Digital Humanities

Ben Heuwing* und Christa Womser-Hacker

Zwischen Beobachtung und Partizipation – nutzerzentrierte Methoden für eine Bedarfsanalyse in der digitalen Geschichtswissenschaft

DOI 10.1515/iwp-2015-0058

Zusammenfassung: Dieser Beitrag zeigt informationswissenschaftliche Ansätze für die domänenspezifische Bedarfsanalyse in den Digital Humanities anhand eines laufenden Projekts. In dem Projekt *Welt der Kinder* werden in einem interdisziplinären Team aus Historikern, Computerlinguisten und Informationswissenschaftlern innovative Methoden entwickelt und evaluiert, mit denen ein großes historisches Schulbuchkorpus mit Mitteln der automatischen Textanalyse vergleichend ausgewertet wird. Aus informationswissenschaftlicher Sicht werden dabei Methoden für die nutzerzentrierte Bedarfsanalyse eingesetzt und bewertet, die in einen Anforderungskatalog für die Systemgestaltung münden sollen.

Deskriptoren: Geschichtswissenschaft, Digital, Benutzerforschung, Informationssystem, Gestaltung, Textanalyse, Maschinell

Between Observation and Participation – Application of User Centered Methods for a Requirements Analysis in a Digital History Project

Abstract: This contribution demonstrates the application of approaches from information science to the domain-specific analysis of user requirements in the digital humanities. The ongoing project “Welt der Kinder” (Children and their World) serves as an example of a project in the area of digital history. In this interdisciplinary project, a team of historians, computer linguists and information scientists develops and tests innovative methods for the comparative analysis of a large corpus of historical text-

books. From an information science perspective, methods for the user centred requirements analysis will be applied and evaluated in the context of the digital humanities, resulting in a set of system requirements.

Descriptors: Digital History, Requirements Analysis, User Centred Design, Textanalysis, Automatic

Entre observation et participation – des méthodes centrées sur l'utilisateur pour l'évaluation des besoins dans l'historiographie numérique

Résumé: L'article décrit les approches des sciences de l'information à l'égard des besoins spécifiques dans le domaine des humanités numériques en se basant sur un projet en cours. Dans le projet « Welt der Kinder » (« Le monde des enfants ») une équipe multidisciplinaire d'historiens, de linguistes et d'informaticiens développe et teste des méthodes novatrices par lesquelles un grand corpus de manuel d'histoire est évalué de façon comparative au moyen d'analyse automatique de texte. Du point de vue des sciences de l'information, des méthodes d'analyse des besoins centrées sur l'utilisateur sont instaurées et évaluées dans le but de créer un catalogue de spécifications pour la conception de systèmes.

Descripteurs: Historiographie numérique, Méthodes des sciences de l'information, Analyse des besoins, Conception centrée sur l'utilisateur, Analyse automatique de texte

Das interdisziplinäre Projekt Welt der Kinder

Das Projekt *Welt der Kinder* untersucht Weltbilder, die Kindern im Deutschen Kaiserreich (1871–1918) in Schulbüchern vermittelt worden sind. Unter anderem soll dabei die Darstellung von Deutschlands Rolle in der Weltpolitik und als Kolonialmacht sowie die Einordnung fremder Völker und die Verwendung des Rassebegriffs untersucht wer-

*Kontaktperson: Dr. Ben Heuwing, Institut für Informationswissenschaft und Sprachtechnologie, Universität Hildesheim, Universitätsplatz 1, 31141 Hildesheim, Deutschland, E-Mail: heuwing@uni-hildesheim.de

Prof. Dr. Christa Womser-Hacker, Institut für Informationswissenschaft und Sprachtechnologie, Universität Hildesheim, Universitätsplatz 1, 31141 Hildesheim, Deutschland, E-Mail: womser@uni-hildesheim.de, www.uni-hildesheim.de/~womser

den. Die Grundlage für die Analysen bildet ein Korpus¹ von Geschichts- und Geografieschulbüchern aus dieser Periode, welches bereits über 3100 Bände umfasst. Für eine vergleichende Analyse soll das Korpus mit ausgewählter Kinder- und Jugendliteratur erweitert werden. Auf der Basis von bibliografischen und inhaltsbezogenen Metadaten werden im Projekt Veränderungen der Inhalte im Zeitverlauf und Unterschiede zwischen Kategorien von Schulbüchern untersucht, etwa für bestimmte Altersklassen oder konfessionelle Ausrichtungen. Dafür werden Werkzeuge zur Unterstützung der vergleichenden quantitativen Analyse und der Interpretation dieses Textkorpus entwickelt.

Ein Ziel des Projekts aus der Sicht der Informationswissenschaft besteht in der Erprobung und Bewertung von nutzerzentrierten Methoden hinsichtlich ihrer Anwendbarkeit in Projekten im Bereich der digitalen Geisteswissenschaften. In der nutzerzentrierten Entwicklung können Anwender unterschiedliche Rollen einnehmen: In einem traditionellen Paradigma helfen Beobachtungsstudien oder Interviews bei der Erfassung der bestehenden Verhaltensweisen von Anwendern. Die Ergebnisse bilden die Grundlage für die Gestaltung neuer Lösungen, die dann unter der Beteiligung repräsentativer Anwender evaluiert werden. Dagegen sieht der Ansatz des partizipativen Designs die Anwender in einer aktiveren Rolle, etwa durch Workshops oder kollaboratives Prototyping (Muller 2012). Im Rahmen dieses Projekts wird versucht, beide Perspektiven zu berücksichtigen (Abbildung 1).

Die primäre Zielgruppe der zu entwickelnden Werkzeuge sind die Historiker, die direkt am Projekt beteiligt sind. Damit handelt es sich um Nutzer mit hoher Expertise in ihrer Fachdomäne und mit spezifischen Anforderungen. Zu Beginn des Projekts wurden im Rahmen einer Kontextanalyse Interviews am Arbeitsplatz mit Wissenschaftlern aus diesem und aus vergleichbaren Projekten geführt, um ihre allgemeinen Vorgehensweisen und Strategien bei der Auswertung von historischen Texten zu erheben. Gleichzeitig sind die Historiker im Projekt *Welt der Kinder* auch aktiv an allen Planungsschritten beteiligt und stehen in direktem Kontakt mit den Entwicklern der Werkzeuge. Daher wurden die Ergebnisse der Studie nicht direkt auf die Entwicklung übertragen, sondern dienen als Grundlage für die Diskussion mit allen Projektbeteiligten über die Anforderungen im Projekt. Auf dieser Grundlage wurde ein interaktiver, funktionaler Prototyp entwickelt, der erste Analysen des Textkorpus ermöglicht.

Parallel wurden erste Modelle für die statistische Textanalyse entwickelt und iterativ durch die Geisteswissen-

Abbildung 1: Aktivitäten und Ergebnisse der nutzerzentrierten Bedarfsanalyse.

schaftler evaluiert. In diesem Prozess wurden vor allem Verbesserungsmöglichkeiten und Evaluierungskriterien für die Modelle diskutiert. Basierend auf dem Prototyp wurde eine interaktive Studie durchgeführt, bei der jeweils ein Historiker mit einem Informationsspezialisten aus dem Projekt für die Analyse existierender Thesen aus der Geschichtswissenschaft kooperierte, um dadurch die statistische Textmodellierung aus fachwissenschaftlicher Sicht zu überprüfen.

In diesem Beitrag werden die angewendeten Methoden vorgestellt und diskutiert. Die nächsten Abschnitte geben dafür zunächst einen Überblick über Nutzerstudien in den digitalen Geisteswissenschaften und über die Anforderungen an die Textanalyse im Bereich der digitalen Geschichtswissenschaft.

Nutzerstudien in den Digital Humanities

In der Informationswissenschaft wird ein breites Spektrum von nutzerzentrierten Methoden angewendet (Kelly 2009, S. 10). Dieses reicht von Evaluierungen der Retrievalalgorithmen durch Relevanzbewertungen von Nutzern, über Studien zur Untersuchung und Optimierung der Informationsinteraktion in bestimmten Systemen bis hin zu Analysen des Informationssuchverhaltens (Womser-Hacker, Mandl 2013) und von Sense-Making-Prozessen (Pirolli, Card 2005) unabhängig von bestimmten Informationssystemen. Während in Projekten zur Entwicklung von Anwendungen in den Digital Humanities die technische Perspektive häufig noch im Vordergrund zu stehen scheint, wird

¹ Zugänglich über <http://gei-digital.gei.de>

auch hier zunehmend die Anwendung nutzerzentrierter Methoden für die zielgerichtete Unterstützung existierender Prozesse gefordert (Gibbs, Owens 2012; Warwick 2012). Die unterschiedlichen Konzeptionen, die von den Anwendern von Werkzeugen im Bereich der Digital Humanities existieren, zeigen sich an der großen Bandbreite dieser Werkzeuge: Sie reichen von spezialisierten Bibliotheken für das Programmieren eigener Skripte,² über konfigurierbare Werkzeuge³ hin zu Anwendungen, die für einen Einsatzzweck und einen festen Datensatz optimiert sind.⁴

Historiker erwarten von Werkzeugen für die Textanalyse primär eine Steigerung der Produktivität ihrer bestehenden Vorgehensweisen (Gibbs, Owens 2012). Die Teilnehmer einer Interviewstudie äußern sich zunächst skeptisch gegenüber Werkzeugen, die ihnen Methoden vorgeben, und erwarten für diesen Fall, dass die neuen Methoden sich in ihre bestehenden Abläufe integrieren. Für Nutzerstudien zur Anforderungserhebung in den Digital Humanities empfiehlt Warwick (2012) daher kontextbezogene Studien, bei denen das Vorgehen während der Arbeit beobachtet werden kann. Prototypen und Wireframes sollten erst im späteren Verlauf von Projekten zum Einsatz kommen und in der tatsächlichen Arbeitsumgebung überprüft werden (Warwick 2012, S. 5). In Bezug auf das Informationsverhalten von Historikern liegen dabei bereits allgemeine Erkenntnisse vor: Rhee (2012) identifiziert als typische Tätigkeiten den Überblick über die vorhandenen Informationsquellen, die kontextuelle Einordnung von historischen Quellen und ihre kritische Bewertung hinsichtlich der untersuchten Fragestellungen.

Anforderungen an die Textanalyse in der digitalen Geschichtswissenschaft

Das Korpus im Projekt *Welt der Kinder* stellt das erste digitale Korpus historischer Schulbücher mit einem derartigen Umfang für den deutschsprachigen Raum dar und eröffnet dadurch neue Zugänge für die historische Forschung. Die allgemeine Verfügbarkeit von digitalisierten historischen Primärquellen verändert die Voraussetzun-

gen der Arbeit von Historikern insgesamt, etwa durch die umfangreiche Verfügbarkeit von historischen Zeitungen und Zeitschriften in digitaler Form (Allen, Sieczkiewicz 2010). In der Disziplin wird jedoch häufig ein allgemeines Misstrauen gegenüber dem Erkenntnispotenzial von quantitativen Studien mit den Mitteln des Text Minings festgestellt (vgl. etwa Gibbs, Cohen 2011, S. 69).

Während in Studien aus der digitalen Literaturwissenschaft mit Stil und Vokabular häufig sprachliche Phänomene im Vordergrund stehen, sind in der Geschichtswissenschaft inhaltliche Aspekte und Aussagen sowie ihre kontextuelle Einordnung von größerer Bedeutung. Nach einer Einschätzung von Robertson (2014) spielen dabei Verfahren zur Lokalisierung von Textelementen die wichtigste Rolle, insbesondere zur Darstellung von Ortsnennungen oder Nennungen historischer Ereignisse auf Karten.⁵ Einen weiteren Ansatz bilden einfache Analysen von Termhäufigkeiten für die Analyse von inhaltlichen Entwicklungen (Gibbs, Cohen 2011). Methoden für die lexikalische Analyse aus der Korpuslinguistik, wie z. B. Kookkurrenzanalysen von Termen und ihre Darstellung im Satzkontext als Konkordanzen bei Baker et al. (2008), ermöglichen zusätzlich auch Aussagen über Wertungen inhaltlicher Konzepte in den Texten.

Ansätze, die mit statistischen Verfahren für das Text Mining arbeiten, gehen über die Untersuchung der sprachlichen Kontexte einzelner Terme hinaus. Die inhaltliche Analyse eines Tagebuchs (Blevins 2010) und die Untersuchung von Zeitungstexten aus dem amerikanischen Bürgerkrieg (Nelson 2010) sind zwei häufig zitierte Beispiele. Beide verwenden sogenannte Topic-Modelle (*Topic Models*) für die Analyse. Topic-Modelle sind ein in den digitalen Geisteswissenschaften verbreitetes, unüberwachtes textstatistisches Verfahren zur Aufdeckung allgemeiner Themen und inhaltlicher Entwicklungen in Textsammlungen. Dafür werden Terme, die häufig gemeinsam auftreten, als Themen interpretiert, welche statistisch als latente Variablen (Topics) modelliert werden. Den Topics sind damit über einen Wahrscheinlichkeitswert Terme zugeordnet, woraus auf das Auftreten von Topics in einem Dokument geschlossen wird (Latent Dirichlet Allocation / LDA, Blei, Ng, Jordan 2003).

Die im Rahmen des Projekts *Welt der Kinder* vorgesehenen Methoden für die Textanalyse setzen insbesondere auf den Ansatz der statistischen Textanalyse mit Hilfe von Topic-Modellen. Durch die nutzerzentrierte Bedarfsana-

² Z. B. die Sammlung von Tutorials auf <http://programminghistorian.org>

³ Z. B. das in CLARIN-D entwickelte *Weblicht* für die Erstellung komplexer Verarbeitungsketten für die automatische Sprachverarbeitung (Hinrichs, Zastrow, Hinrichs 2010).

⁴ Z. B. im Rahmen des Projektes *Shaping the West*: <http://web.stanford.edu/group/spatialhistory/cgi-bin/site/project.php?id=997>

⁵ Beispiele umfassen etwa eine Analyse der Gebiete, über die in einer Zeitung schwerpunktmäßig berichtet wird (Blevins 2014) oder ein Werkzeug zur Analyse von Migrationsbewegungen (Blessing, Kuhn 2014).

lyse können die statistischen Modelle optimiert und die Analysemethoden in Hinblick auf ihre Eignung aus fachlicher Sicht überprüft werden. Als Grundlage werden dafür zunächst allgemeine, werkzeugunabhängige Vorgehensweisen bei der Analyse von Schulmedien in der Geschichtswissenschaft untersucht.

Kontextanalyse im Projekt Welt der Kinder

Die Methode der Kontextanalyse (Contextual Inquiry – Holtzblatt, Wendell, Wood 2004) sieht kontextuelle Interviews am Arbeitsplatz vor, um existierende Vorgehensweisen unter dem Einsatz existierender Werkzeuge und improvisierter Hilfsmittel als Zwischenrepräsentationen im Informationsverarbeitungsprozess zu erfassen. Für das Projekt *Welt der Kinder* wird damit im Sinne einer Domänenanalyse die fachliche Praxis bei der Untersuchung und Auswertung von Schulmedien untersucht. Dabei haben vier Teilnehmer einen geschichtswissenschaftlichen und ein Teilnehmer einen politikwissenschaftlichen Hintergrund. Weiterhin wird auf der Grundlage der Ergebnisse die spezifischen Anforderungen im Projekt *Welt der Kinder* diskutiert. Die Auswertung und Übertragung finden gemeinsam mit allen Projektbeteiligten statt, wodurch im Sinne der partizipativen Entwicklung die interdisziplinäre Kommunikation gefördert wird.

Kontextuelle Interviews werden idealerweise am Arbeitsort begleitend zu einer Beobachtung der tatsächlichen Arbeit eingesetzt. Der Durchführende nimmt dabei eine Rolle als Lehrling (*Apprentice*) ein, stellt Verständnisfragen zu den beobachteten Abläufen und bezieht sich dabei direkt auf die eingesetzten Werkzeuge und Arbeitsobjekte (Holtzblatt, Wendell, Wood 2004, S. 80). Im vorliegenden Fall sind die zu untersuchenden Analyseprozesse langfristig angelegt, was es erschwert, diese direkt zu beobachten. Daher wird der Analyseprozess retrospektiv an einzelnen Beispielen auf der Grundlage von Arbeitsergebnissen und anderen verwendeten Artefakten nachvollzogen. Während der Interviews werden Proben von Artefakten gesammelt, unter anderem Bildschirmansichten zur Verwendung von Werkzeugen, Fotos von Papierdokumenten sowie Auszüge von in digitaler Form vorliegenden Ergebnissen. Die Interviews dauern jeweils zwischen zwei und vier Stunden. Alle Interviews werden aufgezeichnet (Audio) und protokolliert. Bei den Teilnehmern aus dem Projekt werden im Anschluss an den retrospektiven Teil auch Beispiele für Analysen aus dem Projekt anhand des Korpus auf gei-digital.gei.de thematisiert.

Die Auswertung, Interpretation und Übertragung der Ergebnisse der Kontextanalyse erfolgt in mehreren Schritten. Eine *Konzeptanalyse* dient als Grundlage für ein gemeinsames Grundverständnis im Projekt, während eine *Prozessanalyse* die Erweiterung der Anforderungen und die Diskussion der Projektziele ermöglicht.

Konzeptanalyse

Für die Konzeptanalyse werden aus den Interviews zentrale Konzepte der Fachdomäne extrahiert, mit einem besonderen Fokus auf die verwendeten Quellen und Werkzeuge sowie auf die Ergebnisse von Analysen. Beispiele aus den insgesamt 139 identifizierten Konzepten sind *Neuauflagen*, deren Aufnahmekriterien relevant sind für die Beurteilung der Zusammensetzung des Korpus, der *zeitliche Bezug* der Inhalte relativ zu dem *Erscheinungsjahr* von Werken, die Unterscheidung von *Schulbuchtypen*, wie z.B. Lehrerbücher oder Hilfsbücher, relevante *Einheiten für die Analyse* (Sätze, Absätze, Kapitel, Autoren, ...) und unterschiedliche Kontextebenen, wie z.B. das *Thema und die Zielrichtung einer Quellensammlung*.

Mit den gesammelten Konzepten wird im Rahmen eines Treffens der Projektbeteiligten ein offenes *Affinity Diagramming* durchgeführt: Jeweils zwei Teilnehmer aus unterschiedlichen Disziplinen diskutieren dafür zunächst die Zusammengehörigkeit von Konzepten und ordnen diese gemeinsam zu Konzeptgruppen. Im Anschluss werden diese Konzeptgruppen von allen Beteiligten benannt. Aus diesem Vorgehen ergibt sich eine Grundlage für eine gemeinsame Terminologie, sowohl zu Konzepten in der Fachdomäne als auch zu technischen Konzepten aus dem Bereich des Text Minings und der Computerlinguistik. Ausgehend von den in den Interviews erhobenen Konzepten konnten etwa die Rolle und Unterscheidung von Metadatatypen im Projekt sowie Formen der Zusammenarbeit bei der Organisation von Metadaten erörtert werden.

Prozessanalyse

In einem weiteren Analyseschritt werden die Ergebnisse der Interviews hinsichtlich der durchgeführten Arbeitsschritte für die Analyse von historischen Schulmedien ausgewertet. Das Vorgehen ist angelehnt an die *Consolidated Sequence Models* bei Holtzblatt et al. (2004): Vergleichbare Arbeitsschritte unterschiedlicher Teilnehmer werden dafür zunächst zu Aktivitäten zusammengefasst und dadurch von den Bedingungen bei einzelnen Interviewteilnehmern abstrahiert (Abbildung 2). Dabei ist zu beachten, dass die

Abbildung 2: Aktivitätsmodelle als Vorlage für Diskussion.

einzelnen Aktivitäten optional sind und nicht in jedem der untersuchten Projekte auftreten. Die Reihenfolge der einzelnen Aktivitäten kann sich zwischen Projekten unterscheiden und es können Iterationen auftreten.

Die Analyse der Ergebnisse zeigt, dass die Vorgehensweisen abhängig sind von der jeweiligen Fachdisziplin, von dem jeweiligen Forschungsinteresse und von den untersuchten Medientypen. Quellen werden sowohl in digitalen Formaten als auch in gedruckter Form verwendet. Zu Beginn werden häufig quantitative Analysen des vor-

handenen Quellenmaterials durchgeführt (siehe Abbildung 2 – Aktivitätsmodell 1), z. B. zu der Anzahl vorhandener Quellen oder den darin enthaltenen relevanten Textstellen und Medientypen. Als Werkzeuge für diese Analyse kommen z. B. Tabellenkalkulationsprogramme (Excel-Dokumente für Quellenstatistiken) oder Übersichtsfunktionen für Bildsammlungen zum Einsatz. Die Ergebnisse werden jedoch nicht als ein primäres Ergebnis des Forschungsprozesses betrachtet. Eine solche quantitative Analyse kann eher dazu dienen, die Forschungsziele an

dem vorhandenen Quellenmaterial auszurichten oder die zu untersuchende Materialmenge auf eine Untermenge einzuschränken, die vom Umfang her bearbeitet werden kann.

Die Auswertung des Materials (Abbildung 2 – Aktivitätsmodell 2) findet gemeinsam mit einer parallelen Analyse des historischen Kontextes der Quelle statt, etwa in Bezug auf den Sammlungskontext, den Publikationskontext, eine zeitliche Einordnung zu historischen Ereignissen oder den vorherrschenden Einstellungen einer Periode. Die Ergebnisse bestätigen die wichtige Rolle der Kontextualisierung für die Interpretation von Texten (Rhee 2012). Bei der Auswertung wird dafür auf zusätzliche Primär- und Sekundärquellen als auch auf eigenes Hintergrundwissen zurückgegriffen. Die Auswertung und Interpretation einzelner Quellen ist dann eng mit dem Schreibprozess verknüpft: Zwischenergebnisse der Analyse werden meist als Notizen in Dokumenten gesammelt, aus denen dann der interpretierende Text entsteht. Daher lässt sich dieser Prozess nur schwer rekonstruieren.

Die identifizierten Aktivitäten aus dem Forschungsprozess dienen als Grundlage für die Diskussion von Anforderungen im Projekt *Welt der Kinder*. Die Aktivitätsmodelle werden während eines Projekttreffens diskutiert und aus den Aktivitäten Ideen für die Werkzeugunterstützung im Projekt abgeleitet (*Walk the model* bei Holtzblatt et al. 2004). Diskutiert wurde dabei u. a. die Rolle der quantitativen Textanalyse: Wie in den kontextuellen Interviews wird diese von den Fachwissenschaftlern auch im Projekt primär als ein Einstieg verstanden, der dabei helfen kann, erste Hypothesen zu entwickeln und relevante Dokumente für die weitere Einzelauswertung und Interpretation zusammenzustellen. Problematisiert wurde vor allem, dass die Kriterien für die Zusammenstellung des Korpus rückblickend schwer zu rekonstruieren sind. Bei quantitativen Analysen sind demnach Verzerrungen in der Verteilung im Korpus zu berücksichtigen. Die Ergebnisse der Analyse bilden die Grundlage für Beschreibungen von Szenarien zum Vorgehen bei der automatisch unterstützten Textanalyse im Projekt und für erste grafische Prototypen. Parallel dazu werden entwickelte Analysemodelle und ein funktionaler Prototyp evaluiert.

Iterative Evaluierung von Modellen für die Textanalyse durch Nutzer

Im Rahmen des Projekts *Welt der Kinder* sollen Topic-Modelle als primäres Verfahren für die Textanalyse verwendet werden. Um Topic-Modelle für ein Korpus und eine

Forschungsfrage zu optimieren, können eine Reihe von Parametern des Modellierungsprozesses beeinflusst werden (Blei 2014), wobei die vorgegebene Anzahl der Topics einen zentralen Parameter für die Beeinflussung des Modellierungsprozesses darstellt. Für das Projekt sind mehrere Generationen von Modellen erstellt worden, die jeweils Modelle mit unterschiedlichen Anzahlen von Topics umfassen (u. a. mit 50, 100, 200 Topics). Die Unterschiede zwischen den Modell-Generationen liegen in den durchgeführten Vorverarbeitungsschritten (heuristische Korrektur von OCR-Fehlern, Auswahl von Wortarten) und der für die Modellierung angenommene Dokumenteneinheit (Textseiten, einzelne Sätze). Um eine iterative Verbesserung der Modelle zu ermöglichen, müssen die Ergebnisse aus der Perspektive von Anwendern aus der Fachdomäne bewertet werden. Ziel dieser ersten, qualitativ ausgerichteten Bewertung ist die Auswahl und Gewichtung unterschiedlicher Bewertungskriterien. Die Modelle können in drei Dimensionen bewertet werden, hinsichtlich der Qualität einzelner Topics, ihres Bezugs zu den Forschungsfragen des Projekts und ihres Einsatzes für das Retrieval relevanter Dokumente für die weitere Analyse.

In Bezug auf die inhärente Qualität der Topics wird die Kohärenz und Verständlichkeit einzelner Topics bewertet. Als Grundlage für die Bewertung versuchen die Fachwissenschaftler in einem ersten Schritt ein Topic eindeutig zu benennen, entweder nur auf der Basis der wichtigsten, dem Topic zugeordneten Begriffe oder durch das zusätzliche Aufrufen von Dokumenten, die dem Topic zugeordnet sind. Aufgrund des hohen Aufwands werden dabei in späteren Evaluierungsrunden nur die Topics auf diese Weise bewertet, welche für das Projekt als relevant eingestuft wurden. Weiterhin werden systematische Fehler in den Termen, die den Topics zugeordnet sind, identifiziert, die vor allem durch Probleme bei der automatischen Texterkennung entstehen. Bei der qualitativen Analyse der Bewertungen der Topics und von Kommentaren der Fachwissenschaftler fällt auf, dass inhaltliche Kohärenz häufig als ein eindeutiger Bezug zu historischen Ereignissen, Perioden oder räumlichen Gebieten verstanden wird. Gleichzeitig wird aber auch die Fähigkeit gefordert, das Auftreten von zunächst überraschenden, aber im Nachhinein anhand von Beispieltextrn nachvollziehbaren Zusammenhängen zwischen Themen im Korpus abzubilden.

Die Abdeckung der für das Projekt relevanten Themenfelder, etwa „fremde“ Kontinente, Länder und Völker, Kolonien, Entdeckungen, Expeditionen oder Missionserfahrungen, wird zusätzlich jeweils für ein gesamtes Modell eingeschätzt. Für die drei bewerteten Topic-Modelle werden im Mittel acht Prozent der Topics als sehr relevant für das Projekt und 29 Prozent als relevant bewertet. Die Ab-

Abbildung 3: Ansicht des funktionalen Prototyps mit für die Analyse ausgewähltem Topic.

deckung der Themenfelder wird nur in einem Teil der Bereiche als ausreichend bewertet, was in anderen Experimenten mit themenspezifischen Korpora nicht der Fall zu sein scheint (DiMaggio, Nag, Blei 2013; Evans 2014).

Im Verlauf der Evaluierung wird die Möglichkeit, auf relevante Dokumente für ein Topic zuzugreifen, als eine für das Projekt zentrale Anforderung erkannt. Die zugeordneten Dokumente können bei der Interpretation der Topics helfen. Weiterhin ist diese Fähigkeit zentral für die Überprüfung von Ergebnissen der quantitativen Analyse über Topics und für die Auswahl von Dokumenten für die hermeneutisch-interpretative Auswertung. Die Zuordnung von Dokumenten sollte daher für die weitere Optimierung von Topic-Modellen durch nutzerzentrierte Retrievalexperimente und die Bewertung der Zuordnung von Topics zu Dokumenten mit der Methode der Topic Intrusion (Chang, Gerrish, Wang, Boyd-Graber, Blei 2009) systematisch verbessert werden. Zusätzlich zu den beschriebenen Evaluierungskriterien für die interne Qualität der Topic-Modelle müssen jedoch die externe Validität der erstellten Modelle und ihre Effektivität für die Untersuchung von Fragestellungen aus der Geschichtswissenschaft untersucht werden. Für weitere Experimente zur Effektivität der Methoden muss zunächst eine Möglichkeit geschaffen werden, um das Korpus mit der Hilfe der Modelle interaktiv analysieren zu können.

Entwicklung eines funktionalen Prototyps

Um erste Analysen sowohl auf dem Inhalt des gesamten Korpus als auch auf Untermengen zu ermöglichen, wurde ein funktionaler Prototyp entwickelt (Abbildung 3). Dieser wird auf der Basis der Retrievalsoftware Apache Solr⁶ umgesetzt. Neben der Volltext-Indexierung der Texte auf der Ebene von Buchseiten wurden relevante Metadaten und Werte für die zugeordneten Topics in einen Index integriert. Dies ermöglicht dynamische, quantitative Auswertungen über die Statistikfunktionen von Solr. Die Oberfläche basiert auf einem Templating-Werkzeug⁷, welches die einfache Generierung von Suchoberflächen und damit das datengetriebene Prototyping ermöglicht. Dynamische Diagramme von Häufigkeitsverteilungen und Topic Intensität über Zeit wurden mithilfe des Visualisierungs-Frameworks D3⁸ erstellt. Der Prototyp ermöglicht den beteiligten Fachwissenschaftlern einen verbesserten Zugang zum Quellenkorpus und erste quantitative Analysen.

⁶ <http://lucene.apache.org/solr/>

⁷ VelocityResponseWriter: <http://wiki.apache.org/solr/VelocityResponseWriter>

⁸ <http://d3js.org>

Interaktive Validierung der Analysemodelle

Eine taskbasierte, interaktive Anwendung der vorhandenen Analysemodelle auf geschichtswissenschaftliche Thesen durch die Nutzung des interaktiven Prototyps bietet die Möglichkeit, die Effektivität der entwickelten Verfahren für realistische Fragestellungen zu untersuchen. In einer Validierungsstudie werden dabei Thesen untersucht, zu denen bereits belastbare Ergebnisse entweder aus anderen Untersuchungen oder als erwartbare Auswirkungen historischer Ereignisse vorliegen. Da es aufgrund des großen Umfangs des untersuchten Textkorpus nicht manuell möglich ist, die statistischen Modelle zu überprüfen, können Aussagen über die Validität der Modelle getroffen werden, indem die Analyseergebnisse mit den bereits vorliegenden Ergebnissen zu den Thesen verglichen werden. Derartige Validierungen für Topic-Modelle (DiMaggio, Nag, Blei 2013; Evans 2014) können einen wichtigen Beitrag zur Optimierung der Analysemodelle und für die Bildung von Vertrauen in die Aussagekraft von statistischen Modellen leisten. Die folgenden geschichtswissenschaftlichen Thesen wurden in der Studie mithilfe der bestehenden Werkzeuge untersucht:

- These 1: Die Themen *Kolonien* und *Auswanderung* sind ab 1893 stärker in Erdkunde- als in Geschichtsbüchern vertreten.
- These 2: *Kolonien* werden in Volksschulbüchern weniger thematisiert.
- These 3: Der Anteil der Themen *Französische Revolution* und *Befreiungskriege* steigt nach 1890 relativ zum Themengebiet *Antike*.
- These 4: Die *deutsche Kriegsflotte* wird in Erdkunde und Geschichtsbüchern gleich stark thematisiert.

Während der Studie zur interaktiven Validierung kooperiert jeweils ein Fachwissenschaftler mit einem Experten für die zur Verfügung gestellten Analysemöglichkeiten. Dadurch können die Möglichkeiten des funktionalen Prototyps ausgeschöpft und fehlende Funktionalität durch externe Werkzeuge ausgeglichen werden. Einschränkungen, die bei der Untersuchung der Forschungsfragen auftreten, können damit primär auf die verwendeten Topic-Modelle zurückgeführt werden. Gleichzeitig werden auch termbasierte Analysen mit frei durch die Geschichtswissenschaftler formulierten Anfragen durchgeführt. Zu jeder These schätzen die Fachwissenschaftler ein, ob sie eindeutig belegt oder widerlegt werden konnte, oder ob im Gegenteil keine eindeutigen Schlussfolgerungen möglich

waren. Zusätzlich wird die Grundlage für diese Einschätzung diskutiert.

Die Ergebnisse der Studie zeigen, dass die Mehrzahl der Thesen anhand von zeitlichen Entwicklungen und Vergleichen zwischen Untermengen des Korpus zumindest tendenziell bestätigt werden können. Die Belastbarkeit der Ergebnisse wird dann am höchsten eingeschätzt, wenn die term- und die topicbasierte Analyse übereinstimmen. Die Analyse mittels Topics aus den Topic-Modellen führt in einigen Fällen dazu, dass die Fragestellung auf einer allgemeineren Ebene untersucht werden muss, da die vorhandenen Topics nicht spezifisch genug sind. In anderen Fällen sind die Topics zu spezifisch und müssen für die quantitative Analyse miteinander kombiniert werden. Auch mit den formulierten Termanfragen ließen sich bereits erste belastbare Ergebnisse für die Überprüfung von Thesen erzielen, wobei jedoch von Systemseite eine bessere Unterstützung bei der Anfrageformulierung notwendig wäre.

Ein hoher Anteil an Interaktionen entfällt auf die Begutachtung von Dokumenten: Einerseits wurde die Trefferliste durchsucht, um Dokumente zu identifizieren, die relevant sind für die aktuelle Fragestellung. Hier zeigte sich für viele topicbasierte Anfragen eine geringe Genauigkeit in den Ergebnissen, etwa wenn das Themengebiet „Auswanderung“ neben Diskursen über das 19. Jahrhundert auch in Texten über die Antike oder das Mittelalter auftritt. In einigen Fällen waren jedoch gerade die von der Intention einer Anfrage abweichenden Suchtreffer interessant und konnten zu neuen Fragestellungen führen, was ein Teilnehmer so ausdrückte: „[...] dass der Historiker immer im Wechselspiel mit dem Text seine These verschärft und weiterentwickelt.“ Dieses Vorgehen und die kontinuierliche Überarbeitung der Ausgangsthese verdeutlichen den iterativen Charakter der interpretativen Analyse.

Das für das Projekt neu konzipierte Vorgehen der interaktiven Validierung ist aufgrund der interdisziplinären Kooperation vergleichbar zu Studien mit der Methode des *Co-Discovery*, in der zwei Nutzer für die Lösung von Aufgaben miteinander kooperieren. Als Ergebnis wird dabei die Kommunikation zwischen den Teilnehmern ausgewertet (vgl. etwa Rieh et al. 2010). Daher liefert diese Studie als sekundäres Ergebnis auch Hinweise auf Nutzungsanforderungen auf der Ebene der Interaktion. Beispielsweise wird die Suche nach relevanten Topics noch nicht ausreichend unterstützt, was sich daran äußerte, dass häufig die Textsuchfunktion des Browsers für die Suche in umfangreichen Listen von Topics eingesetzt wurde. Außerdem müssen quantitative Vergleiche meist in externen Werkzeugen durchgeführt werden. Diese Aktivitäten sollen in zukünftigen Versionen besser unterstützt werden.

Insgesamt wurde das Vorgehen von allen Beteiligten als in hohem Maße produktiv und kommunikationsfördernd eingeschätzt. Eine Einschränkung der Ergebnisse ergibt sich daraus, dass es sich bei den Fragestellungen in dieser Studie um die Untersuchung bestehender Ergebnisse und damit um *konfirmatorische* Analysen handelt. Ein wichtiges Ziel des Projekts sind jedoch *explorativ* ausgerichtete Untersuchungen: Selbstdefinierte Termanfragen sind etwa für explorative Untersuchungen weniger geeignet. Zudem sollte in dem System aktiv auf signifikante Unterschiede zwischen den aktuell untersuchten Dokumentenmengen und auf auffällige Entwicklungen über Zeit hingewiesen werden.

Bewertung der benutzerzentrierten Methoden

Eine Fokussierung auf technische Aspekte und eine damit einhergehende unzureichende Kommunikation zwischen den Beteiligten können in den digitalen Geisteswissenschaften ein Problem darstellen (vgl. etwa Lin 2012). Die Erfahrungen aus dem Projekt zeigen, dass nutzerzentrierte Methoden die Kommunikation und das gegenseitige Verständnis innerhalb der multidisziplinären Projektgruppe fördern können. Es sollte jedoch vermieden werden, den Fachwissenschaftlern dadurch eine passive Rolle zuzuschreiben. Der Dialog zwischen den Projektbeteiligten kann dadurch gefördert werden, dass nutzerzentrierte Vorgehensweisen möglichst partizipativ gestaltet und die Fachwissenschaftler an den Auswertungsprozessen beteiligt werden.

Durch die Kontextanalyse konnten bereits zu Beginn des Projekts wichtige Anforderungen identifiziert werden. Die Analyse von Konzepten unterstützte dabei den Aufbau einer gemeinsamen Terminologie im Projekt und ein gegenseitiges Grundverständnis für die jeweiligen Vorgehensweisen und Ziele. Die Ableitung von Ideen für neue Ziele aus den bestehenden Abläufen war den Beteiligten zwar zunächst fremd, führte aber dennoch zur Artikulation spezifischer Erwartungen und Anforderungen. Die identifizierten Anforderungen gehen dabei über die Kapazitäten des aktuellen Projekts hinaus.

Die iterative Evaluierung der entwickelten Modelle ermöglicht Einblicke in die Priorisierung von Bewertungskriterien und hilft dabei, Fehlerquellen zu identifizieren. Der Aufwand für die Durchführung der Evaluation durch die Fachwissenschaftler stellte sich als höher heraus als erwartet. Daher fehlen noch quantitative Evaluierungsmaße, die im Rahmen des Projektes auch operationalisierbar sind.

Die kooperative und interaktive Validierung der Modelle anhand von bestehenden Thesen schafft ein gemeinsames Verständnis von der Art der Fragestellung durch die gemeinsame Interpretation und Operationalisierung der untersuchten Thesen. Ermöglicht wird dieses Vorgehen erst durch den Zugang zu den Analysemodellen und zu dem gesamten Korpus über die Interaktionsmöglichkeiten des funktionalen Prototyps. Der Aufwand für den Betrieb dieser Plattform und der gleichzeitigen Konzeption einer Version für den Produktiveinsatz ist jedoch nicht zu vernachlässigen.

Durch die Nutzerstudien sind neue Ideen und innovative Ansätze entstanden, mit denen die geschichtswissenschaftliche Analyse großer Textkorpora unterstützt werden kann. Im weiteren Projektverlauf werden die Anforderungen an die nächste Generation der Werkzeuge aus Nutzersicht formuliert und die Priorität der Umsetzung gemeinsam bewertet. Weiterhin werden neue Interaktionsmöglichkeiten und Visualisierungen für die explorative Analyse entwickelt und in Nutzertests untersucht. Für die Einbindung und Partizipation der Fachwissenschaftler sind weitere kooperative Experimente vorgesehen.

Projektförderung

Das Projekt *Welt der Kinder* (<http://welt-der-kinder.gei.de/>) ist angesiedelt am *Georg-Eckert-Institut*, Braunschweig, und wird gemeinsam mit dem *Deutschen Institut für Internationale Pädagogische Forschung*, der *TU Darmstadt*, der *Universität Hildesheim* und weiteren Projektpartnern von der *Universität Zürich*, der *Bayrischen Staatsbibliothek*, der *Universitätsbibliothek Braunschweig* und dem *Göttingen Centre for Digital Humanities* durchgeführt. Das Projekt wird von der Leibniz-Gemeinschaft unter der Fördernummer SAW-2014-GEL-2 gefördert.

Literatur

- Allen, R. B.; Sieczkiewicz, R. (2010). How historians use historical newspapers. *Proceedings of the American Society for Information Science and Technology*. Bd. 47, Nr. 1, S. 1–4. DOI 10.1002/meet.14504701131.
- Baker, P.; Gabrielatos, C.; KhosraviNik, M.; Krzyzanowski, M.; McNery, T.; Wodak, R. (2008). A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. *Discourse & Society*. Bd. 19, Nr. 3, S. 273–306. DOI 10.1177/0957926508088962.

- Blei, D. M.; Ng, A. Y.; Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*. Bd. 3, S. 993–1022.
- Blei, D. M. (2014). Build, Compute, Critique, Repeat: Data Analysis with Latent Variable Models. *Annual Review of Statistics and Its Application*. Bd. 1, Nr. 1, S. 203–232. DOI 10.1146/annurev-statistics-022513-115657.
- Blessing, A.; Kuhn, J. (2014). Textual Emigration Analysis (TEA). In: *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*.
- Blevins, C. (2010). Topic Modeling Martha Ballard's Diary. *Cameron Blevins* [online]. <http://www.cameronblevins.org/posts/Topic-Modeling-martha-ballards-diary/> [11.08.2015].
- Blevins, C. (2014). Space, Nation, and the Triumph of Region: A View of the World from Houston. *Journal of American History*. Bd. 101, Nr. 1, S. 122–147. DOI 10.1093/jahist/jau184.
- Chang, J.; Gerrish, S.; Wang, C.; Boyd-Graber, J. L.; Blei, D. M. (2009). Reading tea leaves: How humans interpret topic models. In: *Advances in neural information processing systems*. S. 288–296.
- DiMaggio, P.; Nag, M.; Blei, D. (2013). Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of U. S. government arts funding. *Poetics*. Bd. 41, Nr. 6, S. 570–606. DOI 10.1016/j.poe-tic.2013.08.004.
- Evans, M. S. (2014). A Computational Approach to Qualitative Analysis in Large Textual Datasets. *PLoS ONE*. Bd. 9, Nr. 2, S. e87908. DOI 10.1371/journal.pone.0087908.
- Gibbs, F.; Owens, T. (2012). Building Better Digital Humanities Tools: Toward Broader Audiences and User-Centered Designs. *Digital Humanities Quarterly*. Bd. 6, Nr. 2.
- Gibbs, F. W.; Cohen, D. J. (2011). A Conversation with Data: Prospecting Victorian Words and Ideas. *Victorian Studies*. Bd. 54, Nr. 1, S. 69–77. DOI 10.2979/victorianstudies.54.1.69.
- Hinrichs, M.; Zastrow, T.; Hinrichs, E. W. (2010). WebLicht: Web-based LRT Services in a Distributed eScience Infrastructure. In: *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*.
- Holtzblatt, K.; Wendell, J. B.; Wood, S. (2004). *Rapid Contextual Design: A How-to Guide to Key Techniques for User-Centered Design*. Morgan Kaufmann. ISBN 0-12-354051-8.
- Kelly, D. (2009). Methods for Evaluating Interactive Information Retrieval Systems with Users. *Foundations and Trends in Information Retrieval*. Bd. 3, Nr. 1–2, S. 1–224. DOI 10.1561/15000000012.
- Lin, Y. (2012). Transdisciplinarity and Digital Humanities: Lessons Learned from Developing Text Mining Tools for Textual Analysis. In: *Understanding digital humanities*. 1. Basingstoke: Palgrave Macmillan. ISBN 978-0-230-29265-9.
- Muller, M. J. (2012). Participatory Design: the Third Space in HCI. In: *The Human-Computer Interaction Handbook Fundamentals, Evolving Technologies, and Emerging Applications*. Boca Raton, FL: CRC Press. S. 165–185.
- Nelson, R. K. (2010). Mining the Dispatch. <http://dsl.richmond.edu/ispatch/pages/home> [11.06.2015].
- Pirolli, P.; Card, S. (2005). The Sensemaking Process and Leverage Points for Analyst Technology as Identified through Cognitive Task Analysis. In: *Proceedings of International Conference on Intelligence Analysis*. Mitre McLean, VA. S. 2–4.
- Rhee, H. L. (2012). Modelling Historians' Information-Seeking Behaviour with an Interdisciplinary and Comparative Approach. *Information Research*. Bd. 17, Nr. 4.
- Robertson, S. (2014). The Differences between Digital History and Digital Humanities. <http://drstephenrobertson.com/blog-post/the-differences-between-digital-history-and-digital-humanities/> [02.02.2015].
- Warwick, C. (2012). Studying Users in Digital Humanities. *Digital Humanities in Practice*. S. 1–21.
- Womser-Hacker, C.; Mandl, T. (2013). Information Seeking Behaviour (ISB). In: *Grundlagen der praktischen Information und Dokumentation: Handbuch zur Einführung in die Informationswissenschaft und -praxis*. 6. De Gruyter Saur. ISBN 3-11-025822-6.

Dr. Ben Heuwing

Institut für Informationswissenschaft und Sprachtechnologie
Universität Hildesheim
Universitätsplatz 1
31141 Hildesheim
Deutschland
heuwing@uni-hildesheim.de

Ben Heuwing arbeitet als wissenschaftlicher Mitarbeiter am Institut für Informationswissenschaft und Sprachtechnologie der Universität Hildesheim. Er hat eine Promotion im Themengebiet Wissensmanagement für Usability-Ergebnisse abgeschlossen. Der aktuelle Arbeitsschwerpunkt liegt in der Anforderungsanalyse und Interaktionsgestaltung für die Textanalyse in den Digital Humanities. Weitere Forschungsbereiche umfassen nutzerzentrierte Methoden und Entwicklungsprozesse in Organisationen, Studien zum Informationsverhalten sowie die Gestaltung von interaktiven Anwendungen für Information Retrieval und Informationsanalyse.

Prof. Dr. Christa Womser-Hacker

Institut für Informationswissenschaft und Sprachtechnologie
Universität Hildesheim
Universitätsplatz 1
31141 Hildesheim
Deutschland
womser@uni-hildesheim.de
www.uni-hildesheim.de/~womser

Christa Womser-Hacker ist Professorin für Informationswissenschaft an der Universität Hildesheim. Sie leitet als Direktorin das dortige Institut für Informationswissenschaft und Sprachtechnologie. Ihre Forschungsschwerpunkte liegen im Bereich des Internationalen Informationsmanagements, insbesondere auf der Mensch-Maschine-Interaktion und dem mehrsprachigen und domänenspezifischen Information Retrieval. Christa Womser-Hacker promovierte und habilitierte an der Universität Regensburg. Derzeit ist sie Mitglied im Vorstand des Hochschulverbands Informationswissenschaft und in den Leitungsgremien der Fachgruppe Information Retrieval und des Fachbereichs Mensch-Maschine-Interaktion der Gesellschaft für Informatik.