

## Textannotation in der Lehre einsetzen: Ein Einstieg mit CATMA

### **Akazawa, Mari**

mari.akazawa@tu-darmstadt.de  
Technische Universität Darmstadt, Deutschland

### **Gerstorfer, Dominik**

dominik.gerstorfer@tu-darmstadt.de  
Technische Universität Darmstadt, Deutschland

### **Gius, Evelyn**

evelyn.gius@tu-darmstadt.de  
Technische Universität Darmstadt, Deutschland  
ORCID: 0000-0001-8888-8419

### **Guhr, Svenja**

svenja.guhr@tu-darmstadt.de  
Technische Universität Darmstadt, Deutschland

### **Häußler, Julian**

julian.haeussler@tu-darmstadt.de  
Technische Universität Darmstadt, Deutschland

### **Meister, Malte**

meister@linglit.tu-darmstadt.de  
Technische Universität Darmstadt, Deutschland

### **Messner, Stefanie**

s.messner@fu-berlin.de  
Technische Universität Darmstadt, Deutschland

### **Stiemer, Haimo**

haimo.stiemer@tu-darmstadt.de  
Technische Universität Darmstadt, Deutschland

### **von Keitz, Janis**

janis.vonkeitz@stud.tu-darmstadt.de  
Technische Universität Darmstadt, Deutschland

CATMA (Gius et al. 2024) ist eine webbasierte, kollaborative Textannotations- und Analyseplattform, die 2008 an der Universität Hamburg initiiert wurde und seit 2020 an der Technischen Universität Darmstadt kuratiert und weiterentwickelt wird. Hauptzielgruppe sind Geisteswissenschaftler\*innen ohne besondere technische Vorkenntnisse, die über eine intuitiv bedienbare GUI Texte annotieren und analysieren können.

Dieser Workshop richtet sich vor allem an Lehrende mit Vorkenntnissen in digitaler Annotation, die schon erste Erfahrungen mit CATMA gemacht haben und die im Rahmen ihrer universitären Lehre CATMA (und ggf. dessen Erweiterung GitMA) einsetzen möchten, um ihren Studierenden digitale Textannotation und Annotationsauswertung zu vermitteln. Teilnehmenden wird die Möglichkeit geboten, Anwendungsszenarien für die Integration von CATMA in der Lehre kennenzulernen und gemeinsam zu diskutieren.

## Manuelles und kollaboratives Annotieren mit CATMA im Lehrkontext

Das seit Jahrhunderten zu den textwissenschaftlichen Kernpraktiken gehörende Annotieren (vgl. Rapp 2017) lässt sich in sogenannte Highlights, Freitextkommentare sowie taxonomiebasierte Annotationen und Textauszeichnungen unterteilen, wobei die Übergänge häufig fließend sind (vgl. Jacke 2018, § 9). Während CATMA (Computer Assisted Text Markup and Analysis) auch die Möglichkeit für Highlights und Freitextkommentare bietet, ist die taxonomiebasierte Annotation der eigentliche Fokus des Tools — wobei die Taxonomie die Form von sogenannten Tagsets annimmt, denen für kollaborative Annotationsprojekte wahlweise Annotations-Guidelines beigegeben werden können (vgl. u.a. Bögel et al. 2015).

Die Einsatzmöglichkeiten von CATMA sind vielfältig: Neben der Literaturwissenschaft und Linguistik kann CATMA bspw. auch in den Geschichtswissenschaften zur Textquellenanalyse oder in der Soziologie oder Psychologie zur Annotation und qualitativen Auswertung von Interviews verwendet werden. CATMA hat sich als niedrigschwelliger Einstieg in die Digital Humanities in verschiedenen Lehrformaten bewährt, kann aber zum Beispiel auch zur kollaborativen Annotation während eines Hackathons oder ähnlicher Veranstaltungen eingesetzt werden.

Im Workshop werden wir besonders die Funktionalitäten in den Fokus nehmen, die in der Lehre eingesetzt werden: kollaboratives Annotieren und den Vergleich und die Analyse der von verschiedenen Annotator\*innen erzeugten Outputs. Außerdem werden wir eine Beispiel-Annotations-Guideline und vorbereitete Annotationen präsentieren, um die Möglichkeiten im Lehreinsatz aufzuzeigen.

## Von der Annotation zur Analyse: CATMA 7 und GitMA in der computergestützten Lehre

Obwohl der Umfang dessen, was mit dem CATMA-Frontend erreicht werden kann, recht groß ist, gab es in den letzten Jahren einige Neuerungen, die das Tool für die Digital Humanities-Praxis der Textanalyse noch interessanter gemacht haben.

Mit dem Release von CATMA 6 im Jahr 2019 wurde für die Plattform das auf Git basierende Backend (Git-Lab) eingeführt. Dieses ermöglicht unter anderem den direkten Zugriff auf Projektdaten. Für Nutzer\*innen mit fortgeschrittenen CATMA-Kompetenzen und für Interessierte aus der Digital-Humanities-Community mit Programmierkenntnissen und etwas Erfahrung mit Git eröffnet sich dadurch eine Reihe neuer Möglichkeiten, die es in bisherigen CATMA-Versionen nicht gab. Einige dieser Möglichkeiten werden im Laufe dieses Ganztagesworkshops vorgestellt und vermittelt.

Niedrigschwelligkeit und Nähe zu Methoden der Geisteswissenschaften sind nach wie vor wichtige Grundsätze, die in CATMA implementiert sind (vgl. Schumacher/Gius 2022). Doch mit zunehmender Verbreitung des Tools in den Digitalen Geisteswissenschaften sind neben der Möglichkeit der Textanalyse mit unterschiedlichen Methoden oder Interpretationstheorien (vgl. Piez 2010) auch die Einhaltung von Best Practices und Standards, die innerhalb der Digital-Humanities-Community entwickelt werden, von Bedeutung. Eine Verschmelzung von CATMA und dem Datenzugriff über Git zum Python-Package "GitMA" (Vauth et al. 2022) ermöglicht diese Erweiterung (wobei die übrig gebliebenen Buchstaben des CATMA-Akronyms, "M" für Markup und "A" für Analysis, absichtlich im Vordergrund stehen). Dabei bleibt der Annotationsprozess selbst frei gestaltbar. Die resultierenden Daten aber können anhand verschiedener vorgefertigter Visualisierungen (über die Visualisierungsmöglichkeiten in CATMA hinaus) mit GitMA veranschaulicht und miteinander verglichen werden.

Des Weiteren können die im Annotationsprozess erstellten Daten auf die Übereinstimmung zwischen den Annotator\*innen (sogenanntes *Inter-Annotator-Agreement*) mit verschiedenen Methoden ausgewertet werden (vgl. u.a. Artstein/Pesio 2008). Neben der Berechnung von gängigen Agreement-Metriken wie *Scott's pi*, *Cohen's kappa* und *Krippendorfs alpha* (im bekannten NLTK implementiert und von GitMA wiederverwendet) ist es möglich, die mittels verschiedener Parameter als übereinstimmend bewerteten Annotationen in eine neue Kollektion von Gold-Annotationen zu übertragen. Außerdem ist es möglich, die Outputs einer der Annotator\*innen als "silbernen Standard" festzulegen und die Annotationen der anderen an diesem zu messen. Die ermittelte Unstimmigkeit der Annotationen kann die Grundlage für ein Disagreement-Tagset bilden, das über das Backend auch wieder in die CATMA-GUI eingespeist werden kann. Gleiches gilt für die Passagen, die Unstimmigkeiten aufweisen, die wiederum durch Annotationen in der GUI dargestellt werden können.

Beide Workflows, die Bewertung der Übereinstimmung zwischen den Annotator\*innen und die Erstellung von Gold-Annotationen, werden im Workshop kurz präsentiert, wobei der Fokus auf konkrete Vermittlungsszenarien der Annotationsdatenanalyse mit GitMA im universitären Lehrkontext gelegt wird.

## Erprobte Lehrkonzepte

Nach der Beschäftigung mit den Möglichkeiten von CATMA und GitMA werden im Workshop bereits erprobte Lehrkonzepte vorgestellt. Dafür werden ausgewählte Beiträge des im Februar 2025 erscheinenden Themenheftes „Textannotation in der Hochschullehre“ der Online-Zeitschriftenreihe forTEXT herangezogen (vgl. Gerstorfer et al. 2024; siehe auch <https://www.fortext-hefte.de/>). Auf dieser Grundlage können die Teilnehmenden eigene annotationsbasierte Ansätze für Lehrveranstaltungen entwickeln.

## Lernziele

Die Teilnehmer\*innen lernen CATMA und GitMA als Software im Anwendungskontext von universitären Lehrveranstaltungen kennen. Sie erweitern ihr Vorwissen um konkrete Anwendungsszenarien in der Lehre und entwickeln in Gruppen Lehrveranstaltungsskizzen für den Einsatz von CATMA im Kontext der eigenen Lehre.

## Beitragende

Mitglieder des forTEXT labs, die in Forschung und Lehre mit Annotationen arbeiten:

Mari Akazawa, Dominik Gerstorfer, Evelyn Gius, Svenja Guhr, Julian Häußler, Malte Meister, Stefanie Messner, Haimo Stiemer, Janis von Keitz

## Format und Ablauf des Workshops

Der Workshop wird als Ganztagesworkshop angeboten.

Ablauf

9:00 - 9:45 Uhr

Vorstellungsrunde inkl. Abfrage bisheriger Erfahrungen mit CATMA v. a. in der Lehre (Aktivierungsphase, Kurzvorstellung und Mentimeter-Abfragen)

9:45 - 10:30 Uhr

Kurzeinführung in CATMA und in das Python-Package GitMA (theoretisch mit Anwendungsbeispielen) mit Frage-session

10:30 - 11:00 Uhr

Kaffeepause

11:00 - 12:30 Uhr

Reflexion und Ideensammlung im Plenum: Vorstellung und Diskussion konkreter Anwendungsszenarien von CATMA in der Lehre (45 Minuten zu Szenarien mit Studierenden auf Einstiegsniveau und 45 Minuten zu Szenarien mit Studierenden mit fortgeschrittenen Annotations-, CATMA- und GitMA-Kenntnissen)

12:30 - 14:00 Uhr

Mittagspause

14:00 - 14:15 Uhr

Zwischenplattform nach der Mittagspause (u.a. Mentimeter-Abfragen)

14:15 - 15:00 Uhr

Vorstellung ausgewählter Beiträge aus dem *forTEXT* Heft zu "Textannotation in der Hochschullehre": eingeladene Beiträger\*innen berichten: Was haben andere schon gemacht? Erfolgreiche Anwendungsfälle und Herausforderungen

15:00 - 15:30 Uhr

Teil 1: Ausprobieren und/oder Konzeption von Lehrszenarien, Entwicklung eigener Lehrveranstaltungsskizzen der Teilnehmenden mit Unterstützung des forTEXT-Teams, Sammeln von Wünschen und Ideen für Erweiterungen von CATMA zwecks Spezialisierung an den Lehrkontext

15:30 - 16:00 Uhr

Kaffeepause

16:00 - 16:45 Uhr

Teil 2: Ausprobieren und/oder Konzeption von Lehrszenarien, Entwicklung eigener Lehrveranstaltungsskizzen der Teilnehmenden mit Unterstützung des forTEXT-Teams, Sammeln von Wünschen und Ideen für Erweiterungen von CATMA zwecks Spezialisierung an den Lehrkontext

16:45 - 17:30 Uhr

Vorstellung der Ergebnisse, Abschlussdiskussion und Feedback

## Zielgruppe und Vorbereitung

Lehrende, die das Annotieren mit CATMA sowie die Auswertung von Annotationen im universitären Lehrkontext einbinden wollen und dafür konkrete Anwendungsszenarien kennenlernen und diskutieren möchten.

Zur Vorbereitung sollten Teilnehmer\*innen einen CATMA-Account erstellt (unter <https://app.catma.de/>) und sich mit der CATMA-Nutzung bekannt gemacht haben (z. B. mithilfe der forTEXT-Lerneinheit zu CATMA 6: Manuelle Annotation mit CATMA (<https://fortext.net/routinen/lerneinheiten/manuelle-annotation-mit-catma>) oder den aktualisierten Tutorials (<https://catma.de/how-to/tutorials/>) zu CATMA 7 auf [catma.de](https://catma.de)).

Wenn eigene CATMA-Annotationsdaten vorhanden sind, können diese während des Workshops analysiert werden. Teilnehmenden, die nicht mit eigenen Daten arbeiten möchten, wird ein Demo-Projekt zur Verfügung gestellt.

Für GitMA werden Vorinstallationen benötigt, die durch die Bereitstellung eines Docker-Images abgedeckt werden. Ein Link zum Docker-Image wird den Teilnehmer\*innen in der Woche vor dem Workshop per E-Mail zugesendet. Die Teilnehmer\*innen werden gebeten, die Installation von Docker als Vorbereitung auf den Workshop selbst auf einem eigenen Laptop (Touch-Devices werden nicht unterstützt), den sie zum Workshop mitbringen, vorzunehmen.

## Zahl der möglichen Teilnehmer\*innen

50

## Technische Voraussetzungen

Für die Durchführung des Workshops benötigen die Vortragenden einen Beamer und WLAN.

## Bibliographie

**Artstein, Ron und Massimo Poesio.** 2008. "Inter-Coder Agreement for Computational Linguistics." *Computational Linguistics* 34 (4): 555-96. 10.1162/coli.07-034-R2.

**Bögel, Thomas, Michael Gertz, Evelyn Gius, Janina Jacke, Jan Christoph Meister, Marco Petris und Jannik Strötgen.** 2015. "Collaborative Text Annotation Meets Machine Learning: heureCLÉA, a Digital Heuristic of Narrative". *DHCommons Journal* 1. 10.5281/zenodo.3240591.

**Gerstorfer, Dominik, Mari Akazawa und Evelyn Gius.** 2024. "forTEXT-Hefte: Eine Open-Access-Plattform für den Wissensaustausch in den digitalen Literaturwissenschaften." *DHd 2024 Book of Abstracts*. Zenodo. 10.5281/zenodo.10706174.

**Gius, Evelyn, Jan Christoph Meister, Malte Meister, Marco Petris, Dominik Gerstorfer und Mari Akazawa.** 2024. *CATMA (7.1.0)*. Zenodo. 10.5281/zenodo.1470118.

**Jacke, Janina.** 2018. "Manuelle Annotation." *forTEXT – Literatur digital erforschen*. <https://fortext.net/routinen/methoden/manuelle-annotation>.

**Piez, Wendell.** 2010. "Towards Hermeneutic Markup. An Architectural Outline." *ADHO 2010 Book of Abstracts*. 202–205. <http://dh2010.cch.kcl.ac.uk/academic-programme/abstracts/papers/html/ab-743.html>.

**Rapp, Andrea.** 2017. "Manuelle und automatische Annotation." *Digital Humanities: Eine Einführung*, 253–67. Stuttgart: J.B. Metzler Verlag, 2017.

**Schumacher, Mareike und Evelyn Gius.** 2022. "forTEXT.net – Literatur digital erforschen." *Mitteilungen des Deutschen Germanistenverbandes* 69 (2): 121–126. 10.14220/mdge.2022.69.2.121.

**Vauth, Michael, Malte Meister, Hans Ole Hatzel, Dominik Gerstorfer und Evelyn Gius.** 2022. *GitMA (1.4.9)*. Zenodo. 10.5281/zenodo.6330464.