

Vom Begriff über das Phänomen zur Analyse

Ein CRETA-Workshop zur Operationalisierung in den DH

Andresen, Melanie

melanie.andresen@ims.uni-stuttgart.de
University of Stuttgart, Germany

Krautter, Benjamin

Benjamin.Krautter@gs.uni-heidelberg.de
Heidelberg University, Germany

Pagel, Janis

janis.pagel@ims.uni-stuttgart.de
University of Stuttgart, Germany

Pichler, Axel

axel.pichler@fu-berlin.de
FU Berlin, Germany

Der Workshop stellt eine weiterentwickelte und personell anders besetzte Version des auf der DHd 2020 abgehaltenen, beinahe gleichnamigen Workshops dar. Er adressiert eine der zentralen Herausforderungen für Arbeiten in den Digital Humanities – die Operationalisierung geisteswissenschaftlicher Konzepte und Fragestellungen für computergestützte Forschungsansätze (vgl. Jannidis 2010: 109–132; Moretti 2013; Flanders/Jannidis 2015; Jacke 2014: 118–139; Pichler/Reiter 2020, Pichler/Reiter 2021). Während Geisteswissenschaftler*innen vor allem mit komplexen, häufig mehrere Textphänomene umfassenden Konzepten arbeiten und als relevant erachtete Kontexte zu deren Deutung heranziehen, ist die computergestützte Arbeit an identifizierbare Phänomene auf der Textoberfläche gebunden. Die hieraus erwachsende Diskrepanz zwischen theoretischen Erwartungen und konkreten Ergebnissen gilt es über eine adäquate Operationalisierung zu überbrücken (vgl. Moretti 2013: 1). Ziel ist es also, Verfahren zu entwickeln, die theoretische Begriffe über potenziell mehrere Teilschritte auf Textoberflächenphänomene zurückführen. Oder kurz gesagt: **die Erkennung und Messbarmachung von Instanzierungen theoretischer Konzepte**. Mit unserem Workshop wollen wir genau diese Schnittstelle in den Fokus rücken. Anhand ausgewählter Anwendungsfälle zeigen wir, welche Herausforderungen sich aus dem Einsatz computergestützter Methoden für geisteswissenschaftliche Fragestellungen ergeben und wie mit ihnen umgegangen werden kann. In einem praktischen Teil haben die Teilnehmenden die Möglichkeit, an der Operationalisierung vorgegebener exemplarischer Fragestellungen der Textanalyse und der dafür relevanten Konzepte zu arbeiten. Hierfür stellen wir Anwendungsfälle mit geeigneten Tools und Technik-„Baukästen“ zur Verfügung. Programmierkenntnisse werden dabei nicht vorausgesetzt. Ziel des Workshops ist es, das Bewusstsein für die Differenzen zwischen geisteswissenschaftlicher und computergestützter Arbeitsweise zu schärfen, typische Herausforderungen zu adressieren und Herangehensweisen zur Operationalisierung geisteswissenschaftlicher Konzepte aufzuzeigen. Denn

nur durch die reflektierte Auseinandersetzung mit den Operationalisierungsannahmen kann ein angemessener Umgang mit den Ergebnissen gewährleistet werden.

Use Cases

Als Anwendungsfälle stellen wir Phänomene vor, zu denen wir im Rahmen des „Center for Reflected Text Analytics“ e.V. (CRETA) ¹ umfangreiche Erfahrungen gesammelt haben. Die gewählten Beispiele decken verschiedene Aufgabentypen ab: Wir behandeln erstens die Extraktion bestimmter Instanzen aus einem Text und zweitens ein holistisches Textphänomen.

Entitäten und Entitätenreferenzen

In einem ersten Anwendungsfall befassen wir uns mit dem Konzept der Entität und ihrer Referenz in literarischen Texten (vgl. Reiter u.a. 2017: 19–22; Blessing u.a. 2020). Dabei fassen wir den Begriff der Entität sehr weit: „Alles, was man als Einheit denken kann, kann als Entität behandelt werden“ (Jannidis 2017: 103). Zu den Entitäten zählen dementsprechend Personen/Figuren, Orte, Organisationen sowie Ereignisse. Das Konzept ist also für verschiedene Forschungsfragen anschlussfähig. Auf Entitäten kann auf verschiedene Weise referiert werden, etwa über Eigen- und Gattungsnamen (z. B. „Angela Merkel“, „die Kanzlerin“). Um Entitäten in einem Text zu extrahieren, müssen folglich die Entitätenreferenzen annotiert und kookkurrente Ausdrücke aufgelöst werden. Die Herausforderungen bestehen vor allem in der Festlegung der Referenzausdrücke (welche Ausdrücke werden berücksichtigt?), in der Abgrenzung von Entitätenreferenzen gegenüber generischen Ausdrücken sowie im Umgang mit Verschachtelungen, Metonymien und textspezifischen Besonderheiten.

Protagonisten im Drama

Der zweite Anwendungsfall setzt sich mit der Identifikation von Protagonisten im Drama auseinander, fokussiert also ein holistisches Textphänomen. Die verschiedenen Perspektiven der Literaturwissenschaft auf Protagonisten, Hauptfiguren und Helden von Dramen (vgl. die Ausführungen in Krautter u.a. 2018: 6–16 und Wulff 2002: 431–448) haben zur Folge, dass eine Reihe von Definitionen und Identifikationsstrategien koexistieren, die häufig an historische Normvorstellungen geknüpft sind. Diese historische Gebundenheit erschwert die operationale Definition von Protagonisten, wenn man auf größere Abschnitte der Literaturgeschichte blickt.

Direkt anschlussfähig für die Methoden der Digital Humanities erscheint die in den späten 1970er Jahren von Manfred Pfister skizzierte Annahme, dass „quantitative[] Dominanzrelationen“ (Pfister 2001: 227) hilfreich für die Differenzierung des Bühnenpersonals seien. Pfister nennt zwei Kriterien, die dabei helfen können, dramatische Figuren schon aufgrund quantitativer Eigenschaften als Haupt- oder Nebenfiguren zu identifizieren: nämlich die Zeitdauer, die sie auf der Bühne stehen, und ihr Anteil an der gesamten Figurenrede (vgl. Pfister 2001: 226–227). Diese Auffassung Pfisters lässt sich mit digitalen Methoden der Dramenanalyse um weitere Eigenschaften der Figuren, etwa durch Netzwerkmetriken oder Topic Modeling, zu einem multidimensionalen Ansatz ergänzen. Die größte Herausforderung stellt hierbei die Validie-

rung der Ergebnisse dar, da diese an die Gültigkeit der operationalen Definition für die manuelle Annotation gebunden ist.

Ansätze zur Operationalisierung

Im Workshop stellen wir zwei Ansätze zur Operationalisierung vor, die sich – in verschiedenen Phasen des Forschungsprozesses – sehr gut gegenseitig ergänzen. Der erste Ansatz besteht in der **operationalen Definition theoretischer Konzepte durch manuelle Annotationen**. Die Ergebnisse sind also keine Skripte oder Funktionen, sondern klare(re), als Abfolge von Operationen bestimmte Definitionen der fraglichen Konzepte, die von Menschen mit hoher intersubjektiver Übereinstimmung umgesetzt werden, und zudem die theoretische Diskussion bereichern können (vgl. Gius/Jacke 2017; Pagel u.a. 2020; Reiter 2020, Pichler/Reiter 2021). Daneben führt der Annotationsprozess auch zu einer intensiven und kritischen Beschäftigung mit den Texten und den textuellen Indikatoren des Konzeptes und liefert damit auch Ideen für eine computergestützte Operationalisierung.

Als zweiten Ansatz stellen wir eine Vorgehensweise vor, die Zielflächen **indirekt operationalisiert**. Da sich viele geisteswissenschaftliche Konzepte nicht direkt messen lassen, sie also nicht unmittelbar durch mögliche Instanzen auf der Textoberfläche repräsentiert werden, müssen die Konzepte schrittweise auf messbare Eigenschaften zurückgeführt werden. In diesem Fall werden also mehrere messbare Eigenschaften in den Blick genommen, die mit dem Zielkonzept verwandt, aber nicht deckungsgleich sind und das Konzept somit indirekt operationalisieren. Aufschlussreich ist dabei in erster Linie die Gesamtschau der verschiedenen als relevant erachteten Einflussfaktoren (vgl. „instrumental variables“ in Sack 2011; „indirekte Operationalisierung“ in Reiter/Willand 2018). Bei textbasierten Phänomenen lassen sich so insbesondere linguistische und strukturelle Eigenschaften, die größtenteils mit hoher Reliabilität automatisch extrahierbar sind, in die Operationalisierung integrieren.

Ablauf

In einem Theorieteil führen wir in Geschichte und Praxis der Operationalisierung von geisteswissenschaftlichen Fragestellungen und Konzepten für die computergestützte Analyse ein. Anhand der oben genannten Beispiele aus der CRETA-Praxis thematisieren wir die Problematik und stellen Ansätze zur Operationalisierung im Detail vor. Je nach Interesse kann anschließend, im praktischen Teil, einer dieser Anwendungsfälle ausgewählt und bearbeitet werden. Dabei haben die Teilnehmenden die Möglichkeit, beide Operationalisierungsansätze an ihrem gewählten Anwendungsfall zu erproben. Hierfür befassen sie sich zunächst mit dem Konzept, indem sie es anhand eines Textauszugs manuell annotieren und parallel stichpunktartig die Richtlinien schärfen. In einer ersten Diskussionsrunde werden die verschiedenen Ergebnisse gesammelt und diskutiert. Zur Erprobung des zweiten Ansatzes stellen wir für jeden Anwendungsfall einen Operationalisierungs-„Baukasten“ vor. Dieser besteht aus einer Sammlung von Python-Skripten in einem Jupyter-Notebook, das auf das jeweilige Untersuchungsvorhaben zugeschnitten ist und den Teilnehmenden die Möglichkeit gibt, sich dem zu untersuchenden Phänomen über computergestützte Verfahren anzunähern. Die Teilnehmenden können in Kleingruppen in diesem Baukasten verschiedene Parameter einstellen sowie manuell Eigenschaften an- oder abwählen, wobei sie auf ihr Vorwissen über

den Untersuchungsgegenstand aus der ersten Praxisrunde zurückgreifen können. Nachdem die Teilnehmenden die Eigenschaften ausgewählt und ggf. parametrisiert haben, können sie die Ergebnisse visualisieren und mit den Texten abgleichen. Damit erhalten die Teilnehmenden ein direktes Feedback zu den ausgewählten Parametern und können prüfen, ob das Untersuchungsvorhaben mit den festgelegten Einstellungen angemessen umgesetzt wird. Der Baukasten ist zur iterativen Nutzung vorgesehen, sodass der Einfluss verschiedener verwandter Eigenschaften auf die Ausgaben sichtbar wird und die Teilnehmenden sich einer geeigneten technischen Umsetzung sukzessiv annähern können. In einer abschließenden Diskussion werden die Ergebnisse gesammelt und es wird ausgewertet, wie adäquat sich die jeweiligen Zielflächen mittels der gewählten Annahmen abbilden haben lassen.

Lernziele

Ziel unseres Workshops ist es, die Teilnehmenden für die Wichtigkeit der Operationalisierung in den Digital Humanities zu sensibilisieren und ihnen Wege zu ihrer erfolgreichen Realisierung vorzustellen. Durch die interdisziplinäre Ausrichtung von DH-Arbeiten kommt der Operationalisierung eine Schlüsselposition zu, da sie eine Brücke zwischen geisteswissenschaftlichen Konzepten und computergestützter Umsetzung schlägt (vgl. Moretti 2013: 1). Mit den gewählten Anwendungsfällen wollen wir den Teilnehmenden ein „Repertoire“ für die Operationalisierung verschiedener Aufgabentypen mitgeben. Wir zeigen zum einen, dass die Annotation eines Phänomens als Methode seiner Operationalisierung dienen kann (vgl. Gius/Jacke 2017: 233–254); zum anderen führen wir für textbasierte Phänomene eine indirekte Operationalisierung ein (vgl. Reiter/Willand 2018). Beide Verfahrensweisen sind auf andere Anwendungsfälle übertragbar. Gleichzeitig möchten wir deutlich machen, dass es für jedes Untersuchungsvorhaben nicht nur eine, sondern verschiedene Wege der Operationalisierung gibt. Die Spielräume, die bei der Operationalisierung geisteswissenschaftlicher Fragestellungen entstehen, machen es notwendig, Entscheidungen reflektiert zu treffen, sie offenzulegen und ihren Einfluss auf die Ergebnisse als Voraussetzung für eine angemessene Interpretation zu bedenken.

Anhang

Zeitplan

- (insgesamt 3 Stunden + 30 Min. Pause)
- 1. Einführung und Ablauf (10 Min.)
- 2. Theoretischer Teil (insgesamt 30 Min.)
 - Erläuterung der Problemstellung
 - Vorstellung der Anwendungsfälle
- 3. Praktischer Teil
 - Einführung in die Primärtexte und Tools, Ausgabe der skizzierten Guidelines (10 Min.)
 - Erste Praxisrunde (Kleingruppen): Manuelle Annotation eines Phänomens, parallele Erweiterung/Überarbeitung der Guidelines, iterativ (30–40 Min.)
 - Kaffeepause (30 Min.) –
 - Sammeln der Ergebnisse und Diskussion der Herangehensweisen (20 Min.)
 - Zweite Praxisrunde (Kleingruppen): Arbeit am Operationalisierungsbaukasten, Feedback über Ausgabedatei, iterativ (30–40 Min.)

4. Abschlussdiskussion: Sammeln der Ergebnisse, Diskussion der Erfahrungen und Lernziele (30 Min.)

Die Durchführung des Workshops auf der DHd 2020 hat gezeigt, dass das gestraffte dreistündige Format gute didaktische Resultate zeitigt. Der Fokus auf die praktischen Dimensionen der Operationalisierung ist dabei gewollt: Aus der konkreten praktischen Arbeit heraus lässt sich unserer Ansicht nach am besten der theoretische Rahmen und die theoretischen Probleme bei der Operationalisierung reflektieren.

Zahl der möglichen Teilnehmenden

Zwischen 15 und 25

Angaben zur technischen Ausstattung

Abgesehen von Beamer und ausreichend Steckdosen ist keine besondere technische Ausstattung erforderlich. Die Teilnehmenden arbeiten im praktischen Teil an ihrem eigenen Laptop. Informationen zu eventuellen Vorab-Installationen werden rechtzeitig mitgeteilt.

Beitragende

Der Workshop wird von Mitgliedern des Center for Reflected Text Analytics (CRETA) e.V. veranstaltet, die bereits erfahrene Workshop-Leiter*innen im DH-Bereich sind (DHd 2017, DHd 2018, ESU 2018, DHd 2019, HCH 2019, DHd 2020).

CRETA konzentriert sich auf die Entwicklung von Methoden zur kritisch-reflektierten Textanalyse im Forschungsbereich der Digital Humanities. Die Methoden werden fachübergreifend für textanalytische Fragestellungen aus der Literatur-, Sprach-, Geschichts- und Sozialwissenschaft sowie Philosophie erarbeitet und eingesetzt. Das bis 2020 vom BMBF geförderte eHumanities-Zentrum ist Ende 2020 mit der Gründung eines Vereines in eine neue Phase übergegangen. Mit der Vereinsgründung wird der Tatsache Rechnung getragen, dass über Stuttgart hinaus inzwischen Wissenschaftler*innen an ähnlichen Zielen arbeiten und CRETA in vielfältiger Weise verbunden sind, etwa durch gemeinsame Projekte.

Melanie Andresen
melanie.andresen@ims.uni-stuttgart.de
Universität Stuttgart
Institut für Maschinelle Sprachverarbeitung
Pfaffenwaldring 5b
70569 Stuttgart

Melanie Andresen ist Postdoc am Institut für Maschinelle Sprachverarbeitung an der Universität Stuttgart. Sie hat Germanistische Linguistik an der Universität Hamburg studiert und ist dort 2020 im Bereich der Korpuslinguistik promoviert worden. Aus den Projekten *hermA* (Universität Hamburg) und *Q:TRACK* (Universität Stuttgart, Universität zu Köln) bringt sie viel Erfahrung mit der Operationalisierung geistes- und sozialwissenschaftlicher Fragestellungen mit.

Benjamin Krautter
Benjamin.Krautter@uni-koeln.de
Universität zu Köln
Institut für Digital Humanities
Albertus-Magnus-Platz
50931 Köln

Benjamin Krautter ist Promotionsstudent am Germanistischen Seminar der Universität Heidelberg und Mitarbeiter im Projekt *Q:TRACK*. Dort arbeitet er u. a. an der Operationalisierung literaturwissenschaftlicher Kategorien für die quantitative Dramenanalyse. Im Zentrum seines Forschungsinteresses steht dabei die mögliche Verbindung quantitativer und qualitativer Methoden für die Analyse und Interpretation literarischer Texte.

Janis Pagel
janis.pagel@uni-koeln.de
Universität zu Köln
Institut für Digital Humanities
Albertus-Magnus-Platz
50931 Köln

Janis Pagel ist Promotionsstudent am Institut für Maschinelle Sprachverarbeitung der Universität Stuttgart und Mitarbeiter am Institut für Digital Humanities der Universität zu Köln. Er studierte Germanistik und Linguistik in Bochum, sowie Computerlinguistik in Stuttgart und Amsterdam. Er forscht zu Anwendungen von computerlinguistischen Methoden auf literaturwissenschaftliche Fragestellungen und Koreferenzresolution auf literarischen Texten.

Axel Pichler
axel.pichler@ts.uni-stuttgart.de
Universität Stuttgart
Institut für Maschinelle Sprachverarbeitung
Pfaffenwaldring 5b
70569 Stuttgart

Axel Pichler studierte Philosophie und Germanistik in Wien und Graz. Im Sommersemester 2021 war er Gastprofessor für Digital Humanities am EXC "Temporal Communities" der FU Berlin. Zurzeit arbeitet er als Postdoc unter anderem an der Entwicklung und Reflexion von Methoden der computergestützten Textanalyse am Institut für Maschinelle Sprachverarbeitung der Universität Stuttgart.

Fußnoten

1. <https://cretaverein.de/de>

Bibliographie

Blessing, André / Echelmeyer, Nora / John, Markus / Reiter, Nils (2017): „An end-to-end environment for research question-driven entity extraction and network analysis“, in: *Proceedings of the Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, Vancouver.

Flanders, Julia / Jannidis, Fotis (2015): *Knowledge Organization and Data Modeling in the Humanities*. https://opus.bibliothek.uni-wuerzburg.de/opus4-wuerzburg/frontdoor/deliver/index/docId/11127/file/flanders_jannidis_datamodeling.pdf

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, S. 233–254.

Jacke, Janina (2014): „Is There a Context-Free Way of Understanding Texts? The Case of Structuralist Narratology“, in: *Journal of Literary Theory* 8, S. 118–39.

Jannidis, Fotis (2017): „Grundlagen der Datenmodellierung“, in: *Digital Humanities. Eine Einführung*, hg. v. Fotis Jannidis, Hubertus Kohle und Malte Rehbein, Stuttgart, S. 99–108.

Jannidis, Fotis (2010): „Methoden der computergestützten Textanalyse“, in: *Methoden der literatur- und kulturwissenschaftlichen Textanalyse. Ansätze – Grundlagen – Modellanalysen*, hg. v. Vera Nünning, Ansgar Nünning und Irina Bauer-Begerow, Stuttgart, Weimar, S. 109–132.

Ketschik, Nora / Overbeck, Maximilian / Murr, Sandra / Pichler, Axel und André Blessing (2020): „Interdisziplinäre Annotation von Entitätenreferenzen“, in: *Reflektierte algorithmische Textanalyse*, hg. v. Nils Reiter, Axel Pichler und Jonas Kuhn, Berlin. <https://doi.org/10.1515/9783110693973-010>

Krautter, Benjamin / Pagel, Janis / Reiter, Nils / Willand, Marcus (2018): „Titelhelden und Protagonisten – Interpretierbare Figurenklassifikation in deutschsprachigen Dramen“, in: *Litlab Pamphlet 7*.

Moretti, Franco (2013): „Operationalizing”: or, the function of measurement in modern literary theory“, in: *Literary Lab 6*, S. 1–13.

Pagel, Janis / Reiter, Nils / Rösiger, Ina / Schulz, Sarah (2020): „Annotation als flexibel einsetzbare Methode“, in: *Reflektierte algorithmische Textanalyse*, hg. v. Nils Reiter, Axel Pichler und Jonas Kuhn, Berlin. <https://doi.org/10.1515/9783110693973-006>

Pichler, Axel / Reiter, Nils (2020): „Reflektierte Textanalyse“, in: *Reflektierte algorithmische Textanalyse*, hg. v. Nils Reiter, Axel Pichler und Jonas Kuhn, Berlin. <https://doi.org/10.1515/9783110693973-003>

Pichler, Axel / Reiter, Nils (2021): „Zur Operationalisierung literaturwissenschaftlicher Begriffe in der algorithmischen Textanalyse. Eine Annäherung über Norbert Altenhofers hermeneutische Modellinterpretation von Kleists Das Erdbeben in Chili“, in: *Journal of Literary Theory 15*, S. 1–29. <https://doi.org/10.1515/jlt-2021-2008>

Pfister, Manfred (2001): *Das Drama. Theorie und Analyse*, München.

Reiter, Nils (2020): „Anleitung zur Erstellung von Annotationsrichtlinien“, in: *Reflektierte algorithmische Textanalyse*, hg. v. Nils Reiter, Axel Pichler und Jonas Kuhn, Berlin. <https://doi.org/10.1515/9783110693973-009>

Reiter, Nils / Gius, Evelyn / Willand, Marcus (Hg.) (2019): *A Shared Task for the Digital Humanities. Special issue of Cultural Analytics*. November 2019.

Reiter, Nils / Blessing, André / Echelmeyer, Nora / Kremer, Gerhard / Koch, Steffen / Murr, Sandra / Overbeck, Maximilian / Pichler, Axel (2017): CUTE: CRETA Unshared Task zu Entitätenreferenzen“, in: *DHd 2017 Bern*, Conference Abstracts, S. 19–22.

Reiter, Nils / Willand, Marcus (2018): „Poetologischer Anspruch und dramatische Wirklichkeit: Indirekte Operationalisierung in der digitalen Dramenanalyse“, in: *Quantitative Ansätze in den Literatur- und Geisteswissenschaften: Systematische und historische Perspektiven*, hg. v. Toni Bernhart, Marcus Willand, Sandra Richter und Andrea Albrecht, Stuttgart, S. 45–76.

Sack, Graham Alexander (2011): „Simulating Plot: Towards a Generative Model of Narrative Structure“, in: *Papers from the AAAI Fall Symposium (FS-11-03)*.

Wulff, Hans Jürgen (2002): „Held und Antiheld, Prot- und Antagonist: Zur Kommunikations- und Texttheorie eines komplizierten Begriffsfeldes. Ein enzyklopädischer Aufriß“, in: *Weltentwürfe in Literatur und Medien. Phantastische Wirklichkeiten realistische Imaginationen. Festschrift für Marianne Wünsch*, hg. v. Hans Krah und Claus-Michael Ort, Kiel, S. 431–448.