schemata sowie deren Anwendung in ausgewählten Anwendungsfeldern der DH. Die TEI dokumentiert die Entwicklung der Guidelines seit der Einführung 1990 mit der Version P1 in unterschiedlicher Tiefe. Während für die älteren Versionen meist nur deren Endprodukt in Form der Document Type Definition (DTD) bzw. des Textes der Guidelines vorliegt, kann man die Entwicklung des de-facto-Standards seit der Version P4 und dann ab 2007 in listenförmiger Dokumentation der vorgenommenen Änderungen nachvollziehen. Die TEI dokumentiert außerdem die Diskussionen, die in ihren Gremien sowie in der Community über Mailingliste geführt werden, in je eigenen Archiven. Damit lässt sich ein nahezu lückenloses Bild der Entwicklungsschritte hin zu der geltenden Version nachvollziehen. Diese Materialien können vor dem Hintergrund der oben genannten Thesen untersucht und mit den traditionellen Wissenschaften

in Kontext gesetzt werden, um Aufschlüsse über die Wechselwirkungen zu erhalten.

Ein weiterer Gegenstand besteht in der konkreten Anwendung der Guidelines bzw. Schemata der TEI in Projekten. Hier sollen beispielhaft Editionsprojekte sowie die Verwendung der TEI zur Speicherung von Metadaten betrachtet werden, um die Wechselwirkungen zu analysieren. Es soll damit der Frage nachgegangen werden, inwieweit die Definition einer Markup-Sprache, die Anwendung der Sprache bei der Erstellung konkreter Dokumente und die Validierung der Ergebnisse durch technische Rahmenbedingungen vorgegeben oder durch außerhalb der DH liegenden Theoriebildung beeinflusst werden oder die Theoriebildung und Methodenentwicklung selbst vorantreiben.

Theoriebildung

Bereits 1994 wies Sperberg-McQueen auf die theoretischen Implikationen der Textauszeichnung, analog dazu: des Gebrauchs einer Wissenschaftssprache, hin: „Like any notation, the TEI Guidelines inevitably make it easy to express certain kinds of ideas, and concomitantly harder to express other kinds of ideas, about the texts we encode in electronic form. Any notation carries with it the danger that it must favor certain habits of thought --- in the TEI's case, certain approaches to text --- at the expense of others. No one should use TEI markup without being aware of this danger --- any more than we should use the English language, or any other, without realizing that it favors the expression of certain kinds of ideas, and discourages the expression, and even the conception, of other ideas.“ (Sperberg-McQueen 1994) Die TEI-Community versucht allerdings dennoch, diesen Gefahren mit der offenen Diskussion über die semantischen Dimensionen des Markups, mit Best-Practice-Beispielen für die Anwendung entgegen zu wirken. Die Community beweist immer wieder die Fähigkeit, den etablierten Standard im Fall erweiterter Anforderungen an neue Aufgaben anzupassen. Paradigmenwechsel in der Editionsphilologie können dadurch adaptiert, in der Auszeichnungspraxis angewendet und somit wissenschaftlich nutzbar gemacht werden, ohne deshalb vorhandene Dokumente ungültig werden zu lassen oder die bisherige Anwendung des Standards zu konterkarieren. Beispiele hierfür sind die Inkorporation von Markup-Subsystemen für genetische Editionen, zur Beschreibung von Handschriften oder auch zur Dokumentation von Kommunikationsvorgängen, etwa in Briefen, mit der Einführung von <correspDesc>. Das System der TEI wurde dabei prinzipiell erweitert als grundlegend verändert. Die Abschaffung von Elementen oder Attributen, die etwa die Falsifikation nach Popper gleich zu setzen wären, ist daher sehr viel seltener als der Proof of Concept, mit welchem zunächst die Tauglichkeit von Markup zur Repräsentation bestimmter Phänomene getestet wird, bevor die Aufnahme in den Standard erfolgt

und die Angebote durch die Community weiter genutzt werden können.

Gleichzeitig bestimmen die Anforderungen der inhaltlichen bzw. theoretischen Weiterentwicklung eines Faches die technische und theoretische Weiterentwicklung der TEI mit. Ein rezentes Beispiel hierfür ist die theoretische Ausrichtung des *Material Turns* in der Editionsphilologie und den historischen Wissenschaften. Die Hinwendung zum Objekt wird in der TEI mit der Einführung von <facsimile> und dem Modul für genetische Editionen gespiegelt. Auch Überlegungen zur Verallgemeinerung des Beschreibungsstandards von Dokumenten und Objekten mittels der Strukturen von <msDesc> wären hier zu nennen.

Methodenbildung

Ein Beispiel für die Anwendung neuer Methoden der DH ist die Anreicherung von Texten um bestimmte Kontexte. Wenn in einem Text benannte Entitäten wie Personen, Orte, Objekte oder Ereignisse ausgezeichnet werden, dann entspricht dies zunächst der Registerarbeit traditioneller Publikationsverfahren. Wenn diese Entitäten allerdings mit Normdaten angereichert und diese somit zu nachnutzbaren Bestandteilen im Sinne der Linked Open Data (LOD) werden, darf man wohl von der Anwendung einer neuen Methode sprechen. Die entstandenen Dokumente verändern ihren Charakter von traditionell erstellten und händisch ausgewerteten Objekten zu konzeptionell gestalteten und automatisiert nutzbaren. Die Dokumente werden in die Lage versetzt, auch außerhalb bestimmter, inhaltlich vorgegebener Auswertungskontexte mit anderen Methoden ausgewertet zu werden.

Die Konformität zu den TEI-Richtlinien spielt in der Anwendung der neuen Methode eine wesentliche Rolle. Im System der TEI ist die *Customisation* als Normalfall vorgesehen. Unter *Customisation* wird die kontextabhängige Auswahl von Modulen, Elementen und Attributen oder die Vorgabe von Wertemengen für Attribute verstanden. Diese Vorgaben können in der TEI in sogenannten „One Document Does it all“-Dateien (ODD) definiert und dokumentiert werden, um dann anschließend daraus eigene Datenstrukturdefinitionen in Form von Schema-Dokumenten zu generieren. Zunehmend werden hierbei neben den klassischen Schemasprachen zur Validierung von Datenstrukturen (DTD, RelaxNG, XML Schema) auch Regelwerke definiert, welche eine Validierung des Inhalts eines Dokumentes ermöglichen. (Schematron) ODD sind durch die Dokumentationsleistung und eine normierte Beschreibungssprache ein starkes Hilfsmittel, um die Verwissenschaftlichung von textuellem Markup und dem unterliegenden Textverständnis zu erreichen. Wie aber der Grad der TEI-Konformität zu messen und zu beschreiben sei, ist in der TEI bereits seit längerem Gegenstand von Diskussionen.

Tool-Entwicklung

Wo nun mehr Texte in TEI-konformer Auszeichnung vorliegen, umso wichtiger wird die Existenz von Tools, um diese Daten zu nutzen. Wo Text- und Image-Digitalisierung Quellen einfacher zugänglich macht, werden diese Materialien potentiell als Gegenstand der Digital Humanities nützlich. Ähnlich wie bei der Theorie- oder Methodenbildung befruchten sich zur Verfügung stehende Materialien und daran anzulegende Fragestellungen gegenseitig. Die Kritik der digitalen Methoden wird hier ansetzen müssen, wo das Markup vom Ende her zu denken ist, weil bestimmte Funktionalitäten gewünscht sind: Dokumente sollen nach bestimmten Kriterien sortiert werden? Was bedeutet dies für die Aufbereitung der Daten? – Dokumente sollen im Rahmen des LOD genutzt werden? Welche Eigenschaften müssen sie hierfür haben?

DH, eine Wissenschaft wie jede andere?

Die in einer Wissenschaft zur Verfügung stehenden Quellen und deren Verarbeitungsmechanismen haben sich immer schon gegenseitig sowie die Theorie- und Methodenbildung beeinflusst. An den Beispielen der Untersuchung soll gezeigt werden, dass sich dies in den DH nicht anders ausnimmt.

Bibliographie

Lou Burnard / Sebastian Rahtz (2004): "RelaxNG with Son of ODD", in: *Proceedings of Extreme Markup Languages 2004*. <http://conferences.idealliance.org/extreme/html/2004/Burnard01/EML2004Burnard01.html>

Lou Burnard / C. Michael Sperberg-McQueen (1995): "The Design of the TEI Encoding Scheme", in: *Computers and the Humanities* 29 (1) p. 17–39. 10.1007/BF01830314

Guidelines for Electronic Text Encoding and Interchange, edited by TEI Consortium. 1990-, P1–P5. <http://www.tei-c.org/Guidelines/> , <http://www.tei-c.org/Vault/>

Patrick Sahle (2013): *Digitale Editionsformen. Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels*. 3 Bde. Norderstedt: BoD. <https://www.i-d-e.de/publikationen/schriften/s7-9-digitale-editionsformen/>

C. Michael Sperberg-McQueen (1994): *Textual Criticism and the Text Encoding Initiative*. (First draft of a paper presented at MLA, San Diego, 1994) <http://www.tei-c.org/Vault/XX/mla94.html>

C. Michael Sperberg-McQueen / Henry Thompson (2000): *XML-Schema*. <https://www.w3.org/XML/Schema>

Jörg Wettlaufer (2016): "Neue Erkenntnisse durch digitalisierte Geschichtswissenschaft(en)? Zur hermeneutischen Reichweite aktueller digitaler Methoden in informationszentrierten Fächern", in: *Zeitschrift für digitale Geisteswissenschaften*. DOI: 10.17175/2016_011.

RelaxNG (2008): *ISO/IEC 19757-2:2008. Information technology -- Document Schema Definition Language (DSDL) -- Part 2: Regular-grammar-based validation -- RELAX NG*. [http://standards.iso.org/ittf/PubliclyAvailableStandards/c052348_ISO_IEC_19757-2_2008\(E\).zip](http://standards.iso.org/ittf/PubliclyAvailableStandards/c052348_ISO_IEC_19757-2_2008(E).zip)

Schematron (2016): *ISO/IEC 19757-3:2016. Information technology -- Document Schema Definition Languages (DSDL) -- Part 3: Rule-based validation -- Schematron*. http://standards.iso.org/ittf/PubliclyAvailableStandards/c055982_ISO_IEC_19757-3_2016.zip