

Modellierung von Forschungsdaten durch Annotation

Zinsmeister, Heike

heike.zinsmeister@uni-hamburg.de
Universität Hamburg, Deutschland

Annotationen im Sinne von „Content“ (vgl. Agosti / Ferro 2007) modellieren Forschungsdaten anhand dreier Dimensionen: der Konzepte, der Repräsentationsformate und zusätzlich auch durch den Annotationsprozess.

Der Schwerpunkt dieses Beitrags liegt auf der konzeptuellen Modellierung textueller Daten durch Abstraktionen und die Hervorhebung charakterisierender Eigenschaften bis hin zum Ergänzen assoziativer Bezüge. Der Beitrag spannt dabei einen Rahmen von intuitiven Randkommentaren (vgl. Blustein et al. 2011) bis hin zu wohldefinierten Annotationsprojekten (z. B. Stede / Neumann 2014) und gibt einen Ausblick auf die Umsetzung in Repräsentationsformaten (Dipper 2005; Piez 2011).

Das Besondere an der Modellierung eines Textes oder einer Textsammlung durch Annotation ist, dass die Auszeichnungen und Kommentare nicht einem holistischen Ganzen zugewiesen werden, wie es normalerweise bei beschreibenden Metadaten der Fall ist, sondern dass sie mit bestimmten Bestandteilen des dekomponierten Textes verknüpft werden und damit eine gewissen Distribution im Text aufweisen, die der Modellierung grundsätzlich einen quantitativen Aspekt verleiht und für weiterführende Auswertungen und Visualisierungen verwendet werden kann.

In der geisteswissenschaftlichen Tradition bestehen unterschiedliche Zugänge zur Modellierung von Forschungsdaten durch Annotation. Beispielhaft stellt der Beitrag den prozessorientierten, hermeneutischen Ansatz der Literatur- und Kulturwissenschaft (Bradley / Vetch 2007, Bradley 2008) dem produktorientierten, deduktiven Ansatz der Korpus- und Computerlinguistik (Leech 1997, Pustejovsky / Stubbs 2012) gegenüber und schlägt eine Synthese der Modellierung in Anlehnung an die zusammenfassenden Darstellungen in Burghardt (2014) und Gius / Jacke (2015) vor. Zusätzlich gibt er einen Ausblick auf Repräsentationsmöglichkeiten von ambigen Annotationen (vgl. Barteld et al. 2014).

incomplete language phenomena“, in: Levin, Lori / Stede, Manfred (eds.): *Proceedings of LAW VIII - The 8th Linguistic Annotation Workshop* 99–104.

Blustein, James / Rowe, David / Graff, Ann-Barbara (2011): „Making Sense in the Margins: A Field Study of Annotation“, in: *Research and Advanced Technology for Digital Libraries*. Berlin: Springer 252–259.

Bradley, John (2008): „Pliny: A model for digital support of scholarship“, in: *Journal of Digital Information (JoDI)* 9, 1.

Bradley, John / Vetch, Paul (2007): „Supporting Annotation as a Scholarly Tool – Experiences from the Online Chopin Variorum Edition“, in: *Literary and Linguistic Computing* 22, 2.

Burghardt, Manuel (2014): *Engineering Annotation Usability-Toward Usability Patterns for Linguistic Annotation Tools*. Doktorarbeit. Universität Regensburg.

Dipper, Stefanie (2005): „XML-Based Stand-off Representation and Exploitation of Multi-Level Linguistic Annotation“, in: *Proceedings der Berliner XML-Tage 2005 (BXML 2005)* 39–50.

Gius, Evelyn / Jacke, Janina (2015): „Informatik und Hermeneutik. Zum Mehrwert interdisziplinärer Textanalyse“, in: Baum, Constanze / Stäcker, Thomas (eds.): *Grenzen und Möglichkeiten der Digital Humanities* (= Sonderband der Zeitschrift für digitale Geisteswissenschaften, 1). http://www.zfdg.de/sb001_006 [letzter Zugriff 09. Januar 2016].

Leech, Geoffrey (1997): „Introducing Corpus Annotation“, in: Garside, Roger / Leech, Geoffrey / McEnery, Tony (eds.): *Corpus Annotation*. Linguistic Information from Computer Text Corpora. London / New York: Longman 1–18.

Piez, Wendell (2010): „Towards Hermeneutic Markup. An Architectural Outline“, in: *Digital Humanities 2010. Conference Abstracts, präsentiert auf der Digital Humanities Conference 2010 (DH 2010)* 202–205.

Pustejovsky, James / Stubbs, Amber (2012): *Natural Language Annotation for Machine Learning*. O'Reilly Media, Inc.

Stede, Manfred / Neumann, Arne (2014): „Potsdam Commentary Corpus 2.0: Annotation for Discourse Research“, in: *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)* 925–929.

Bibliography

Agosti, Maristella / Nicola Ferro (2007): „A Formal Model of Annotations of Digital Content“, in: *ACM Transactions on Information Systems* 26, 1.

Barteld, Fabian / Ihden, Sarah / Schröder, Ingrid / Zinsmeister, Heike (2014): „Annotating descriptively