Nils Reiter und Axel Pichler

CRETA, ein erstes Fazit

Herausforderungen, Erfolge und Empfehlungen

Das Stuttgarter *Center for Reflected Text Analytics* (CRETA) entwickelt textanalytische Methoden für die digitalen Geisteswissenschaften. Wie die Beiträge dieses Bandes zeigen, führt diese Zielsetzung weder zur Entwicklung einer explizierten, elaborierten und logisch konsistenten Interpretationstheorie – wie man sie aus den traditionellen Geisteswissenschaften, insbesondere jedoch aus den Literaturwissenschaften kennt, – noch zur Entwicklung neuer Textanalysetools.¹ Kernergebnis von CRETA ist stattdessen die Etablierung und theoretische Reflexion disziplinenunabhängiger regelgeleiteter Arbeitsablaufpraktiken (= Methoden).² Am Ende dieses Buches wollen wir die Herausforderungen, die mit einer solchen Zielsetzung einhergehen, und die Resultate, die unser Umgang mit diesen Herausforderungen gezeitigt haben, resümieren. Damit sollen zukünftigen Projekten, die CRETA in ihrem disziplinären und institutionellen Aufbau ähnlich sind oder vergleichbare Zielsetzungen verfolgen, hilfreiche Strategien auf den Weg mitgegeben werden.

Eine zu erwartende, jedoch trotzdem in ihrer Bedeutung nicht zu unterschätzende Herausforderung stellte für CRETA die interdisziplinäre Zusammenarbeit dar (siehe dazu auch die Einleitung des Bandes sowie den Beitrag zur Interdisziplinarität ab Seite 467). Deren Spezifika – insbesondere die Notwendigkeit der Entwicklung einer gemeinsamen Sprache oder zumindest eines Bewusstseins für mögliche ‚falsche Freunde‘ – wurden von Anfang an bei der Planung berücksichtigt, was sich unmittelbar auf die konkrete Forschungsarbeit auswirkte. So wurden zum Beispiel die zweiwöchigen Arbeitstreffen im ersten Förderjahr von CRE-

1 Solche wurden zwar ebenfalls im Rahmen von CRETA entwickelt, siehe zum Beispiel das Annotationstool CRETAnno (S. 207), allerdings eher als *proof of concept* denn als Bereitstellung digitaler Infrastruktur.

2 CRETA folgt dabei einem Verständnis von Algorithmus, nach dem es sich bei einem solchen um eine Handlungsvorschrift zur Beantwortung einer Frage bzw. zur Lösung eines Problems handelt. Die Handlungsvorschriften zur Lösung textanalytischer Probleme können sowohl in Code implementiert sein als auch ‚manuell‘, d. h. von Menschen, realisiert werden (siehe hierzu auch den Beitrag von Pichler und Reiter in diesem Band).

Nils Reiter, Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart

Axel Pichler, Stuttgart Research Center for Text Studies, Universität Stuttgart

TA zur gemeinsamen Arbeit an für alle Beteiligten relevanten Fragestellungen genutzt, was wesentlich zur Etablierung einer gemeinsamen Sprache beitrug.³ Daneben wurden weitere Strategien entwickelt und umgesetzt, die den Dialog zwischen den CRETA-Mitarbeiter*innen förderten und vertieften, die wir weiter unten noch ausführen. Eine Konsequenz dieses Vorgehens war, dass es insbesondere am Anfang des Förderzeitraums zu einer ‚Entschleunigung‘ der individuellen Forschungsarbeit kam: Nicht die zügige individuelle Beantwortung von fachspezifischen Fragen standen im Mittelpunkt der interdisziplinären Aktivitäten, sondern die Erarbeitung eines gemeinsamen Fundaments. Darauf aufbauend konnten dann die fachspezifischen Fragen in Fragen überführt werden, die dem aktuellen Stand der digitalen Geisteswissenschaften entsprechen. Derartige Aktivitäten nehmen viel Zeit in Anspruch. Die temporäre ‚Entschleunigung‘ wird aber durch ein gemeinsam erarbeitetes interdisziplinäres Wissen ausgeglichen, was letztendlich – so unser Eindruck – zu einem schnelleren Arbeiten nach der Einarbeitungsphase beiträgt. Gleichzeitig wird dabei dem ‚*scheduling conflict*‘ entgegengewirkt, der in DH-Projekten leicht auftritt: Während der Entwicklung von Textanalyse-Tools herrscht auf geisteswissenschaftlicher Seite viel Leerlauf – und wenn die Tools fertig sind, ist häufig ein Großteil der Projektlaufzeit vorbei (siehe dazu auch die Einleitung dieses Bandes). Durch eine frühzeitige enge Kooperation von Computer- und Geisteswissenschaften verkürzen sich diese Entwicklungszyklen.⁴

Auswirkungen hat das interdisziplinäre Arbeiten auch auf die Publikationspraxis der Beteiligten. Dies betrifft sowohl die Publikationsarten als auch die Publikationsfrequenz. In Hinblick auf den Modus sind zum Beispiel in den Geisteswissenschaften Beiträge mit mehr als einer*m Verfasser*in sehr selten, in den Computerwissenschaften sind sie hingegen üblich. Ebenfalls unüblich ist es in den Geisteswissenschaften, Beiträge zu publizieren, die sich ausschließlich mit der Entwicklung von Methoden beschäftigen,⁵ in vielen Computerwissenschaften hingegen wird ausschließlich über Methoden publiziert und auch in den Digital Humanities sind derartige Beiträge zunehmend verbreitet.⁶ Hinzu kommt, dass es prinzipiell nicht üblich ist Zwischenresultate von Qualifikationsprojekten zu

³ Resultat dieser Zusammenarbeit sind unter anderem die CRETA-Richtlinien zur Entitätenauszeichnung, die in diesem Band ab Seite 204 besprochen werden.

⁴ Siehe hierzu auch Seite 81 ff.

⁵ Jenseits der Wissenschaftssoziologie und -theorie werden Methoden in den Geisteswissenschaften zumeist im Zusammenhang mit den ihnen zugrunde liegenden Theorien verhandelt, nur sehr selten jedoch *per se*.

⁶ Die Tatsache dass Digital-Humanities-Beiträge sich mehr und mehr methodisch orientieren, liegt sicher auch daran, dass die Methodik bzw. das ‚Digitale‘ das ist, was die DH verbindet.

publizieren, viele der CRETA-Mitwirkenden jedoch an solchen arbeite(te)n. Bereits diese wenigen Beispiele zeigen, dass auch das Publizieren von ‚Ergebnissen‘ interdisziplinärer Forschendengruppen eine Herausforderung darstellt, unter anderem auch weil in jedem Fach unterschiedliche Vorstellungen darüber herrschen, was denn ein Forschungsergebnis ist. Erschwerend kommt hinzu, dass es im deutschsprachigen Raum kaum wissenschaftlich etablierte Fachzeitschriften für die digitalen Geisteswissenschaften gibt,⁷ eine Publikation von DH-Beiträgen in den fachspezifischen Organen jedoch selten versucht wird. Dies mag Folge einer Erfahrung sein, die auch ein Großteil der CRETA-Mitarbeitenden machen musste: Die Präsentation von Erkenntnissen, die mithilfe digitaler Tools gewonnen wurden, wird in nicht DH-affinen, traditionelleren fachspezifischen Kontexten – wenn überhaupt – nur sehr kritisch zur Kenntnis genommen. Wie eine diesbezügliche Umfrage im CRETA-Kontext gezeigt hat, reichen die diesbezüglichen Kommentare von „Das ist interessant, aber uns als Vertreter*innen des Faches interessiert eigentlich etwas anderes.“ über „Das wissen wir doch schon alles.“ bis zu Fragen wie „Ist das nicht trivial?“. Auch der vergleichsweise große Fokus auf die Methoden – die unseres Erachtens in jedem Fall Teil solcher Beiträge sein müssen, wie wir z. B. ab Seite 43 dargelegt haben – führt seitens der Herausgebenden und Gutachtenden häufig zu Irritationen.

Trotz dieser Herausforderungen, vor denen viele DH-Projekte stehen, sehen wir CRETA als Erfolg an. Ein Indikator dafür mag das reichhaltige ‚Ökosystem‘ sein, das sich um CRETA herum gebildet hat: Der interdisziplinäre Austausch in CRETA hat zu einer Reihe von weiteren Projekten geführt, die inhaltlich und strukturell von CRETA profitier(t)en, und für deren Durchführung teilweise auch Mittel eingeworben werden konnten. Wir stellen hier eine Auswahl vor:⁸

- Schwerpunktprogramm *Computational Literary Studies* (**SPP CLS**). Gefördert durch die Deutsche Forschungsgemeinschaft (DFG). Projektleitung: Fotis Jannidis (Computerphilologie, Universität Würzburg), Evelyn Gius (Digital Philology, TU Darmstadt), Jonas Kuhn (Computerlinguistik), Nils Reiter (Computerlinguistik), Christof Schöch (Digital Humanities, Universität Trier), Simone Winko (Literaturwissenschaft, Universität Göttingen).
URL: <https://dfg-spp-cls.github.io>
- Modelling Argumentation Dynamics in Political Discourse (**MARDY**). Gefördert durch die DFG (Schwerpunktprogramm RATIO). Projektleitung: Jonas Kuhn (Computerlinguistik), Sebastian Padó (Computerlinguistik), Sebastian

⁷ Eine Ausnahme stellt die *Zeitschrift für digitale Geisteswissenschaften* (ZfdG), dar: <http://www.zfdg.de>.

⁸ Personen ohne Nennung einer Affiliation sind Angehörige der Universität Stuttgart.

- Haunss (Sozialwissenschaften, Universität Bremen).
 URL: <https://sites.google.com/view/mardy>
- Tracing Global Information Networks In Historical Newspaper Repositories (**Oceanic Exchanges**). Gefördert durch die DFG („Digging into Data Challenge“). Projektleitung: Ryan Cordell (Northeastern University), Marc Prieue (Amerikanistik), Sebastian Padó (Computerlinguistik), Steffen Koch (Visualisierung).
 URL: <https://oceanicexchanges.org>
 - Quantitative Drama Analytics (**QuaDrama**). Gefördert durch die Volkswagen-Stiftung. Projektleitung: Nils Reiter (Computerlinguistik), Marcus Willand (Neuere Deutsche Literaturwissenschaft).
 URL: <https://quadrama.github.io>
 - Quantitative Drama Analytics: Tracking Character Knowledge (**Q:TRACK**). Gefördert durch die DFG (Schwerpunktprogramm ‚Computational Literary Studies‘). Projektleitung: Nils Reiter (Computerlinguistik), Marcus Willand (Neuere Deutsche Literaturwissenschaft, Universität Heidelberg).
 URL: <https://quadrama.github.io>
 - Umfassende Modellierung von Redebeiträgen in Prosatexten (**QUOTE**). Gefördert durch die DFG. Projektleitung: Sebastian Padó (Computerlinguistik), Roman Klinger (Computerlinguistik).
 URL: <https://www.ims.uni-stuttgart.de/forschung/projekte/quote/>
 - Dhimmis and Muslims – analysing multi-religious spaces in the Medieval Muslim World (**Dhimmis and Muslims**). Gefördert durch die Volkswagen-Stiftung. Projektleitung: Steffen Koch (Visualisierung), Dorothea Weltecke (Mittelalterliche Geschichte, Universität Frankfurt/Main).
 URL: <https://www.vis.uni-stuttgart.de/en/projects/dhimmis-and-muslims/>
 - Structured Multi-Domain Emotion Analysis from Text (**SEAT**). Gefördert durch die DFG. Projektleitung: Roman Klinger (Computerlinguistik).
 URL: <https://www.ims.uni-stuttgart.de/forschung/projekte/seat/>
 - Science Data Center für Literatur (**SDC4Lit**). Gefördert durch das Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg. Projektleitung: Roland S. Kamzelak (Deutsches Literaturarchiv Marbach), Thomas Bönnisch (Höchstleistungsrechenzentrum Stuttgart), Gabriel Viehhauser (Digital Humanities), Jonas Kuhn (Computerlinguistik).
 URL: <http://www.sdc4lit.org>
 - Merkmale ästhetischer Reflexionsfiguren: Systematische Annotation und quantitative Analyse. Gefördert durch die DFG (Sonderforschungsbereich 1391, „Andere Ästhetik“). Projektleitung: Nils Reiter (Computerlinguistik), Angelika Zirker (Anglistik, Universität Tübingen).
 URL: <https://uni-tuebingen.de/de/160783>

- Realisierung einer Plattform und begleitender Dienste zum Forschungsdatenmanagement für die Fachcommunity Digital Humanities (**RePlay-DH**). Gefördert durch das Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg. Projektleitung: Jonas Kuhn (Computerlinguistik), Helge Steenweg (Universitätsbibliothek Stuttgart), Stefan Wesner (KIZ Ulm). URL: <https://www.ims.uni-stuttgart.de/forschung/projekte/replay-dh/>
- Automatic Fact Checking for Biomedical Information in Social Media and Scientific Literature (**FIBISS**). Gefördert durch die DFG. Projektleitung: Roman Klinger (Computerlinguistik).
- Text Mining auf geschützten Werken durch Auszüge transparent erschließen (**Xsample**). Gefördert durch das Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg im Rahmen des Programms „Wissenschaftliche Bibliotheken gestalten den digitalen Wandel (BW-BigDIWA)“. Projektleitung: Helge Steenweg (Universitätsbibliothek Stuttgart), Jonas Kuhn (Computerlinguistik), Thomas Dreier, (Zentrum für Angewandte Rechtswissenschaft, KIT Karlsruhe).
- Intertextuelle Referenz bei Nietzsche. Projektleitung: Axel Pichler (Philosophie/Literaturwissenschaft), Nils Reiter (Computerlinguistik).

Neben diesem ‚Ökosystem‘ und den in diesem Buch dargestellten inhaltlichen und methodischen Ergebnissen ist die Ausbildung und Schulung der in CRETA Mitarbeitenden ein wesentliches Ergebnis unserer Bemühungen der letzten Jahre. Um diese Kompetenzen weiterzuentwickeln, ist geplant, CRETA als Methoden-Netzwerk weiterzuführen und auch CRETA-Werkstätten weiterhin durchzuführen.

Abschließend möchten wir unsere in den letzten Jahren gesammelten Erfahrungen in Form von sechs Empfehlungen verdichten, um so die enge und interdisziplinäre Zusammenarbeit bei der Entwicklung textanalytischer Methoden für die digitalen Geisteswissenschaften auch zukünftigen interdisziplinären DH-Projekten zu vermitteln:

Externe Veranstaltungen. Ein frühzeitiges Commitment zu externen Events fördert das Zusammenwachsen in Arbeitsgruppen und dient auch – durch die Arbeit an gemeinsamen Materialien wie Vortragsfolien, Handouts oder Übungsaufgaben – als Katalysator für Begriffsklärungen. Nicht zuletzt gibt eine erfolgreiche Veranstaltung einen wichtigen Referenzpunkt für gemeinsame Arbeit und liefert dafür auch ein Erfolgserlebnis.

Förderung des persönlichen Austauschs. Kaum zu unterschätzen ist der persönliche Austausch, gerade in der Anfangsphase eines Projektes. Unsicherheiten und Nichtwissen zuzugeben, fällt wissenschaftlich Arbeitenden oft schwer. Diese lassen sich daher am leichtesten im informellen Rahmen klä-

ren. Die Schaffung von Gelegenheiten für informellen Austausch (etwa durch Stammtische oder gemeinsame Mensagänge) hat die Arbeit in CRETA erleichtert und für die Beteiligten auch angenehmer gestaltet.

Keine frontalen Projekttreffen. Projekttreffen, gerade in größeren Verbänden, sollten nicht nur als Vortragsreihe gestaltet werden. Raum für spontane Aktivitäten, bilaterale Absprachen oder praktische Arbeit hat zum Beispiel die CRETA-Werkstätten (siehe dazu die Seiten 28 und 468 in diesem Band) zu den produktiven und positiven Terminen gemacht, als die die Beteiligten sie wahrgenommen haben. Die Bandbreite möglicher Aktivitäten ist dabei recht groß und reicht vom gemeinschaftlichen Annotieren über die interaktive gemeinsame Datenanalyse bis hin zu Diskussionen in Kleingruppen. Gerade eine Aufteilung von großen Projektgruppen in kleinere Einheiten erleichtert es auch dem wissenschaftlichen Nachwuchs, sich einzubringen, und sorgt generell dafür, dass mehr Teilnehmende sich beteiligen können.

Freiraum für die Arbeitsebene. Projekte in den Digital Humanities lassen sich oft nicht bis ins letzte Detail planen, da sie letztlich von empirischen Ergebnissen abhängen. Es ist daher einerseits hilfreich, sich nicht zu fest an die Projektpläne zu klammern. Andererseits ist es in DH-Projekten oft so, dass praktische und pragmatische Aspekte einen größeren Stellenwert einnehmen, als sie das sonst in geisteswissenschaftlicher Forschung tun. Projektleitende sollten sich ihre disziplinären Grenzen und Vorannahmen bewusst machen, und – gerade beim interdisziplinären Zusammenarbeiten – ihr Engagement dementsprechend ausrichten.

Gegenseitige Weiterbildung. Ein großer Vorteil von DH-Projekten ist, dass die Beteiligten aus unterschiedlichen Disziplinen kommen, und daher Expertinnen und Experten für unterschiedliche Dinge sind. Eine gegenseitige Weiterbildung, bei der Mitarbeitende aus den Computerwissenschaften z. B. Programmierkurse und Mitarbeitende aus den Literaturwissenschaften einen Einblick in z. B. Interpretationstheorien geben, kann die Projektbeteiligten motivieren und ebenfalls zusammenführen. Im besten Falle führen die erlernten Grundlagen zu einem besseren Verständnis der jeweils ‚anderen Seite‘ und damit zu produktiverer Zusammenarbeit.

Technische Kenntnisse. Eine häufig gestellte Frage ist die, ob Forschende aus den Geistes- und Sozialwissenschaften, die sich für Digital-Humanities-Methoden interessieren bestimmte Technologien lernen sollten, insbesondere jedoch, ob sie programmieren können sollten. Basierend auf den Erfahrungen in CRETA lässt sich sagen, dass gewisse Programmierkenntnisse für einen mündigen und selbstbestimmten Einsatz von DH-Methoden nötig sind. Aufgrund der komplexen und wandelbaren Natur einzelner Module der reflektierten Textanalyse, sind sowohl eine graphische Benutzeroberfläche

als auch eine kompatible Sammlung von Tools, die den gesamten *workflow* unterstützen, nicht zu erwarten. Dieser (und andere) *workflows* werden auf absehbare Zeit aus einer Vielzahl von Einzel-Werkzeugen zusammengesetzt, die unterschiedliche Ein- und Ausgabeformate erfordern. Eine Programmierkenntnis, die aus und in verschiedene/n Dateiformaten konvertieren kann, scheint uns eine – nicht sehr hohe – Mindestanforderung zu sein.

Trotz der von uns und anderen Projektgruppen erzielten Erfolge in den letzten Jahren bleibt die Integration von geistes- und sozialwissenschaftlichen Fragestellungen, Annahmen und Vorgehensweisen mit computerwissenschaftlichen Methoden, Tools und Formaten eine Herausforderung. Spricht hieraus Resignation? Das glauben wir nicht. „Niemals setzt sich die Wissenschaft das Phantom zum Ziel, endgültige Antworten zu geben oder auch nur wahrscheinlich zu machen; sondern ihr Weg wird bestimmt durch ihre unendliche, aber keineswegs unlösbare Aufgabe, immer wieder neue, vertiefte und verallgemeinerte Fragen aufzufinden und die immer nur vorläufigen Antworten immer von neuem und immer strenger zu prüfen.“ (Popper, *Logik der Forschung*)