

Beitrag aus:

Sonderband 3 der ZfdG: Wie Digitalität die Geisteswissenschaften verändert: Neue Forschungsgegenstände und Methoden. Hg. von Martin Huber und Sybille Krämer. 2018. DOI: [10.17175/sb03](https://doi.org/10.17175/sb03)

Titel:

Digital Humanities – Eine Herausforderung an die Informatik und an die Geisteswissenschaften

Autor/in:

Klaus-Georg Deck

Kontakt:

klaus-georg.deck@mosbach.dhbw.de

Institution:

Duale Hochschule Baden-Württemberg, Digitalität und Data Science

GND:

[123092124](https://nbn-resolving.org/urn:nbn:de:hbz:5:1-63862-p0011-9)

ORCID:

[0000-0002-2032-1923](https://orcid.org/0000-0002-2032-1923)

DOI des Artikels:

[10.17175/sb003_002](https://doi.org/10.17175/sb003_002)

Nachweis im OPAC der Herzog August Bibliothek:

[1008144347](https://nbn-resolving.org/urn:nbn:de:hbz:5:1-63862-p0011-9)

Erstveröffentlichung:

27.06.2018

Lizenz:

Sofern nicht anders angegeben 

Medienlizenzen:

Medienrechte liegen bei den Autoren

Letzte Überprüfung aller Verweise:

19.06.2018

GND-Verschlagwortung:

[Digital Humanities](#) | [Informatik](#) |

Zitierweise:

Klaus-Georg Deck: Digital Humanities – Eine Herausforderung an die Informatik und an die Geisteswissenschaften. In: Digitale Metamorphose: Digital Humanities und Editionswissenschaft. Hg. von Roland S. Kamzelak /Timo Steyer. 2018 (= Sonderband der Zeitschrift für digitale Geisteswissenschaften, 2). PDF Format ohne Paginierung. Als text/html abrufbar unter DOI: [10.17175/sb003_002](https://doi.org/10.17175/sb003_002).

Klaus-Georg Deck

Digital Humanities – Eine Herausforderung an die Informatik und an die Geisteswissenschaften

Abstracts

Dieser Beitrag stellt das Verhältnis der an Projekten der digitalen Geisteswissenschaften beteiligten Disziplinen dar und geht dabei insbesondere auf die Rolle der Informatik ein. Anschließend werden sich daraus ergebende Anforderungen für eine erfolgreiche Zusammenarbeit von Wissenschaftlerinnen und Wissenschaftlern unterschiedlicher disziplinärer Herkunft präsentiert und abschließend Potentiale dargestellt, die sich aus dem digitalen Kontext für die Geisteswissenschaften eröffnen. Dabei werden die Digital Humanities als Herausforderung und gleichzeitig als Chance verstanden, die zu lange kultivierte Trennung der geistes- und naturwissenschaftlichen Wissenschaftsbereiche fruchtbar zu überbrücken.

This paper presents the relationship between the disciplines involved in the digital humanities, focusing in particular on the role of computer science. Subsequently, requirements for a successful cooperation between scientists of different disciplinary backgrounds are presented, and finally, potentials arising from the digital context for the humanities are formulated. We understand digital humanities as a challenge and at the same time as an opportunity to bridge the separation of the sciences and humanities, which has been cultivated for too long.

»Die vorschnelle Aufteilung der Wissenschaft auf zwei Hemisphären hat vergessen lassen, daß es sich bei manchem, was Theoretiker der Geisteswissenschaften als vermeintliches Charakteristikum dieses Wissenschaftsbereiches herausgestellt haben, in Wahrheit um Bestimmungen handelt, die nicht nur speziell für historische Wissenschaften, sondern für Wissenschaft überhaupt gelten. So kommt es dazu, daß als Gegenposition in solchen Fällen leicht das Bild einer Naturwissenschaft und eines Methodenideals entworfen wird, die es so niemals gegeben hat [...].«¹

1. Einleitung

Die Positionierung der Informatik innerhalb der Wissenschaften ist aus verschiedenen Gründen mit Schwierigkeiten verbunden. Zum einen bewegt sie sich hinsichtlich der Forschungsziele und der Gegenstandsbereiche wie auch ihrer Methoden innerhalb eines sehr breiten Spektrums, das sich thematisch von der Theoretischen Informatik über die Technische und Praktische Informatik bis zur Angewandten Informatik erstreckt.² Sie berührt nicht nur Grenzen zu zahlreichen anderen Wissenschaften, teilweise ist die fachliche Zuordnung einer Teildisziplin – wie etwa die der Theoretischen Informatik zur Informatik oder zur Mathematik –

¹ Wieland 1970, S. 33.

² Vgl. Rechenberg / Pomberger 2002, S. 6. John Nerbonne verortet die Informatik – wie er selbst zugibt, etwas unseriös – in die Geisteswissenschaften, vgl. Nerbonne 2015, passim, insbesondere unter »5. Schlussbemerkungen«.

nicht eindeutig und deren institutionelle Verortung auch von den historischen Entwicklungen einer Hochschule abhängig.

Häufig verschwimmen dabei die disziplinären Konturen. So lassen sich Medizinische Informatik, Wirtschaftsinformatik, Bioinformatik etc. als spezielle Ausprägungen der Angewandten Informatik verstehen, bei denen informatische Methoden und Werkzeuge in informatikfremden Bereichen zur Anwendung kommen, aber auch als Teilbereiche der Medizin, Betriebswirtschaft bzw. Biologie. Dabei bilden sich im Laufe der Zeit relativ eigenständige Disziplinen mit eigenen Forschungsfragen und -methoden heraus, die dann mit den ursprünglichen Wissenschaften thematisch und methodisch interferieren.³

Vor diesem Hintergrund lassen sich die Digital Humanities als Bereich zwischen der Informatik und den Geisteswissenschaften begreifen, in dem zunächst mit bekannten Methoden und Werkzeugen der Informatik Fragestellungen der Geisteswissenschaften behandelt werden, aus dem sich jedoch durchaus neue Methoden und Problemstellungen entwickeln können. Dies betrifft jedoch nicht nur die Geisteswissenschaften, sondern auch die Informatik, die aus solchen Fragestellungen neue Impulse erhält und daraus neue Konzepte entwickeln kann.

2. Unterschiedliche Forschungsparadigmen

Die erfolgreiche Anwendung computergestützter Verfahren und Methoden setzt eine adäquate Modellierung des zu betrachteten Bereichs voraus. Die Erfolge, welche die Informationstechnologie innerhalb der quantitativen Wissenschaften genauso verzeichnet wie in Technik und Wirtschaft, hängen wesentlich damit zusammen, dass dort diese Modellierung gut gelingt. Die Begrifflichkeiten und die Gegenstände sind klar definiert und die angewandten Verfahren standardisiert, so dass diese direkt operationalisiert werden können. Als geradezu paradigmatisches Beispiel sei hier der Finanzsektor genannt, wo man sich per se innerhalb eines digitalen Kontextes befindet und damit in einem Umfeld, in dem die erwähnte Modellierung und digitale Kodierung im großen Umfang bereits geleistet wurden. In IT-affinen Disziplinen liegt ein sehr hoher Spezialisierungsgrad vor, verbunden mit Annahmen und Voraussetzungen, über die innerhalb des disziplinären Rahmens eher selten reflektiert wird. Soll etwa ein technisches Verfahren oder ein Geschäftsprozess »optimiert« werden, dann ist in der Regel klar, welche Größen – nämlich Zeit und Kosten – die »Verbesserung« bestimmen. Es wird nicht thematisiert, was Optimierung überhaupt bedeutet, ob und gegebenenfalls welche Alternativen sich anbieten und welche weiteren Implikationen sich aus einer derartigen Optimierung ergeben.⁴

³ Dies ist kein Phänomen, das nur die Informatik betrifft. Jürgen Mittelstraß beschreibt dieses wissenschaftsphilosophische Dilemma einer zunehmenden disziplinären Partikularisierung einerseits, verbunden mit der Beobachtung, dass sich manche Probleme dem Zugriff einer einzelnen Disziplin entziehen. Unter dem Begriff »Transdisziplinarität« schlägt er eine Lösung vor, welche die Wissenschaftssystematik als dynamische Struktur begreift und die Identität der Disziplinen und deren gegenseitige Abgrenzungen selbst in den Forschungs- und Wissenschaftsprozess mit einbezieht, vgl. Mittelstraß 2007, S. 7.

⁴ Michael Hampe weist mit Bezug auf Charles Sanders Peirce auf die Gefahr hin, dass rein ökonomisch motivierte Deutungen und Interpretationen des Handelns eine Gesellschaft herbeiführen können, die sich nach genau diesen Prinzipien verhält, vgl. Hampe 2014, S. 40ff.

In vielen Bereichen der Geisteswissenschaften sieht dies jedoch anders aus. Sie besitzen im Allgemeinen nicht die methodische und begriffliche Affinität zur Informationstechnologie, ihre Methoden sind aus der Perspektive der Informationsverarbeitung weniger standardisiert, ihre Aussagen eher qualitativ und in der Regel weniger formalisiert. Naturgemäß geht es mehr um Interpretation und Bewertung von Phänomenen als um quantitativ exakte Aussagen. Kommen dabei digitale Methoden und Verfahren zum Einsatz, dann stellen deren Resultate weniger das Forschungsergebnis selbst dar, sondern sie sind innerhalb des gesamten Kontextes einzubetten, zu interpretieren und zu bewerten. Die bei einer konkreten digitalen Modellierung explizit oder implizit getroffenen Voraussetzungen besitzen innerhalb der Geisteswissenschaften eine größere Relevanz als dies innerhalb eines IT-affinen Kontextes der Fall ist. Dies betrifft nicht nur die Festlegung, welche Attribute in welchen Datentypen und in welcher Genauigkeit oder Granularität erfasst werden sollen, sondern bereits den Gegenstandsbereich selbst. Gerade wenn von der ›digitalen Faktenlage‹ die Rede ist und der Begriff der Daten als ›das Gegebene‹ suggeriert, als handle es sich dabei um Informationen über Fakten, die nur noch einzusammeln wären, darf man nicht aus den Augen verlieren, dass eine Datenstrukturierung »bereits einen ersten hermeneutischen Akt darstellt.«⁵ Die Methoden und Werkzeuge der Informatik stellen keinen neutralen Weltzugang her, indem sie bestehende Sachverhalte digital abbilden und maschinell verarbeiten, sondern sie sind selbst konstitutiv an der Erzeugung von Wirklichkeit beteiligt.

Dabei ist noch zu erwähnen, dass die Entwicklung der Informatik selbst innerhalb eines historischen Kontextes zu begreifen ist. Systematisch betrachtet entstand sie zusammen mit der formalen Logik aus der Mathematik, wesentliche Treiber aber waren und sind vorwiegend politische und wirtschaftliche Interessen. Beispiele sind etwa die Entschlüsselungsprojekte der Alliierten während des 2. Weltkrieges und auch die aktuellen Aktivitäten von Google, Apple, Facebook und Amazon. Dass sich diese Motivatoren nicht nur auf praktische Systeme und Anwendungen sondern auch auf grundlegende Konzepte und Technologien auswirken, sollte man gerade in Zeiten von »Big Data« und »Web 2.0« nicht aus den Augen verlieren.

Mit dem vielfach formulierten Gegensatz zwischen Geisteswissenschaften und Naturwissenschaften⁶ wurde eine Demarkationslinie gezogen, die auf beiden Seiten zu lange kultiviert wurde. Die Digital Humanities stellen die Herausforderung und bieten gleichzeitig die Chance, diese Grenze an der einen oder anderen Stelle zu durchbrechen. Dies bedeutet zum einen, quantitative, empirische und formale Methoden digital operationalisiert für die Geisteswissenschaften fruchtbar zu machen, dann aber auch eine Methodologie zu entwickeln, die den Ansprüchen einer Wissenschaftstheorie (im Sinne einer ›Philosophy of Science‹) gerecht wird oder sich zumindest dazu positioniert. Zum anderen bieten sich auch für die Informatik Chancen, ihre etablierten Methoden in einem neuen Kontext anzuwenden, diese anzupassen und weiter zu entwickeln.⁷

⁵ Vgl. Scheuermann 2016, S. 61. Im dortigen Beitrag werden auch die Grundsätze einer hermeneutisch orientierten Methodologie der Digital Humanities thematisiert.

⁶ Vgl. Scheuermann 2016, S. 59f., Schemmer 1990, S. 48ff. und Wieland 1970, passim.

⁷ Typische Anforderungen sind etwa die zu berücksichtigende historische Dimension über einen langen Zeitraum. Die Informationstechnologie bietet zwar Verfahren zur historisierten Verwaltung von Information, man geht aber – etwa bei einer klassischen Entity-Relationship Datenmodellierung (vgl. Kemper / Eickler 2011, S. 37ff.) – davon aus, dass sich im Betrachtungszeitraum nur Attributwerte von Entitäten ändern und keine strukturellen Änderungen oder begriffliche Verschiebungen stattfinden.

Dies fördert bei den Akteuren seitens der Informatik ein reflektierteres Verständnis von ihrer Fachdisziplin und ihrer Tätigkeit, mehr jedenfalls, als es etwa innerhalb eines rein finanzwirtschaftlichen Kontextes zu erwarten wäre. Im Idealfall treten in Projekten der Digital Humanities verschiedene Forschungsmomente gemeinsam auf, ohne dass deren Zugehörigkeit zu unterschiedlichen Paradigmen im Vordergrund steht.⁸

Ein Ansatzpunkt für eine solche Weiterentwicklung der Informatik stellt etwa der Umgang mit unsicheren oder ungenauen zeitlichen Informationen dar. Während ein Historiker in der Lage ist, mit Datierungen wie ›im frühen 12. Jahrhundert‹ oder ›Frühjahr 1770 bis Sommer 1771‹ umzugehen, bringt diese Art von Information ein Digitalisierungsvorhaben schnell an seine Grenzen. Dies betrifft die Modellierung, aber auch die digitale Codierung und insbesondere die Operationalisierung von Ereignissen mit solchen zeitlichen Bezügen. So ist die zeitliche Relation zwischen ›Mitte 14. Jahrhundert‹ und ›26. Juni 1348‹ insofern unklar, dass bei diesen unterschiedlichen Zeitangaben weder die erste vor der zweiten liegt noch umgekehrt. Fasst man dagegen die Angaben als Zeitperioden auf, dann ist letztere in der ersten enthalten, aber nicht umgekehrt. Auch wenn auf Seiten der Modellierung bereits Lösungsansätze existieren⁹, steht dies für eine adäquate Datenverwaltung noch aus. Dabei ist dieses Problem nicht allein auf die Digitalen Geisteswissenschaften beschränkt: In klinischen Studien werden Ereignisse mit zeitlichem Bezug (Geburts-, Erkrankungs-, Sterbedatum) in der Regel mit dem Standarddatentyp ›Date‹ modelliert. Da dieser Datentyp die Eingabe einer tagesgenauen Zeitangabe erzwingt, führt dies dazu, dass ungenaue Informationen mit einem vereinbarten Standardwert (etwa ›1.07.1920‹ für ›1920‹) erfasst werden. Wird der Umstand, dass es sich bei diesem Datumseintrag um einen symbolischen Wert handelt, nicht durch ein weiteres Attribut im Datensatz vermerkt, führt dies nicht selten zu Inkonsistenzen oder zu falschen statistischen Aussagen.¹⁰

3. Erfolgsfaktoren

An Projekten der Digitalen Geisteswissenschaften sind allein schon wegen der Verschiedenheit der erforderlichen Kompetenzen häufig mehrere Personen unterschiedlicher wissenschaftlicher Provenienz beteiligt. Neben den bekannten Faktoren einer erfolgreichen Teamarbeit wie transparente Zielsetzung, klare Aufgabenteilung, transparente Entscheidungsstrukturen, gegenseitiges Vertrauen und offene Kommunikation spielen in solchen Projekten gegenseitige Wertschätzung sowie sich ergänzende, aber auch überschneidende Kompetenzen der Mitglieder eine besondere Rolle.

⁸ Dass digitale Methoden zu Erkenntnissen führen, die bisheriges Wissen ergänzen oder erweitern, ist dabei eher unspektakulär. Methodologisch und wissenschaftstheoretisch spannend wird es dann, wenn ›digital erworbenes Wissen‹ zu einer Neubewertung bisheriger Methoden und Annahmen führt. Ein Beispiel liefern Charlotte Schuberts Untersuchungen zur Autorschaft Plutarchs mittels visueller Textanalyse, vgl. Schubert 2016, passim.

⁹ Vgl. Stokes 2015, passim.

¹⁰ Inkonsistenzen können entstehen, wenn im Nachhinein Informationen mit exakten Zeitangaben bekannt werden, z. B. Arztbriefe mit Diagnosen nach dem vermeintlichen Sterbedatum. Falsche statistische Informationen sind etwa Aussagen über die Häufung von Geburten oder Todesfällen am symbolischen Datumswert.

Im Idealfall kennen die beteiligten Informatiker nicht nur die bereits auf ihre fachliche Ebene projizierte (und damit häufig verkürzte) Sicht der geisteswissenschaftlichen Fragestellungen, sondern ihnen sind die originären Forschungsfragen genauso bekannt wie die Historie bisheriger Lösungsansätze und Antworten. Denn nur dann ist es ihnen möglich, die Anforderungen der Geisteswissenschaften an die Informatik zu erkennen und daraus gegebenenfalls neue Methoden zu entwickeln.

Umgekehrt sind die Wissenschaftler aus den Geisteswissenschaften soweit über die Methoden und Werkzeuge der Informatik informiert, dass sie diese nicht als »Black-Box« betrachten (müssen); sie kennen die Möglichkeiten, die Randbedingungen und Annahmen, aber auch die Grenzen der verwendeten oder zur Disposition stehenden digitalen Methoden. Wichtige Themenfelder in diesem Zusammenhang sind die Modellierung und Strukturierung von Information aber auch der Begriff der Digitalisierung selbst: Die Unterscheidung von Kodierung von Information einerseits, das maschinelle Verarbeiten dieser digitalen Daten mit Algorithmen und Programmen andererseits¹¹ und nicht zuletzt die Visualisierung der Resultate.

Wenn alle Beteiligten die Bereitschaft und die Fähigkeit mitbringen, den eigenen Standpunkt zu erkennen, diesen zu reflektieren und gegebenenfalls zu modifizieren, sich aber auch auf andere Positionen innerhalb des Projektes einzulassen, so fördert dies ein gemeinsames Grundverständnis über das Forschungsprojekt, führt zu weniger begrifflichen und methodischen Missverständnissen und im Idealfall zu neuen Methoden oder Verfahren. Gerade über das Verbindende oder Gemeinsame können Brücken geschlagen und beschriftet werden.

Mit dem Anspruch der gegenseitigen fachlichen Wertschätzung in heterogen besetzten Forschungsteams geht auch die Forderung nach einer adäquaten Würdigung der wissenschaftlichen Leistung der Beteiligten einher. Dies betrifft die Autorenschaft bei Publikationen, aber auch die Unterstützung durch die universitären Organe und Gremien bei der Möglichkeit zur wissenschaftlichen Qualifikation innerhalb solcher Forschungsprojekte, insbesondere dann, wenn Graduierte aus Disziplinen verschiedener Fakultäten beteiligt sind.¹²

4. Potentiale

Das Internet/Web 2.0 vereinfacht ortsunabhängige Formen kollaborativer Forschung. Diese reichen von der verstärkten Nutzung von Home-Office einzelner Wissenschaftler über Kooperationsprojekte mehrerer Institutionen bis zur Bildung international besetzter multidisziplinärer Forschungsteams. Dem gegenüber stehen zu überwindende

¹¹ Hierzu zählt nicht unbedingt die Kenntnis einer bestimmten Programmiersprache. Algorithmische Verfahren lassen sich genauso gut (wenn nicht besser) in natürlicher Sprache oder in Pseudocode formulieren.

¹² Fast anekdotenhaft klingen meine persönlichen Erfahrungen (vor über 20 Jahren), als mir an der neuphilologischen Fakultät die Annahme als Doktorand in Computerlinguistik verwehrt wurde, da mir dafür trotz Latinum, Altgriechisch- und Hebräischkenntnissen ein Seminarschein in Mittelhochdeutsch fehlte. – Erschwerend kommt hinzu, dass Projektstellen der Geisteswissenschaften finanziell häufig nicht so gut ausgestattet sind als solche im technischen/naturwissenschaftlichen Bereich, und daher für Absolventen der Informatik weniger attraktiv scheinen.

Hemmnisse und offene Fragen: Wer ist Eigentümer der Daten, wer darf welche Daten, welche Teilresultate wie nutzen? Wie sieht eine adäquate Beteiligung aller Partner aus? Ähnliches gilt für Themen wie Open Data, Open Access und Open Source, die häufiger eingefordert als letztlich eingehalten werden. In diesem Umfeld muss sich eine entsprechende wissenschaftssoziologische Kultur und Praxis erst noch ausprägen und entwickeln.

Das Web bietet darüber hinaus die Möglichkeit, Recherchen und Datenabgleich mit (Web/Cloud-basierten) Archivdaten sowie Datenvalidierung bereits bei der Feldarbeit durchzuführen, was zu einer zeitnahen Datenverfügbarkeit und zu einer erhöhten Datenqualität führt. Dies gilt auch für Analysewerkzeuge, die als Webservice ortsunabhängig verfügbar sind.¹³ Im Bereich der Linguistik und Angewandten Sprachwissenschaft werden Sprachressourcen wie Korpora oder verschiedene linguistische Werkzeuge (Parser, Tokenizer, Tagger) mittlerweile webbasiert angeboten, die auch für nicht rein linguistische Forschungsvorhaben genutzt werden können.¹⁴

Was die Verarbeitung von großen Datenmengen aus heterogenen Datenquellen betrifft, stellt die Informationstechnologie aus dem kommerziellen Umfeld Werkzeuge zu deren Verarbeitung, Integration, Analyse und Visualisierung – oft in einer Open Source Version kostenfrei – zur Verfügung. Grundsätzlich lohnt es sich, dass sich die Digitalen Geisteswissenschaften an bestehenden (Industrie-)Standards orientieren und die dort entwickelten Werkzeuge für eigene Szenarien nutzen. Beispiele sind etwa das Data Mining Werkzeug RapidMiner, das in zahlreichen linguistischen Analysen verwendet wird, oder Graphendatenbanksysteme, wie Neo4J, zur Modellierung und Verarbeitung flexibler Beziehungsstrukturen.¹⁵ Häufig lassen sich die Werkzeuge durch Konfiguration anpassen oder über offene Schnittstellen erweitern. Gerade hier ist technologische wie fach- (in diesem Fall geistes-) wissenschaftliche Kompetenz vonnöten, um zu beurteilen, was ein Werkzeug leistet und was nicht, und ob es – gegebenenfalls modifiziert – für die eigene Fragestellung verwendet werden kann. Der folgende Abschnitt skizziert eine erfolgreiche Erweiterung und Modifikation von RapidMiner um Text-Mining Verfahren für die Linguistik.

5. RapidMiner und KobRA

Das Softwaresystem **RapidMiner**¹⁶ gilt im kommerziellen wie im akademischen Bereich als eines der führenden Werkzeuge für die Verarbeitung, Analyse und Visualisierung großer Datenmengen. Ursprünglich am Lehrstuhl für Künstliche Intelligenz der Technischen Universität Dortmund entwickelt, wird die baukastenartige Umgebung für Maschinelles Lernen, Data-Mining und prädiktive Analysemethoden von einem kommerziellen Unternehmen in verschiedenen Lizenzvarianten bereit gestellt. Das System beinhaltet mittlerweile über 1000 Operatoren, aus denen Data-Mining Prozessabläufe in einer graphischen Umgebung zusammengestellt und konfiguriert werden können. Neben Standardoperatoren, etwa für die

¹³ Ein Beispiel aus der Archäologie ist die Möglichkeit der Validierung von Fundpositionen direkt auf dem Grabungsfeld mittels digitaler, webbasierter Stratigraphie durch Harris-Matrizen, vgl. Bobowski 2005, passim.

¹⁴ Vgl. Weblicht und Clarin-D.

¹⁵ Vgl. Bartz et al. 2014, S. 2ff. und die Akademie der Wissenschaften und der Literatur Mainz.

¹⁶ Vgl. RapidMiner.

Datenerfassung, Datenvalidierung und Visualisierung der Resultate, stehen Operatoren für die unterschiedlichsten Algorithmen des Maschinellen Lernens zur Verfügung. Darüber hinaus erlaubt das offene System die Entwicklung und Integration eigener Operatoren für spezielle fachliche oder technologische Anforderungen.

Im BMBF-Verbundprojekt Korpus-basierte linguistische Recherche und Analyse mithilfe von Data-Mining (KobRA)¹⁷ werden in einer Kooperation von Wissenschaftlern aus Informatik, Linguistik und Sprachtechnologie¹⁸ linguistische Recherche- und Analysewerkzeuge entwickelt und in RapidMiner integriert. Schwerpunkte des Projekts sind unter anderem die korpusbasierte automatische Disambiguierung von Wörtern mit mehreren Bedeutungen (Polyseme und Homonyme) sowie die Beschreibung und Visualisierung von diachronen Bedeutungsentwicklungen. Hierbei wird das Verfahren der Latent Dirichlet Allocation (LDA)¹⁹ dahingehend modifiziert, dass die so genannten Topics, die Kontextwörter eines Wortes, als Repräsentationen von dessen Bedeutung aufgefasst und im Modell zur Disambiguierung verwendet werden.²⁰ Bedeutung eines Wortes wird also als latente Information des sprachlichen Kontextes verstanden.

Dieser Ansatz hat sich bei verschiedenen linguistischen Evaluationen als fruchtbar erwiesen. Mit diesem Verfahren können nicht nur automatisierte Disambiguierungen erfolgreich durchgeführt werden, sondern durch die Einbeziehung zusätzlicher korpuslinguistischer Daten (Wortarten- oder Syntaxannotationen) und weiterer Metainformation (etwa Textsorten oder Zeiträume) ergeben sich, jenseits des eigentlichen Textes, weitere Merkmale zur Bildung von Topics und damit eine Erweiterung des Bedeutungsbegriffs. Dadurch eröffnen sich Forschungsfragen, die an die ursprünglichen Fragen anschließen, aber auch darüber hinausgehen.²¹

6. Digitalität als Thema für die Wissenschaft

Abschließend möchte ich noch auf zwei Themen aufmerksam machen, die sich gerade für die Geisteswissenschaften im Zusammenhang mit der Digitalität eröffnen:

Zum einen betrifft dies den Umstand, dass in beiden Bereichen Text und Literalität eine bedeutende Rolle spielen. Texte sind in den Geisteswissenschaften häufig Gegenstand und so gut wie immer Medium des wissenschaftlichen Diskurs. Ein wesentlicher Aspekt von Digitalität besteht nun darin, dass Phänomene ›in der Welt‹ digital kodiert, das heißt in eine lineare Folge endlich vieler unterscheidbarer Zeichen – in der Regel sind dies zwei Zeichen, häufig bezeichnet mit 0 und 1 – transformiert werden. Diese Digitalisate werden mittels Programmen – dies sind

¹⁷ Vgl. KobRA sowie Bartz et al. 2014, passim.

¹⁸ Vgl. KobRA Team: Die beteiligten Personen und Forschungseinrichtungen.

¹⁹ LDA wurde ursprünglich zur thematischen Partitionierung von Dokumenten entwickelt. Vgl. Blei et al. 2003, passim.

²⁰ Dies ist nur eine sehr grobe Darstellung des Verfahrens. Für eine detaillierte Beschreibung vgl. Bartz et al. 2014, passim und KobRA-Disambiguierung.

²¹ Mit diesem Verfahren können neben metaphorischen Verwendungen von Begriffen auch nicht lexikographisch erfasste Bedeutungen ausgemacht werden, sowie ›Bedeutungs-Topics‹ mit im Vergleich zur lexikographischen Bedeutung feinerer Granularität, vgl. KobRA-Disambiguierung.

nichts anderes als linear kodierte Algorithmen – untersucht und Muster, Korrelationen oder Konturen ausgemacht, die schließlich visualisiert werden. Sowohl bei den kodierten Daten als auch bei den Handlungsanweisungen an die Computer, den Programmen, handelt es sich um lineare Zeichenfolgen, also im weiteren Sinn um Texte.²² Letztere entstammen in der Regel einer Sprache, der Programmiersprache, mit einer Syntax und einer – im Gegensatz zu natürlichen Sprachen – eindeutigen Semantik. Hier gibt es möglicherweise einiges Potential, wenn man die Gemeinsamkeiten, aber auch die auf einer solchen Basis auszumachenden Unterschiede untersucht.

Das zweite Phänomen betrifft die digitale Durchdringung aller unserer Lebensbereiche und deren Folgen, vor denen so viele warnen. Mit dem Aufkommen prädiktiver Analyseverfahren, die mit Algorithmen des maschinellen Lernens Muster in großen Datenmengen erkennen und ohne zunächst begründbaren Zusammenhang verlässliche Vorhersagen treffen, wurde bereits 2008 eine neue Ära prognostiziert und das »Ende jeglicher Theorie« beschworen²³. Ohne an dieser Stelle auf diese Problematik im Detail einzugehen und dies zu bewerten, existiert dafür mindestens eine weitere Lesart: Zwar schafft es die Digitalität, Korrelationen auszumachen und Muster auf der digitalen Ebene zu konstatieren, sie liefert aber keine Erklärung oder Begründung. Das mit vielen Daten trainierte System scheint eine Art von (zumindest implizitem) Wissen zu besitzen, das über die bloße »digitale Faktenlage« hinausgeht. Mehr noch: Es werden Muster durch digitale Methoden hervorgebracht und letztlich verstärkt, die wir ohne diese Methoden möglicherweise nie erkennen würden. Vor diesem Hintergrund stellt sich die Frage nach einer Begriffsbestimmung von »Wissen«, »Erklären«, »Verstehen« aus der Perspektive des Digitalen.

Geisteswissenschaften beanspruchen einen hohen Grad der Reflexion über ihre Methoden. Dies betrifft auch den Begriff des Digitalen aus philosophischer und medientheoretischer Perspektive: Vor dem Hintergrund, dass Medien die Welt nicht einfach abbilden sondern auch konstituieren, stellt sich die Frage, welche Besonderheiten es bei der Digitalität im Vergleich zu konventionellen Medien gibt. Inwieweit vermitteln digitale Medien – etwa wenn diese in einem bereits etablierten Projekt neu hinzukommen – eine andere Realität? Was machen Digitalität im Kontext von Wissenschaft und die digitale Durchdringung unserer Lebenswelt mit uns als wahrnehmendes und erkennendes Subjekt? Inwieweit ist der Begriff der Wissensgenerierung und -verbreitung sowie der von Wissenschaft von diesen Entwicklungen betroffen?²⁴ Man kann nun nicht erwarten, dass diese Fragen von den Digital Humanities quasi im Vorbeigehen beantwortet werden, genauso wenig, dass sie selbst zunächst ihre eigene Methodologie klären, um danach sich an dieser orientierend wissenschaftlich zu arbeiten. Hier sind die Wissenschaftsphilosophie und Wissenschaftsgeschichte – also geisteswissenschaftliche Disziplinen – gefragt, diese Entwicklungen zeitnah zu begleiten.

²² Sybille Krämer hebt diese Doppelfunktion von Kalkülen hervor: Sie repräsentieren die Gegenstandsdomäne und dienen gleichzeitig als Instrument zur Problemlösung, vgl. Krämer 1998, S. 29.

²³ Als locus classicus gilt der provokative Essay des Wissenschaftsjournalisten Chris Anderson, vgl. Anderson 2008, passim.

²⁴ Vgl. Hagner 2014, passim.

Bibliographische Angaben

- Digitale Akademie. Graphentechnologien. Hg. von der Akademie der Wissenschaften und der Literatur Mainz. 2017. [\[online\]](#)
- Chris Anderson: The end of theory: The data deluge makes the scientific method obsolete. In: Wired Magazine. Beitrag vom 23. Juni 2008. [\[online\]](#)
- Thomas Bartz / Christian Pölitz / Katharina Morik / Angelika Storrer: Using Data Mining and the CLARIN Infrastructure to Extend Corpus-based Linguistic Research. [\[online\]](#) In: CLARIN 2014 Selected Papers. Linköping Electronic Conference Proceedings 116 (2014), S. 1–13.
- David M. Blei / Andrew Y. Ng / Michael I. Jordan: Latent Dirichlet Allocation. In: Journal of Machine Learning Research 3 (2003), S. 993–1022. [\[online\]](#)
- Bogdan Bobowski: Archaeoscope-Interface for 3D Stratigraphy Visualisation. In: Computer Applications & Quantitative Methods in Archaeology 33 (2005), S. 151–153. [\[online\]](#)
- Clarín-D: Auswerten. Web-basierte Analysewerkzeuge. Hg. von Clarín-D, Universität Tübingen. 2016. [\[online\]](#)
- Michael Hagner: Geisteswissenschaften oder Digital Humanities? [\[online\]](#) In: Forschung & Lehre 21 (2014), H. 2, S. 85. [\[Nachweis im GBV\]](#)
- Michael Hampe: Die Lehren der Philosophie. Eine Kritik. 2. Auflage. Berlin 2014. [\[Nachweis im GBV\]](#)
- Alfons Kemper / André Eickler: Datenbanksysteme. Eine Einführung. 8., aktualisierte und erweiterte Auflage. München 2011. [\[Nachweis im GBV\]](#)
- Wiki KobRA: Disambiguierung: Disambiguierung von Suchtreffern: Leiter und zeitnah. Hg. von Technischen Universität Dortmund. 2016. [\[online\]](#)
- Wiki KobRA: Hauptseite. Korpus-basierte linguistische Recherche und Analyse mit Hilfe von Data-Mining. Hg. von Technischen Universität Dortmund. 2016. [\[online\]](#)
- Wiki KobRA: Team. Hg. von Technischen Universität Dortmund. 2016. [\[online\]](#)
- Sybille Krämer: Zentralperspektive, Kalkül, Virtuelle Realität. Sieben Thesen über die Weltbildimplikationen symbolischer Formen. In: Medien, Welten, Wirklichkeiten. Hg. von Gianni Vattimo / Wolfgang Welsch. (Internationaler Workshop Media Transforming Reality, Magdeburg, 16.-18.12.1994) München 1998, S. 27–37. [\[Nachweis im GBV\]](#)
- Jürgen Mittelstraß: Methodische Transdisziplinarität. In: LIFIS ONLINE. Internet-Zeitschrift des Leibniz-Instituts für interdisziplinäre Studien e.V. (LIFIS). 5. November 2007. PDF. [\[online\]](#)
- John Nerbonne: Die Informatik als Geisteswissenschaft. DOI: [10.17175/sb001_006](#) In: Grenzen und Möglichkeiten der Digital Humanities. Hg. von Constanze Baum / Thomas Stäcker. Wolfenbüttel 2015. (= Sonderband der Zeitschrift für digitale Geisteswissenschaften, 1) DOI: [10.17175/sb01](#)
- RapidMiner: Data Science Platform | RapidMiner. Real Data Science, Fast and Simple. Hg. von RapidMiner Inc. 2017. [\[online\]](#)
- Informatik-Handbuch. Hg. von Peter Rechenberg / Gustav Pomberger. 3., aktualisierte und erweiterte Auflage. München u.a. 2002. [\[Nachweis im GBV\]](#)
- Oswald Schemmer: Die Philosophie und die Wissenschaften. Zur Kritik einer Abgrenzung. Frankfurt/Main 1990. [\[Nachweis im GBV\]](#)
- Leif Scheuermann: Die Abgrenzung der digitalen Geisteswissenschaften. DOI: [10.11588/dco.2016.1.22746](#) In: Digital Classics Online 2 (2016), H. 1, S. 58–67. [\[online\]](#)
- Charlotte Schubert: Die Visualisierung von Quellennetzwerken am Beispiel Plutarchs. DOI: [10.11588/dco.2016.1.23825](#) In: Digital Classics Online 2 (2016), H. 1, S. 68–87. [\[online\]](#)
- Peter Anthony Stokes: The Problem of Digital Dating: A Model for Uncertainty in Medieval Documents. [\[online\]](#) In: Conference Abstracts. DH2015 Global Digital Humanities. Hg. von University of Western Sydney. (DH 2015, Sydney, 29.06.-03.07.2015). Sydney 2015. [\[online\]](#)
- Wiki WebLicht: Weblicht. Main Page. Hg. von Clarín-D, Universität Tübingen. Stand 09.06.2017. [\[online\]](#)
- Wolfgang Wieland: Möglichkeiten der Wissenschaftstheorie. In: Hermeneutik und Dialektik. Aufsätze. Hg. von Rüdiger Bubner / Konrad Cramer / Reiner Wiehl. Bd. 1: Methode und Wissenschaft. Lebenswelt und Geschichte. Tübingen 1970, S. 33–56. [\[Nachweis im GBV\]](#)