

Textmodellierung

Galka, Selina; selina.galka@uni-graz.at

Texte können unterschiedliche Textstrukturen, Inhalte, Erzählebenen, Lesarten, historische Informationen, Datierungen oder topografische Informationen enthalten. Diese impliziten Strukturen werden bei der Modellierung je nach Zweck und Forschungsfrage identifiziert und mit Hilfe von Auszeichnungssprachen für den Computer explizit gemacht, damit dieses Wissen maschinell weiterverarbeitet werden kann. Bei der Textmodellierung versucht man also, Strukturen und Informationen von Texten herauszuarbeiten.

Bei Brieftexten könnte man beispielsweise als modellierbare Elemente das Absendedatum und den Absendeort sehen, Sender, Empfänger und Grußformeln, aber auch z. B. im Brieftext selbst erwähnte Personen und Orte (vgl. Beispiel im Artikel zu Markup). Bei Gedichten könnte man einzelne Strophen oder Verse modellieren; bei Tagebuchtexten z. B. als größte Struktur das Tagebuch und darin eventuell die einzelnen Tagebucheinträge mit Datumsangaben und erwähnten Entitäten wie Personen, Orten oder Institutionen. Bei linguistischen Forschungsfragen werden einzelne Wörter und Wortteile getaggt.

Die Modellierung betrifft sowohl inhaltliche als auch strukturelle Ebenen – so können bei mehrere Seiten umfassenden Texten auch die einzelnen Seiten als Entitäten erfasst werden. Auch der Modellierungstiefe sind grundsätzlich keine Grenzen gesetzt; Texte könnten sogar bis hin zu einzelnen Wörtern oder sogar Zeichen modelliert werden. Dabei sollte jedoch immer der Zweck hinterfragt und das Forschungsziel nicht aus den Augen verloren werden.

Die modellierten Entitäten, Strukturen, Informationen usw. werden bei Digitalen Editionen mit Markup explizit gemacht und können somit mithilfe des Computers analysiert, visualisiert oder ausgewertet – allgemein gesagt also weiterverarbeitet – werden. Als Quasi-Standard hat sich in den Digitalen Geisteswissenschaften die TEI zur Codierung von Texten etabliert. Für das Einfügen des Markups existieren spezielle Annotationsumgebungen.

Literatur:

- The shape of data in digital humanities: modeling texts and text-based resources. Hrsg. von Julia Flanders und Fotis Jannidis. London: 2019.

Verweise:

Modellierung, Annotationsstandards, Annotationsumgebungen, Annotation, Markup, TEI, Interpretation, XML

Themen:

Annotation und Modellierung

Zitervorschlag:

Galka, Selina. 2021. Textmodellierung. In: KONDE Weißbuch. Hrsg. v. Helmut W. Klug unter Mitarbeit von Selina Galka und Elisabeth Steiner im HRSM Projekt "Kompetenznetzwerk Digitale Edition". URL: <https://gams.uni-graz.at/o:konde.195>