

Digital vs. Humanities. Didaktische Aufbereitung digitaler Methoden für die klassischen Geisteswissenschaften im Projekt forTEXT

Jacke, Janina

janina.jacke@uni-hamburg.de
Universität Hamburg, Deutschland

Horstmann, Jan

jan.horstmann@uni-hamburg.de
Universität Hamburg, Deutschland

Meister, Jan Christoph

j-c-meister@uni-hamburg.de
Universität Hamburg, Deutschland

Einleitung

Computergestütztes Arbeiten kann geisteswissenschaftliches Forschen auf unterschiedlichste Weise befördern und bereichern. Dennoch müssen wir in unserem Arbeitsalltag und in Gesprächen mit Kolleginnen und Kollegen¹ immer wieder feststellen, dass viele traditioneller arbeitende Geisteswissenschaftler digitalen Methoden noch immer mit Skepsis begegnen.² Dies liegt nicht zuletzt daran, dass in den Geisteswissenschaften zahlreiche Methoden zum Einsatz kommen, von denen nur einigen wenigen eine derart formalisierte Arbeitsweise naheliegt, wie sie im Rahmen der Digital Humanities oft verfolgt wird.

Das lässt sich gut am Beispiel der Literaturwissenschaft illustrieren: Digitale Methoden werden bisher vornehmlich von Literaturwissenschaftlern genutzt, die an strukturellen oder anderen formalen Aspekten literarischer Texte interessiert sind (beispielsweise an narrativen Strukturen, Figurennetzwerken etc.).³

Ihrem traditionellen Selbstverständnis nach ist die Literaturwissenschaft allerdings zentral an komplexen und innovativen *Interpretationen* literarischer Texte interessiert⁴ – und wie diese durch digitale Methoden der Textanalyse befördert werden können, ist nicht evident.

Damit digitale Methoden eine breitere Akzeptanz finden, ist es deswegen notwendig, den Nutzen dieser Methoden auch für stärker hermeneutisch ausgerichtete geisteswissenschaftliche Forschungsfragen

zu reflektieren. Unserem (weiten) Verständnis von “Hermeneutik” entsprechend handelt es sich bei hermeneutischen Forschungsfragen um Fragen, die auf die (holistische) Auslegung bzw. Deutung von Texten gerichtet sind (vgl. bspw. Spörl 2004: 128). In literaturwissenschaftlichen Zusammenhängen spielen dabei insbesondere Fragen nach Funktion bzw. Wirkung bestimmter Textelemente oder des Gesamttextes eine Rolle, ebenso wie die In-Beziehung-Setzung des Textes mit bestimmten Kontexten. Diese Forschungsfragen sollten exponiert im Zusammenhang mit der Entwicklung von Tools, digitaler Forschungsumgebungen und vor allem didaktischer Konzepte zur Vermittlung von DH-Methoden berücksichtigt werden. Diese Forderungen werden bisher jedoch nicht in zureichendem Maße erfüllt.⁵

Wir möchten in diesem Beitrag das aktuelle Projekt forTEXT (2017–2020) vorstellen, das der Vermittlung, Aufbereitung und Bereitstellung von Mitteln zur computergestützten Textanalyse insbesondere für hermeneutisch arbeitende Geisteswissenschaftler gewidmet ist. Im Folgenden sollen in diesem Zusammenhang zunächst die unterschiedlichen konzeptionellen Dimensionen (Abschnitt 2) sowie anschließend erste inhaltliche Ergebnisse des Projekts präsentiert werden (Abschnitt 3).

Dimensionen des forTEXT-Projekts

Paradigmen

Das Anfang 2017 gestartete DFG-Projekt *forTEXT. Literatur digital erforschen* (<http://www.fortext.net>) hat die Entwicklung einer digitalen Forschungsumgebung zum Ziel, die im Rahmen der qualitativen Analyse und Interpretation von Texten genutzt werden kann. Das Augenmerk bei der Gestaltung dieser Umgebung liegt insbesondere auf zwei Aspekten:

(a) Orientierung an genuin geisteswissenschaftlichen Arbeitsweisen: Es geht, ganz im Sinne des geisteswissenschaftlichen Selbstverständnisses, um die Unterstützung der genuin *interpretativen* Auseinandersetzung mit Texten. In anderen Worten: forTEXTs Fokus liegt *nicht* ausschließlich auf der statistischen Auswertung von Texten, wie es sonst im Rahmen von DH-Methoden zur Textanalyse oft der Fall ist.⁶ Auf diese Weise soll gewährleistet werden, dass traditioneller arbeitende Geisteswissenschaftler die Umgebung tatsächlich zur digitalen Unterstützung vertrauter Methoden der Textanalyse und -interpretation nutzen können und ihnen keine Hinwendung zur statistischen Textanalyse abverlangt wird.

(b) Niedrigschwelliger Zugang: Geisteswissenschaftler sollen die digitale Forschungsumgebung *intuitiv* und weitgehend ohne technische Vorkenntnisse nutzen können. Hierzu trägt zum einen die Tatsache bei, dass forTEXT ein individualisiertes Empfehlungssystem zur Verfügung

stellt, das geisteswissenschaftlichen Nutzern Vorschläge unterbreitet, welche digitalen Ressourcen, Routinen und Tools für ihr Projekt hilfreich sein könnten (siehe auch Abschnitt 2.2). Nutzer werden also nicht einfach mit einem unüberschaubaren digitalen Angebot alleingelassen, dessen potenziellen Nutzen für die eigene Fragestellung sie sich erst noch selbst erschließen müssen. Zum anderen stellt forTEXT leicht verständliche Beschreibungen zu digitalen Methoden und Korpora zur Verfügung, ebenso wie intuitiv bedienbare Benutzeroberflächen für digitale Werkzeuge zur Textanalyse und -interpretation (siehe auch Abschnitt 2.3). Auf diese Weise können DH-Methoden ohne technisches Know-how sowie ohne das aufwändige Studieren von Nutzerhandbüchern eingesetzt werden.

In den folgenden Unterabschnitten sollen sowohl forTEXTs Empfehlungssystem als auch die drei Komponenten des Informationsrepositoriums (*Routinen*, *Ressourcen* und *Tools*) kurz etwas detaillierter vorgestellt werden.

Individualisiertes Empfehlungssystem

Für Geisteswissenschaftler, die noch nicht wissen, auf welche Weise digitale Methoden der Textanalyse und -interpretation ihre eigene Forschung unterstützen können, bietet forTEXT ein individualisiertes Empfehlungssystem in Form eines digitalen Fragebogens an (siehe Abb. 1).

Abb. 1: forTEXTs individualisiertes Empfehlungssystem für digitale Textuntersuchung (Ausschnitt des Prototyps)

Hier können die Nutzer beispielsweise angeben, ob sie schon Vorerfahrungen mit digitalen Methoden der Textanalyse gemacht haben, in welchem Zustand sich ihr Textkorpus befindet, unter welcher Fragestellung sie ihre Texte untersuchen wollen und welcher literaturtheoretischen Schule sie sich zuordnen.⁷ Als

Output erhalten die Forscher, angepasst an die von ihnen gemachten Angaben, eine Liste mit Vorschlägen zu digitalen Korpora, Methoden und Werkzeugen, die zu ihrer Fragestellung und Arbeitsweise passen. Das Empfehlungssystem also eine individualisierte Kompilation aus forTEXTs Inhalten und Verzeichnissen, die im Folgenden kurz vorgestellt werden sollen.

Routinen, Ressourcen und Tools

forTEXTs digitale Forschungsumgebung ist in drei Bereiche gegliedert.

(a) Routinen: Im Teilbereich *Routinen* finden sich zum einen Informationstexte zu digitalen Methoden, die der Analyse und Interpretation von Texten dienen (bspw. zu taxonomiebasiertem Annotieren, zur Textanalyse mittels individualisierter Abfragen auf Text- und Annotationsdaten, zu Topic Modeling etc.), sowie zu vorbereitenden Prozeduren wie der Digitalisierung von Texten. In diesen Informationstexten finden sich darüber hinaus Links zu digitalen Tools (s.u.), mithilfe derer die fraglichen Methoden ausgeführt werden können.

Zum anderen werden unter *Routinen* auch didaktische Texte (d.h. Lerneinheiten und Lehrmodule) zur Verfügung gestellt. Die Lerneinheiten dienen der selbstständigen Aneignung bestimmter digitaler Methoden und Tools, während die Lehrmodule didaktisches Material für Lehrende zur Verfügung stellen, die auf 90-minütige Workshopsituationen zugeschnitten sind (siehe auch Abschnitt 3). Die Entwicklung neuer Lehrmodule wird sich dabei an den im Projektverlauf akquirierten Bedarfen der Nutzer orientieren.

(b) Ressourcen: Unter *Ressourcen* ist ein Verzeichnis digital nutzbarer Korpora zu finden. Hierunter fallen sowohl hochqualitativ digitalisierte Textkorpora als auch inhaltlich annotierte Korpora, die nachgenutzt werden können. Einige annotierte Korpora werden im Rahmen von forTEXT selbst bzw. affilierten Projekten produziert.⁸ Das Verzeichnis enthält darüber hinaus informative Beschreibungen der gelisteten Ressourcen.

(c) Tools: Im Bereich *Tools* findet sich schließlich eine kommentierte Liste digitaler Werkzeug-Suites bzw. Funktionskomponenten, mithilfe derer unterschiedliche textanalytische und -interpretatorische Operationen durchgeführt oder unterstützt werden können. Darüber hinaus sollen in forTEXT auch eigene Funktionskomponenten entwickelt werden. Hierzu gehört vornehmlich die Weiterentwicklung des Textannotations- und Analyseprogramms CATMA (<http://www.catma.de>) – aber auch die Entwicklung von graphischen Step-by-step-Benutzerschnittstellen für bestehende Tools.⁹ Dies soll es technisch weniger versierten geisteswissenschaftlichen Nutzern ermöglichen, hochfunktionale Tools einzusetzen, ohne sich umfangreich einarbeiten zu müssen.

Alle Verzeichnisse und Einträge aus den drei forTEXT-Bereichen Routinen, Ressourcen und Tools können von Nutzern eigenständig durchsucht und aufgerufen werden –

oder es erfolgt ein angeleiteter Zugriff durch die Nutzung des Empfehlungssystems.

Im verbleibenden Teil dieses Beitrags möchten wir etwas genauer auf einen Lehrmodulentwurf eingehen, das dem forTEXT-Bereich *Routinen* zuzuordnen ist. Anhand dieses Moduls soll beispielhaft gezeigt werden, wie in forTEXT die Paradigmen der Orientierung an einer genuin geisteswissenschaftlichen Arbeitsweise und des niedrigschwelligen Zugangs umgesetzt werden.

Lehrmodule Manuelles Annotieren

Zwei von forTEXTs 90-minütigen Lehrmodulen sind der Vermittlung der digitalen Methode des *manuellen Annotierens* gewidmet. Die Methode des Annotierens stellt einen guten Brückenschlag zur traditionelleren geisteswissenschaftlichen Arbeitsweise dar – schließlich gehört das Anbringen von Notizen in zu interpretierenden literarischen Texten seit jeher zur literaturwissenschaftlichen Arbeitspraxis.¹⁰ In den Lehrmodulen zum *digitalen Annotieren* gilt es nun, zum einen an die bereits bekannte Praxis anzuknüpfen und zum anderen deutlich zu machen, inwiefern die digitale Unterstützung es ermöglicht, bekannte Arbeitsprozesse deutlich effektiver durchzuführen, oder gar ganz neue Arbeitsweisen eröffnet, die das jeweilige Forschungsziel befördern.

Um diese Anforderungen umzusetzen, sieht forTEXT zwei Lehreinheiten zum manuellen Annotieren vor, von denen wir die erste im Folgenden kurz vorstellen möchten.

In der Einheit *Taxonomiebasiertes Annotieren* wird schrittweise gezeigt, wie mithilfe des Annotations- und Analysetools CATMA freie Kommentare in literarischen Texten angebracht sowie analysiert und systematisiert werden können. Die Systematisierung freier Kommentare kann wiederum als Grundlage dienen, um eine Annotationstaxonomie zu entwickeln, die dann für eine noch feinere und zielgerichtete Analyse des literarischen Textes genutzt werden kann.

Das verwendete Programm CATMA ist hierbei in zweifacher Hinsicht auf die forTEXT-Paradigmen abgestimmt: Es bietet eine intuitiv bedienbare Benutzeroberfläche und unterstützt den freien, undogmatischen und genuin interpretativen Zugang zu Texten, während es zugleich Optionen stärkerer Formalisierung bereithält (vgl. Abb. 2).¹¹

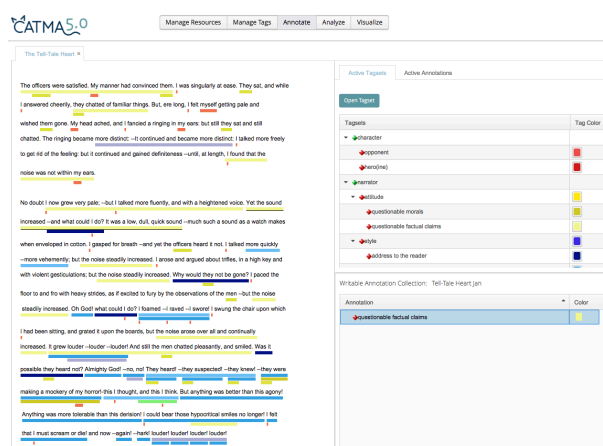


Abb. 2: Intuitives, nicht-deterministisches Annotieren in CATMA

Die im Rahmen des Lehrmoduls verfolgte didaktische Strategie hat mehrere Vorteile: Der erste Schritt, d.h. das digitale Anbringen freier Kommentare, stellt einen vollkommen explorativen und potenziell unstrukturierten Zugang zu literarischen Texten dar. Er erzwingt also kein formalistisches Umdenken und bildet die traditionellere geisteswissenschaftlich-hermeneutische Arbeitsweise gut ab. Im Vergleich zum analogen Arbeiten birgt er aber dennoch den Vorteil, dass die freien Kommentare durch digitale Unterstützung effektiver *nachgenutzt* werden können.

Als zusätzliches Angebot an Literaturwissenschaftler, die für eine etwas stärkere Formalisierung ihres Zugangs offen sind, zeigt das Lehrmodul, welche weiteren Vorteile und Optionen *taxonomiebasiertes* Annotieren mit sich bringt (bspw. bessere Vergleichbarkeit, detailliertere und vereinfachte Analyse, ggf. Reproduzierbarkeit etc.)¹² und wie dieses digital umgesetzt werden kann. Diese Herangehensweise kann in der Folge auch die Nützlichkeit noch stärker formalistisch anmutender DH-Techniken plausibilisieren – wie beispielsweise der kollaborativen, guidelinegestützten Annotation¹³ oder der automatisierten Annotation bzw. Informationsextraktion.

Im Lehrmodul zum manuellen digitalen Annotieren sind also die Paradigmen umgesetzt, die auch die weitere Arbeit am forTEXT-Projekt bestimmen sollen: Durch stärkere Orientierung an der traditionell-geisteswissenschaftlichen Arbeitsweise und erleichterten Zugang können digitale Methoden einer breiteren Nutzergemeinschaft nähergebracht werden.

Fußnoten

1. Wir werden aufgrund besserer Lesbarkeit im Folgenden nur noch die männliche Form verwenden – unsere Ausführungen beziehen sich aber selbstverständlich dennoch auf alle Geschlechter.

2. Auf diese persistente Skepsis verweisen beispielsweise auch Fiedler und Weiß in ihrem Tagungsbericht zur DHd-Konferenz 2015 in Graz (vgl. Fiedler/Weiß 2015).

3. Beispiele hierfür sind u.a. die folgenden Projekte aus dem Bereich der digitalen Literaturwissenschaft: "heureCLÉA", ein Projekt zur automatischen Annotation von Zeitphänomenen in narrativen Texten (<http://www.heureclea.de> , vgl. Bögel et al. 2015); "Redewiedergabe. Eine literatur- und sprachwissenschaftliche Korpusanalyse", ein Projekt zur automatisierten Annotation von Figurenrede (<http://www.redewiedergabe.de> , vgl. auch Brunner 2015); die "Computational Stylistics Group" (<https://sites.google.com/site/computationalstylistics> , vgl. Rybicki/Eder/Hoover 2016); das "Rhythmicalizer"-Projekt (<http://www.rhythmicalizer.net> , vgl. Baumann/Meyer-Sickendiek 2016), das ein Tool zur Erkennung von Prosodie in freien Versen entwickelt; ebenso wie Projekte zur Analyse von Figurennetzwerken wie etwa "Digital Literary Network Analysis" (<https://dlina.github.io/about> , vgl. Fischer et al. 2017).

4. So betonen beispielsweise Kindt und Köppe, dass literaturwissenschaftliche Interpretationen weniger auf strukturelle Aspekte oder bloßes Sprachverstehen gerichtet sind als vielmehr auf mannigfaltige, komplexe und oft nicht eindeutig umrissene Verstehensziele (vgl. Kindt/Köppe 2008: 12–14).

5. So sind beispielsweise bestehende Verzeichnisse für Tools zur Textanalyse aufgrund der vornehmlich technisch ausgerichteten Beschreibungen für Geisteswissenschaftler ohne DH-Vorwissen äußerst schwer zugänglich (z.B. TAPoR, <http://www.tapor.ca>). Einen vielversprechenden Ansatz zur nutzbaren Aufbereitung und Vermittlung digitaler Methoden – allerdings vornehmlich in schulischen Kontexten – stellt das Projekt "Digitalität in den Fachdidaktiken" dar (<http://dhd-blog.org/?p=6812>).

6. So basieren zahlreiche DH-Praktiken zur Textanalyse auf der automatischen Verarbeitung von Textoberflächendaten wie Wortfrequenzen, beispielsweise stilometrische Untersuchungen oder Topic Modeling (vgl. Brett 2012).

7. Der Fragebogen bietet Nutzern zudem die Option, auf fehlende Antwort- oder sogar Fragemöglichkeiten hinzuweisen. Auf diese Weise kann der Fragebogen – und mit ihm forTEXTs Empfehlungssystem – im Laufe des Projekts weiter optimiert werden.

8. Affilierte Projekte sind beispielsweise "3DH" (<http://www.threedh.net>) und "SANTA: Shared Task on the Analysis of Narrative levels Through Annotation" (<https://sharedtasksinthedh.github.io/>), in deren Rahmen narrative Ebenen in Texten annotiert werden.

9. Ein Vorbild hierfür ist der bereits in CATMA integrierte "Query Builder", mithilfe dessen Nutzer komplexe Abfragen über Text- und Annotationsdaten laufen lassen können, ohne eine Abfragesprache lernen zu müssen.

10. Vgl. bspw. Bauer/Zirker 2015: Absatz 1.

11. So hat sich CATMA beispielsweise dem Konzept des *hermeneutischen Markups* verschrieben – darunter verstehen wir nach Piez Markup das bewusst interpretativ und flexibel ist (vgl. Piez 2010). Auf diese Weise erlaubt CATMA beispielsweise mehrfache und sogar "widersprüchliche" Annotationen derselben Textstelle – dennoch sind die Annotationen im standardisierten TEI-Format exportierbar und somit flexibel nachnutzbar. Diese Markup-Eigenschaften werden im Backend ermöglicht durch Standoff-Markup und die Nutzung von TEI Feature Structures (<http://www.tei-c.org/release/doc/tei-p5-doc/de/html/FS.html>).

12. Vgl. zu den Vorteilen von Klassifikationssystemen wie Taxonomien und Typologien auch Bailey 1994.

13. Zur Methode und zum Nutzen kollaborativen Annotierens, siehe auch Gius/Jacke 2015 und Gius/Jacke 2017.

Bibliographie

- Bailey, Kenneth D.** (1994): *Typologies and Taxonomies. An Introduction to Classification Techniques*. Thousand Oaks/London/New Delhi: Sage Publications.
- Bauer, Matthias / Zirker, Angelika** (2015): "Whipping Boys Explained. Literary Annotation and Digital Humanities", in: *MLA Commons. Literary Studies in the Digital Age* <https://dlsanthology.mla.hcommons.org/whipping-boys-explained-literary-annotation-and-digital-humanities/> [letzter Zugriff 06. September 2017].
- Baumann, Timo / Meyer-Sickendiek, Burkhard** (2016): "Large-scale Analysis of Spoken Free-verse Poetry", in: *Proceedings of LT4DH-Workshop 2016* http://edoc.sub.uni-hamburg.de/informatik/volltexte/2016/228/pdf/baumann_large_scale_analysis.pdf [letzter Zugriff 06. September 2017].
- Bögel, Thomas / Gertz, Michael / Gius, Evelyn / Jacke, Janina / Meister, Jan Christoph / Petris, Marco / Strötgen, Jannik** (2015): "Collaborative Text Annotation Meets Machine Learning. heureCLÉA, a Digital Heuristic of Narrative", in: *DHCommons Journal 1* <http://dhcommons.org/journal/issue-1/collaborative-text-annotation-meets-machine-learning-heureclea> [letzter Zugriff 06. September 2017].
- Brett, Megan R.** (2012): "Topic Modeling. A Basic Introduction", in: *Journal of Digital Humanities 2*(1) <http://journalofdigitalhumanities.org/2-1/topic-modeling-a-basic-introduction-by-megan-r-brett/> [letzter Zugriff 06. September 2017].
- Brunner, Annalen** (2015): *Automatische Erkennung von Redewiedergabe*. Berlin / Boston: de Gruyter (= Narratologia Bd. 47).
- Fiedler, Maik / Weiß, Andreas** (2015): "Von Daten zu Erkenntnissen. Digitale Geisteswissenschaften als Mittler zwischen Information und Interpretation.

DHd-Jahrestagung 2015” <http://www.hsozkult.de/conferencereport/id/tagungsberichte-6059> [letzter Zugriff 06. September 2017].

Fischer, Frank / Göbel, Mathias / Kampkaspar, Dario / Kittel, Christopher / Trilcke, Peer (2017): “Network Dynamics, Plot Analysis. Approaching the Progressive Structuration of Literary Texts”, in: *Digital Humanities 2017. Conference Abstracts* <https://dh2017.adho.org/abstracts/DH2017-abstracts.pdf> [letzter Zugriff 06. September 2017].

Gius, Evelyn / Jacke, Janina (2015): “Informatik und Hermeneutik. Zum Mehrwert interdisziplinärer Textanalyse”, in: *Zeitschrift für digitale Geisteswissenschaften*, Sonderband 1 http://www.zfdg.de/sb001_006 [letzter Zugriff 06. September 2017].

Gius, Evelyn / Jacke, Janina (2017): “The Hermeneutic Profit of Annotation. On Preventing and Fostering Disagreement in Literary Analysis”, in: *International Journal of Humanities and Arts Computing* 11(2) 233–254.

Kindt, Tom / Köppe, Tilmann (2008): “Einleitung”, in: dies. (eds.): *Moderne Interpretationstheorien. Ein Reader*. Göttingen: Vandenhoeck & Ruprecht 7–26.

Piez, Wendell (2010): “Towards Hermeneutic Markup. An Architectural Outline”, in: *Digital Humanities 2010. Conference Abstracts* <http://piez.org/wendell/papers/dh2010/> [letzter Zugriff 06. September 2017].

Rybicki, Jan / Eder, Maciej / Hoover, David (2016): “Computational stylistics and text analysis”, in: Crompton, Constance / Lane, Richard J. / Siemens, Ray (eds.): *Doing Digital Humanities*. London / New York: Routledge 123–144.

Spörl, Uwe (2004): *Basislexikon Literaturwissenschaft*. 2., durchges. Aufl. Paderborn [u.a.]: Schöningh.

Websites

3DH (<http://www.threedh.net>)

CATMA (<http://www.catma.de>)

Computational Stylistics Group (<https://sites.google.com/site/computationalstylistics>)

Digital Literary Network Analysis (<https://dlina.github.io/about>)

Digitalität in den Fachdidaktiken. Projektpräsentation im DHd-Blog (<http://dhd-blog.org/?p=6812>)

forTEXT (<http://www.fortext.net>)

heureCLÉA (<http://www.heureclea.de>)

Redewiedergabe (<http://www.redewiedergabe.de>)

Rhythmicalizer (<http://www.rhythmicalizer.net>)

SANTA: Shared Task on the Analysis of Narrative levels Through Annotation (<https://sharedtasksinthedh.github.io/>)

TAPoR (<http://www.tapor.ca>)

TEI Feature Structures (<http://www.tei-c.org/release/doc/tei-p5-doc/de/html/FS.html>)