


Ein Gefühl für die Daten entwickeln. Eine ethnografische Annäherung an große Textdaten am Beispiel digitaler Chats

Libuše Hannah Vepřek

Zusammenfassung

Die Bewältigung und Analyse großer digitaler Textdaten, wie digitaler Chat- oder Social Media-daten, stellt Ethnograf:innen noch immer vor Herausforderungen. Trotz zahlreicher Entwicklungen in den Digital Humanities und Computational Social Sciences stehen qualitative, induktive und subjektzentrierte Ansätze zum Umgang mit großen Textdaten bislang weniger im Fokus methodischer Lösungsansätze. Am Beispiel digitaler Chatdaten stellt dieser Beitrag ein Vorgehen vor, das große Textdaten strukturiert und analysiert, und ordnet dieses Vorgehen in den ethnografischen Forschungsprozess ein. Das Vorgehen greift auf relationale Datenbanken und SQL als Datenbankabfragesprache zurück, um Chatdaten zu verwalten, umstrukturieren und befragen zu können, sodass diese für die qualitative Analyse handhabbar werden. Der Einbezug von Chatdaten ermöglicht nicht nur neue Erkenntnisse. Vielmehr ermöglicht das Vorgehen Ethnograf:innen gerade zu Forschungsbeginn, ein Gefühl für die Daten zu entwickeln und Nähe zum Feld herzustellen. Der Text diskutiert Möglichkeiten und Grenzen sowohl des Vorgehens als auch der Analyse schriftlicher Chatdaten im ethnografischen Forschungsprozess.

Schlagwörter: Relationale Datenbanken, SQL, qualitative Analyse, Exploration, Chatdaten

Libuše Hannah Vepřek, Institut für Empirische Kulturwissenschaft und Europäische Ethnologie, Ludwig-Maximilians-Universität München, Deutschland 

Große Datenmengen in ethnografischer Forschung

Die Digitale Anthropologie und das Forschen im und mit dem Internet haben sich inzwischen als fester Bestandteil einer ethnografisch arbeitenden empirischen Kulturwissenschaft etabliert. Dennoch scheint Thomas Hengartners Feststellung, die er Anfang der 2000er Jahre formulierte, auch heute noch zu greifen: „Mit einigermaßen gezielter Suche [...] lassen sich heute Informationen in Überfülle über das Netz recherchieren und finden, so dass weniger deren Beschaffung, als vielmehr die Bewältigung, Sichtung und Beurteilung vor Herausforderungen stellt“ (Hengartner 2007 [2001]: 190–1). Informationen können hierbei unterschiedlichster Art sein. Sie können von Nachrichtenbeiträgen, Artikeln und Blogposts über Prozess- und Metadaten oder Quellcode zu Foreneinträgen und Chatnachrichten reichen. In diesem Beitrag möchte ich einen Vorschlag zur Bewältigung dieser Herausforderung, die ich

während meiner ethnografischen Feldforschung zu Human-Computation-basierten Citizen-Science-Spielen selbst erfahren habe, beitragen. In meiner Forschung stellten sich die textlichen Chatdaten des Fallbeispiels *Stall Catchers* als hilfreicher Zugang und als wertvolle Quelle heraus, nach dem ich das hier vorgestellte Vorgehen entwickelt habe. Mit Blick auf den ethnografischen Forschungsprozess lässt sich dieses Vorgehen zwischen einer Strukturierung von Daten und deren Analyse einordnen. Ich nehme dabei den Umgang mit besonders großen Textdaten aus Chats mit dem Fokus auf schriftliche Chatnachrichten in den Blick, in denen die Kommunikation grafisch realisiert ist (vgl. Beißwenger 2000: 39). Das vorgestellte Vorgehen ist jedoch ebenso auf andere digitale Daten anwendbar, sofern diese grundsätzlich in Tabellen gespeichert werden können. Das Verfahren richtet sich an alle Ethnograf:innen, ob mit oder ohne Informatikkenntnissen. Zwar mag das Arbeiten mit relationalen Datenbanken sowie die Syntax und technischen Spezifika zunächst unvertraut wirken. Ähnlich wie bei der Auseinandersetzung mit neuen Apps oder Computerprogrammen bietet es jedoch nach anfänglicher Gewöhnung viele Möglichkeiten. Um den Einstieg zu erleichtern, stelle ich im Text deshalb auch Werkzeuge vor, die keine Programmier- oder Datenbankabfragesprachenkenntnisse voraussetzen.

Die Analyse von großen digitalen Textdaten ist in den Sozialwissenschaften kein neues Problem. In den Digital Humanities und den Computational Social Sciences wurden besonders in den letzten Jahren große Fortschritte in der Entwicklung (semi-)automatischer Analysemethoden von digitalen Textdaten erzielt, die sich dabei besonders für diskursanalytische und andere textintensive Vorgehen eignen (vgl. Lemke & Wiedemann 2016). Weniger im Fokus scheinen bislang qualitative, induktive und subjektzentrierte Ansätze zu stehen, die für ein kulturwissenschaftliches, ethnografisches Vorgehen besonders interessant sind und daher im Fokus dieses Beitrags stehen. Denn wie ich im Folgenden zeigen werde, können Chatkonversationen Themen und Bedeutungen zu Tage führen, die im Forschungsprozess mit anderen Methoden unartikuliert bleiben oder nicht teilnehmend beobachtbar sind. Zudem bietet die Umstrukturierung der Daten mittels relationaler Datenbanken einen explorativen Zugang und eine tiefgehende Auseinandersetzung mit den Daten, die über das lineare Lesen und Analysieren von Chatdaten, wie oft in der qualitativen Analyse üblich, hinausgeht.

Konkret möchte ich zweierlei diskutieren: Zum einen die Potenziale und Grenzen der qualitativen Analyse schriftlicher Chatdaten für die ethnografische Forschung sowie deren Besonderheiten; und zum anderen relationale Datenbanken, d.h. Datenbanken, in denen Informationen in tabellarischer Form organisiert werden, wobei verschiedene Tabellen miteinander in Beziehung stehen können (siehe unten für eine ausführlichere Erläuterung), als Möglichkeit, diese großen Chatdaten für eine qualitative Analyse zu organisieren und handhabbar zu machen. Denn aufgrund des meist großen Datenvolumens stößt die rein manuelle Analyse von Chatnachrichten oftmals an ihre Grenzen. Anhand eines Beispiels aus meiner Forschung zu Human-Computation-basierten Citizen-Science-Projekten zeige ich exemplarisch, wie dies in der Praxis aussehen kann. Darüber hinaus gehe ich auf ausgewählte ethische Fragen, die bei der Analyse von Chatdaten beachtet werden sollten, ein. Der Beitrag versteht sich als ein Ausgangspunkt für weitere Beschäftigungen mit digitalen Chatdaten und relationalen Datenbanken in der Empirischen Kulturwissenschaft.

Qualitative Chatanalyse mittels relationaler Datenbanken im Methodenspektrum der Empirischen Kulturwissenschaft

Während das Vorgehen eine Handhabung großer Textdaten für die qualitative Analyse ermöglicht und als explorativer erster Zugang zu empirischen Daten dient, stellt es jedoch kein umfassendes, eigenständiges Analyseverfahren dar. Vielmehr ist es als ein Teil eines breiteren methodischen und analytischen Prozesses zu verstehen. Die Chatanalyse spielte besonders am Anfang meiner Forschung eine wichtige Rolle, um einen indirekten Zugang zu den Praktiken um und Bedeutungen von *Stall Catchers* zu erhalten. Die Chatanalyse, die mittels Datenbanken qualitativ realisierbar wird, kann jedoch nicht als einziger Zugang ethnografischer Forschung dienen. Michi Knecht beschreibt treffend: „Methoden sind [...] nicht einfach herauszulösen aus einem Gesamtzusammenhang von Ethnografie, dessen ontologische, epistemologische, methodologische, ethische und pragmatische Dimensionen ein Gefüge bilden, welches sich im Forschungsprozess immer wieder neu austariert.“ (Knecht 2013: 82) Ethnografische Forschung lebt von der Kombination verschiedener Methoden. Während die Analyse des *Stall-Catchers*-Chats Bedeutungen aufdeckte, die während qualitativer Interviews mit Teilnehmer:innen nicht angesprochen wurden, eröffneten Interviews, Teilnehmende Beobachtungen sowie die Analyse verschiedenen Quellenmaterials weitere Perspektiven auf *Stall Catchers*. Diese Perspektiven wurden wiederum nicht über die Chatanalyse deutlich. Außerdem benötigt das für das eigene Forschungsvorhaben sinnvolle und erkenntnisbringende Abfragen der Chatdaten in der Datenbank Vorwissen, das meist iterativ (vgl. Koch & Franken 2020) und unter Einbezug verschiedener Methoden gesammelt wird. Gerade in der Kombination der verschiedenen methodischen Zugänge wird *Stall Catchers* in seiner Mehrdimensionalität, seinen multiplen Bedeutungen und dem Zusammenspiel verschiedener menschlicher Akteur:innen, nicht-menschlicher Entitäten und Materialitäten erfahrbar. Um einen Rahmen zu bilden, beschreibe ich im nächsten Abschnitt zunächst knapp den Forschungskontext, in dem das hier diskutierte Vorgehen entwickelt wurde. Dabei zeige ich auch, weshalb die Umstrukturierung und Analyse von Chatdaten gewinnbringend für den ethnografischen Forschungsprozess sein kann.

Das Forschungsbeispiel *Stall Catchers*

In dem webbasierten Citizen-Science-Projekt *Stall Catchers* tragen Teilnehmer:innen spielerisch zur Analyse von Forschungsdaten aus der Alzheimer-Forschung eines US-amerikanischen biomedizinischen Techniklabors bei (<https://stallcatchers.com/>). *Stall Catchers* beschreibt gleichzeitig ein sogenanntes Human-Computation-System, da (noch) nicht maschinell lösbare Aufgaben an Teilnehmende ausgelagert werden und diese gemeinsam mit Algorithmen und Software das Analyseproblem lösen. Die Aufgabe, als kurze Filmabschnitte präsentierte mikroskopische Bildaufnahmen zu klassifizieren, ist dabei von einem spielerischen Punktesystem gerahmt. Die repetitive Aufgabe wird zudem durch die Möglichkeit, mit anderen Teilnehmer:innen und Teammitgliedern zu chatten, etwas aufgebrochen. In meiner Forschung gehe ich der Frage nach, wie sich Human-Computation-basierte Citizen-Science-Projekte wie *Stall Catchers* im Zusammenspiel verschiedener Akteur:innen und Materialitäten im Spannungsfeld Spiel und Forschung formieren. Dazu verfolge ich ein multiperspektivisches und auf Methodentriangulation beruhendes Vorgehen, das etablierte Me-

thoden mit explorativen Ansätzen wie der „fokussierten Codeanalyse“ (siehe auch den Beitrag von [Vepřek et al. in diesem Band](#)) und eben dem hier dargelegten Vorgehen kombiniert. Die Erforschung von *Stall Catchers* bestand vor allem aus Teilnehmender Beobachtung, etwa im alltäglichen Spiel oder bei besonderen Veranstaltungen, und der Mitarbeit im *Human Computation Institute*, das *Stall Catchers* entwickelt. Zudem führte ich qualitative Interviews mit Teilnehmer:innen, Teammitgliedern sowie Forscher:innen durch und analysierte verschiedene mediale Beiträge von und über das Projekt. Während ich die Perspektive der Entwickler:innen und Forscher:innen hinter dem Projekt bei einem späteren Feldaufenthalt in Ithaca, NY, großteils in Person erheben konnte, blieb mein Zugang zu den meisten Teilnehmer:innen auf das Internet und Telefon beschränkt, da diese global verteilt sind. Schnell wurde beim eigenen Spielen und dem explorativen Lesen des im Spiel integrierten Chats deutlich, dass dieser einen bedeutenden Aspekt von *Stall Catchers* bildete. Denn, und dies gilt ebenso für viele weitere Online Spiele, das Chatfenster stellt den primären Kommunikationsweg dar. Teilnehmer:innen kommunizieren während des Spiels miteinander oder mit dem Team des *Human Computation Institutes*, stellen Fragen oder berichten von Problemen mit der Plattform. Auch das Team nutzt den Chat, um über Entwicklungen des Projekts, neue Datensätze, Wartungsarbeiten, etc. zu informieren und Feedback einzuholen.

Der Einbezug der Chatverläufe in meine Forschung erwies sich in vielerlei Hinsicht als gewinnbringend: Zum Beispiel spielt er für viele Teilnehmer:innen, die den Chat aktiv nutzen, eine wichtige Rolle für die Motivation zur Teilnahme, da sie sich gegenseitig unterstützen und anfeuern können. Gleichzeitig dient er auch der Herstellung eines Wir-Gefühls. Die Praxis der Einzelnen wird zu einem geteilten Erlebnis. Außerdem begegnete ich durch das Lesen des Chatverlaufs Aspekten und Themen, die bis dahin in meiner Forschung nicht aufgetaucht waren. Oft handelt es sich dabei um Dinge, die mit dem Spiel und *Stall Catchers* direkt zusammenhängen, in meinen Gesprächen mit Teilnehmer:innen jedoch nicht aufkommen waren. So tauschten sich Teilnehmer:innen beispielsweise über alternative Interpretationen der zu analysierten Bilder aus. Anstelle der mikroskopischen Aufnahmen wurden die Bilder spielerisch als Galaxien interpretiert und das Finden einer ‚Supernova‘ im Chat mitgeteilt. Des Öfteren war auch die Rede vom Sammeln und Einlösen von Punkten und wie viel es Teilnehmer:innen bedeuten würde, ein neues Level zu erreichen oder einen eigenen Rekord zu brechen. Dies wurde in den Interviews kaum erwähnt. Die Analyse der Chatnachrichten stellte sich daher als ein wertvoller indirekter Zugang zum Feld dar, der es erlaubt, neue Einsichten und eine erste Vertrautheit mit den Daten zu gewinnen. Um diese Einblicke zu erhalten, musste ich jedoch zunächst die Datenmenge bewältigen, denn mir standen zum Zeitpunkt der Analyse Anfang 2020 Chatdaten von über einem Jahr zur Verfügung (und in anderen Forschungen stehen möglicherweise Daten aus Jahrzehnten zur Verfügung). Die Daten rein manuell wie andere empirische Daten zu kodieren, stellte sich schnell als nicht realisierbar heraus. Aus dieser Feststellung und der Frage, wie ich diese vorhandenen Daten für meine Forschung nutzbar machen kann, ist schließlich das im Folgenden vorgestellte Vorgehen entstanden. Bevor ich mich diesem zuwende, möchte ich jedoch kurz auf die medienspezifischen Affordanzen und Charakteristika von Chats eingehen, die beim Einbezug dieser Daten mitzudenken sind.

Chatspezifische Affordanzen und Charakteristika

Digitale (textliche) Chats haben spezifische Charakteristika und Materialitäten, die es in der Forschung zu beachten gilt. Es ist daher zunächst wichtig, diese technischen und medien-spezifischen Aspekte zu beachten, um zu verstehen, wie beispielsweise (welche Form von) Kommunikation vom Medium afforziert wird (Gibson 1982; Bareither 2020). Michael Beißwenger geht in *Kommunikation in virtuellen Welten: Sprache, Text und Wirklichkeit* (2000) der technischen Funktionsweise und der Medialität von Chats nach. Medienspezifische Aspekte wirken ihm zufolge auf die möglichen Kommunikationsformen ein, unterstützen die einen, während es die anderen erschwert. Zudem unterscheiden sich textliche Chats von mündlichen Kommunikationsformen. Gleichzeitig werden jedoch die Strukturen des Alltags in Internetkommunikation allgemein nicht etwa aufgehoben, sondern vielmehr reproduziert oder sogar verstärkt (vgl. Schönberger 2015; Ullrich & Schieck 2015). Zu den spezifischen Kommunikationsformen zählt, dass die Unterhaltungen in digitalen Chats wie in Webforen zumeist „polylogical“ (Pink et al. 2016: 109) verlaufen. In den Worten von Sarah Pink et al. bedeutet das, dass „es weder monologisch noch dialogisch ist, sondern drei oder mehr Konversationspartner:innen involviert“ (ebd.; Übersetzung LHV). Im Fallbeispiel *Stall Catchers* handelt es sich um ein kleines Chatfenster, das in einer Ecke des Spiels platziert ist und im eingeklappten Modus einem ungeöffneten Browser Tab ähnelt. Wenn es dagegen mit einem Klick geöffnet wird, ist seine Präsenz kaum zu ignorieren, da jede eingehende Nachricht den Chatverlauf nach oben schiebt und somit Bewegung in diesem Bildschirmbereich stattfindet. Zudem lädt das Chatfenster, das nur eine Zeile des getippten Textes anzeigt, bevor auf Senden gedrückt wird (Abbildung 1), zu kurzen Nachrichten ein.



Abbildung 1: Das Eingabefenster des Stall Catchers Chats.

Quelle: Bildschirmaufnahme vom 20.07.2023.

In seiner aktuellen Fassung erlaubt der *Stall-Catchers*-Chat keine Darstellung von Bildern oder Filmen, wie es bei gegenwärtigen Messengerdiensten der Fall ist. Letztere müssen stattdessen als Link eingefügt werden, während einige Smilies mittels Copy-and-Paste angenommen werden. Daneben ist der *Stall-Catchers*-Chat *single-threaded*: Das heißt, dass Benutzer:innen parallele Gespräche in einem Fenster führen müssen und nicht in private Threads ausweichen können. Auch die hinter den Chats liegenden Infrastrukturen und die Kommunikation ermöglichenden technischen Protokolle strukturieren die mögliche Kommunikation und geben u.a. vor, wie und wo die Chatinhalte und Metadaten gespeichert werden oder wer wie Zugriff auf welche Daten erhält.

Neben diesen Aspekten wird die Kommunikation in digitalen Chats oft durch ausgewählte Moderator:innen oder Administrator:innen moderiert. Das ist für die Analyse von Chatdaten bedeutend, da diese vorgeben, wie Konversationen ablaufen dürfen und jederzeit in diese eingreifen können. Dadurch ergeben sich spezifische Machtstrukturen, die es zu beachten gilt. Im Fallbeispiel *Stall Catchers* verfolgen Administrator:innen regelmäßig die Nachrichten und können zum Beispiel Nachrichten verstecken oder Nutzer:innen blocken, wenn sie ungewünschte Inhalte teilen oder gegen die Regeln, denen alle Teilnehmer:innen bei der Registrierung zu *Stall Catchers* zustimmen mussten, verstoßen. In digitalen Citizen-Science-Projekten wird die Rolle der Moderation oft auch an ausgewählte Teilnehmer:innen

übertragen. Bei *Stall Catchers* übernehmen sie oft auch die Aufgabe, Fragen zum Spiel oder der Forschung im Chatfenster zu beantworten oder Probleme an das Team weiterzugeben. Sie wirken daher auch als Mittler:innen zwischen Teilnehmer:innen und dem *Stall-Catchers*-Team. Des Weiteren ist die Frage der Zugänglichkeit und Sichtbarkeit der Chats und damit auch der Chatdaten zu bedenken. Der *Stall-Catchers*-Chat ist für alle offen und sichtbar, die sich auf der Plattform registriert haben. Somit kann grundsätzlich jede:r, der:die über eine E-Mail-Adresse verfügt, Zugang zum Chatfenster erhalten. Nach der Registrierung müssen alle Teilnehmer:innen den AGBs und Datenschutzbestimmungen zustimmen, die u.a. beinhalten, dass das Chatfenster für alle eingeloggten *Stall-Catchers*-Teilnehmer:innen öffentlich und sichtbar ist (vgl. Terms and Privacy Policy *Stall Catchers*). Die genannten Spezifika digitaler Chats bilden den Kontext und Rahmen für die Inhalte der Konversationen. Je nach Forschungsfeld können dabei weitere Spezifika eine Rolle spielen, die besonders durch die Kombination verschiedener Methoden beleuchtet werden können. Im Folgenden möchte ich mich nun dem Vorgehen zur Analyse solcher textlichen Chatnachrichten zuwenden.

Von quantitativen zu qualitativen Ansätzen

Es gibt bereits unterschiedliche Ansätze, große textliche Datensätze zu analysieren. Wie eingangs erwähnt, wurden besonders in den Digital Humanities und Computational Social Sciences (semi-)automatisierte Ansätze zur Analyse von digitalen Textdaten entwickelt. Die meisten dieser Ansätze folgen einem quantitativen Vorgehen und basieren auf Text Mining, das die (semi-)automatisierte Extraktion von Informationen wie Strukturen, Regelmäßigkeiten etc. aus unstrukturierten Textdaten erlaubt (siehe zur Anwendung aus der Fachperspektive der Empirische Kulturwissenschaft [Franken et al. in diesem Band](#); weitere Literatur siehe Manderscheid 2019; Franken 2023). Franco Moretti (2016) hat für diese algorithmischen Ansätze zur Textanalyse den Begriff *distant reading* im Gegensatz zum *close reading* eingeführt. Letzteres bezieht sich auf die menschliche Analyse von Textdaten und ist durch die Menge der Daten, die in einer bestimmten Zeit analysiert werden können, eingeschränkt. Andere Autor:innen, wie etwa Matthias Lemke (2017) haben mit dem Konzept *blended reading* dagegen einen Kompromiss zwischen beiden Ansätzen vorgeschlagen, indem es sowohl *distant reading* als auch das Lesen individueller, repräsentativer Texte kombiniert. Ähnlich versucht Martin Müller mit *scalable reading* eine Synthese von Nähe und Distanz zu erreichen, um die neuen Möglichkeiten des Pendelns zwischen beiden zu nutzen (Müller o.J.).

Neben diesen quantitativ ausgerichteten Ansätzen existieren Überlegungen und methodische Vorschläge für qualitative Ansätze zur Analyse von schriftlichen online Konversationen (siehe beispielsweise Schirmer et al. 2015; Nam 2019). Insgesamt bleibt die Frage, wie große Datenmengen für eine qualitative Analyse handhabbar gemacht werden können, jedoch noch wenig diskutiert. Das hier präsentierte Vorgehen ist mit solchen qualitativen Analysemethoden kombinierbar, da es sich gerade zwischen der Strukturierung und Analyse von digitalen Daten verortet. Es nutzt relationale Datenbanken, um die erhobenen Chatdaten zu strukturieren und organisieren, sodass diese leichter für eine qualitative inhaltliche Analyse und Kodierung zugänglich werden. Es lässt sich zudem mit anderen softwaregestützten Analysen verbinden. Während bei Machine-Learning-basierten Text-Mining-Modellen die Modelle selbst und/oder die Daten, auf denen sie trainiert wurden, nicht immer transparent sind, schlage ich ein Verfahren vor, in dem Ethnograf:innen direkter und flexibler, wenngleich vermittelt über ein Datenbankmanagementsystem, mit den Daten interagieren können. Zudem lädt es im Vergleich zu klassifikatorischen Text-Mining-Modellen dazu

ein, die Daten lupenhaft und im Detail durch unterschiedliche Abfragen und Strukturierungen kennenzulernen. Es eignet sich daher besonders für einen explorativen Modus, der dann auch zur Entscheidung, weitere softwarebasierte Analyseverfahren aus den Digital Humanities heranzuziehen, führen kann.

Bei dem hier vorgestellten Vorgehen handelt es sich also um eine explorative Annäherung an empirische Daten, das besonders dann, wenn die Daten nicht selbst erhoben wurden, die Möglichkeit bietet, *ein Gefühl für die Daten zu entwickeln*. Gerade zu Beginn der Forschung oder des Analyseprozesses ist es sinnvoll, um sich mit dem vorhandenen Material unterschiedlich auseinanderzusetzen. Das Vorgehen ist im Forschungsprozess zwischen Datenbeschaffung und Datenanalyse zu verorten, da die Auswahl von Material und das Strukturieren von Daten immer auch analytisch relevant ist und Richtungen sowie Perspektiven leitet. Dieser Zugang zu und Umgang mit den Daten erlaubt schließlich eine andere Perspektive auf das Forschungsfeld und eine Vertrautheit zu entwickeln, die beispielsweise Themen in das Blickfeld der:des Ethnograf:in rückt, die in qualitativen Interviews oder Teilnehmenden Beobachtungen möglicherweise nicht aufkommen würden. Durch eine Umstrukturierung der Daten ist es außerdem möglich, nicht nur größere Datenmengen auszuwerten, sondern auch für die Forschung weniger wichtige Teile auszuschließen und auf interessante Inhalte zu fokussieren.

Damit schließt dieser Beitrag zum Beispiel auch an Gertraud Koch und Lina Franks Überlegungen zum Prinzip des Filterns an. Dieses bildet immer einen wichtigen Teil der Forschung zur Reduktion der Komplexität des Forschungsfelds. Die Autorinnen setzen hier besonders für die wissenssoziologische Diskursanalyse manuelle und maschinelle Vorgehen gewinnbringend in Bezug (2020). Das Hin- und Herwechseln zwischen den qualitativ zu analysierenden Inhalten und dem Datenkorpus und seine Strukturierung könnte dabei wie Müllers *scalable reading* als Pendeln zwischen verschiedenen Skalen verstanden werden. Gleichzeitig ermöglicht die Relationalität der Datenbank besonders gut, einen akteurszentrierten Ansatz zu wählen und so im Sinne kulturwissenschaftlicher Fallbeispiele oder Porträts (Bourdieu & Accardo 1998; Gutekunst & Rau 2017) die Bedeutungen, mit denen Akteur:innen ihre Lebenswelten und ihr Handeln aufladen, nachzuvollziehen. Das Vorgehen erlaubt es, auf individuelle Bedeutungen und Interpretationen innerhalb großer Datenmengen und somit auf bestimmte Perspektiven zu fokussieren. Dabei stellt das im Folgenden präsentierte Vorgehen eben nicht die ‚eine‘ neue Methode vor. Vielmehr möchte ich mit dem Vorgehen auf die Nützlichkeit relationaler Datenbanken für die qualitative Analyse von großen Textdaten aufmerksam und diese handhabbar machen. Im nächsten Abschnitt erläutere ich relationale Datenbanken und Datenbanksprachen und diskutiere, wie diese für die Analyse von Chatdaten nützlich sein können.

Was sind relationale Datenbanken?

Datenbanken spielen in digitalen Speicher- und Datenverarbeitungssystemen grundlegende Rollen zum dauerhaften Speichern und Ordnen von Daten. Sie sind aber auch mehr als bloße Datenspeicher, denn sie sind einheitlich organisierte Datensätze, deren Datenstrukturen es uns ermöglichen, verschiedene Fragen an die Daten zu stellen. Harvey Quamen und Jon Bath haben treffend formuliert, wann für Forscher:innen in den Sozial- und Kulturwissenschaften eine Datenbank nützlich sein kann:

Infobox: Relationale Datenbanken. Der vorliegende Beitrag kann kein vollständiges mathematisches und theoretisches Verständnis von relationalen Datenbanken und Datenbankabfragesprachen vermitteln. Der Fokus liegt stattdessen auf Grundlagen, die zur Verwendung von Datenbanken für die Strukturierung von Chatdaten wichtig sind. Weiterführende Literatur und Handbücher zur funktionalen Einführung in relationale Datenbanken, SQL und den *DB Browser for SQLite* sind zum Beispiel:

- Tutorial zu Datenmanagement mit SQL für Sozialwissenschaftler*innen: <https://datacarpentry.org/sql-socialsci/index.html> (hier ist zu beachten, dass das Tutorial nicht aktiv aktualisiert wird)
- SQL-Tutorial: <https://www.w3schools.com/sql/>
- Als Beispiel für eine einführende Liste mit SQL-Befehlen siehe <https://www.its-lot.de/2013/12/sql-befehle-fur-anfanger.html>
- Daneben sind Auflistungen mit SQL-Befehlen über gängige Suchmaschinen abrufbar.
- Zu anderen Ansätzen wie graphbasierten Datenbanken siehe z.B. Passoth & Wehner (2018).
- Einführung zu SQLite siehe z.B. Litovski et al. (2010)
- Einführung in Datenbanken siehe z.B. Schubert (2007)
- Einführung in relationale Datenbanken siehe z.B. Studer (2019)

„[I]f you are embarking upon a project in which you will be actively engaging with your data, pushing its limits, and asking challenging questions of it – finding patterns, seeing changing dynamics over time, locating anomalies, looking for missing information – then you will need a database“ (Quamen & Bath 2016: 145).

Digitale Datenbanken stellen keine „neutralen“ Speichermedien dar, vielmehr verändern sie auch Perspektiven auf Daten, die Daten selbst sowie die Arbeitspraktiken, die mit der digitalen Datenbank oder den Daten verbunden sind (Hine 2006). Marcus Burkhardt untersucht Datenbanken medientheoretisch als „spezifische, technische Infrastruktur[en] und universelle Metapher[n] digitaler Informationssammlungen“ (Burkhardt 2015: 8). Mich interessieren in diesem Beitrag dagegen weniger Datenbanken als Untersuchungsgegenstand, wie es Burkhardt vorschlägt, als vielmehr ein praktisch-reflektierter Umgang mit Datenbanken.

Digitale Datenbanken, und hier besonders relationale Datenbanken, begegnen uns im Internetalltag täglich indirekt, wie etwa bei der Suche nach neuer Musik oder Filmen auf Streaming-Plattformen, die möglicherweise dem eigenen Lieblingsfilm ähneln. Die erstellten Listen mit Filmempfehlungen basieren meist auf in Datenbanken gespeicherten Informationen (vgl. Passoth & Wehner 2018), die je nach Abfrage unterschiedlich zu Listen strukturiert werden. Die Informationen sind hierbei in den Datenbanken jedoch nicht in dem Sinne selbst als Listen gespeichert. Vielmehr können die

„sichtbaren Repräsentationen, also die Tabellen, in denen z. B. Musiktitel, Künstler, Nutzer abgelegt sind, sowie die Tabellen, die Verknüpfungen zwischen ihnen in der Form von NutzerID, TitelID oder KünstlerID, TitelID darstellen, [als] Ermöglichung der Manipulation, Selektion und variablen Zusammenstellung von Informationen, die logisch, nicht physisch verknüpft sind“ (ebd.: 63),

verstanden werden. Datenbanken können unter Beachtung forschungsethischer Aspekte als befragbare Depots empirischer Daten genutzt werden. Oft sind Daten dabei bereits vorhanden und werden nicht von dem:der Ethnograf:in erhoben, sodass Fragen der Nachnutzung von Daten aufkommen (siehe dazu [Franken et al. in diesem Band](#)).

Der heute wohl verbreitetste Ansatz und eine der stabilsten Datenbanken ist die relationale Datenbank, die seit dem Ende der siebziger Jahre besteht (vgl. Schubert 2007: 35). Diese zeichnet sich dadurch aus, dass ihr ein relationales Modell (im Sinne der relationalen Algebra) zugrunde liegt, sodass Daten in Form von sogenannten Relationen oder Tabellen abgespeichert werden (Abbildung 2). In der Praxis kann mensch sich diese zunächst als Tabellen mit Zeilen und Spalten vorstellen. Eine Tabelle ist im Prinzip eine Sammlung von Objekten der gleichen Art, wie beispielsweise eine Tabelle für Nutzer:innen. Eine andere Tabelle könnte einzelne Nachrichten mit Informationen über Sender:in und Datum enthalten. Jede Zeile beschreibt dabei ein Objekt, eben zum Beispiel eine:n Nutzer:in oder eine Nachricht. Jede Spalte beschreibt eine Eigenschaft (Attribut) des Objekts, also den Namen der:des Nutzer:in oder Textinhalt einer Nachricht. Manche Spalten enthalten Verweise auf andere Objekte in anderen oder der gleichen Tabelle, in der Regel in der Form von Identifikationsnummern; die ID der:des Nutzer:in, der:die diese Nachricht geschickt hat. Das ist hilfreich, um die Information über den:die Nutzer:in nur einmal speichern zu müssen, statt den Namen neben der Nachricht zu duplizieren. Technisch werden Tabellen selbst auch als Relationen beschrieben, die mit anderen Tabellen oder Relationen wiederum in Beziehung oder Relation stehen können.

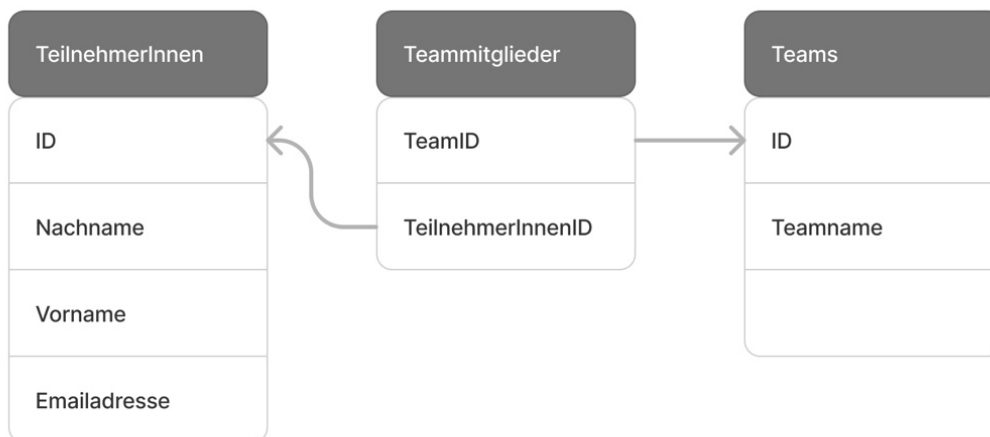


Abbildung 2: Exemplarische Darstellung dreier Relationen/Tabellen, die miteinander verknüpft sind.

Quelle: Libuše Hannah Vepřek, 2023.

Walkthrough: Analyse digitaler Chatdaten mithilfe relationaler Datenbanken am Beispiel *Stall Catchers*

Anhand des *Stall-Catchers*-Chats stelle ich nun Schritt für Schritt die Erhebung, Strukturierung und Handhabarmachung großer Chatdatensätze für die spätere Analyse mittels relationaler Datenbanken vor. Das Ziel ist es die Datenbank als Werkzeug zu nutzen, um relevante und fokussierte Ausschnitte aus den Daten zu generieren, welche dann manuell analysiert werden können. In diesem Beispiel arbeite ich mit der Datenbank [SQLite](#) und nutze den [DB Browser for SQLite](#), um die Datenbankdateien zu verwalten und Abfragen auszuführen. Die gleichen Ergebnisse können aber prinzipiell mit jeder SQL-fähigen Datenbank wie [MySQL](#) und [PostgreSQL](#) erreicht werden. Das hier präsentierte Vorgehen stellt keine fertige und sicherlich nicht die einzige Methode zur Strukturierung und Analyse digitaler Chatnachrichten dar, vielmehr ist es als eine Art Laborbericht zu verstehen, an dem ich anschließend die Möglichkeiten und Grenzen der Aufarbeitung und qualitativen Analyse von digitalen Chatdaten diskutiere.

Das Beispiel stellt eine inhaltliche Analyse der Narrationen von Spielpraktiken und Bedeutungen von *Stall Catchers* vor. Möglich wäre auch, wenngleich dies im Folgenden nicht im Zentrum steht, eine dezidiert kommunikationstheoretische Analyse von oder Kommunikation via textlicher Chats (vgl. Beißwenger 2000). Die einzelnen Schritte des Vorgehens bestehen aus 1. Datenerhebung, 2. Datenbereinigung, 3. Erstellen einer Datenbank für die vorliegenden Daten, 4. Umstrukturierung des ausgewählten Chatmaterials mit Abfragen, 5. Qualitative Analyse. In jedem dieser Schritte trifft der:die Ethnograf:in Entscheidungen, die die weitere Forschung und Analyse beeinflussen. Beispielsweise kann die Analyse umstrukturierter Chatdaten neue Einsichten generieren, die den Fokus einer Forschung verschieben. Um dies auch im Nachhinein nachvollziehen zu können empfiehlt es sich von Beginn an ein Forschungstagebuch zu führen, das die einzelnen Schritte, Entscheidungen und ihre Begründungen festhält. Zur besseren Nachvollziehbarkeit webe ich im Folgenden, wo möglich, kleine Analogien zu einem manuellen Umgang mit gedrucktem empirischem Material ein. Deutlich werden dabei sowohl Gemeinsamkeiten als auch grundlegende Unterschiede. Digitale Datenbanken ermöglichen zum Beispiel eine flexible Handhabung und Umstrukturierung von Daten, ohne den Originaldatensatz zu verändern, was bei manuellen Vorgehensweisen meist schnell an seine Grenzen stößt.

Datenerhebung

Die Erhebung von Daten zur Analyse digitaler Chats beschreibt ein nicht-reaktives Vorgehen, d.h. die Daten sind ursprünglich ohne das Eingreifen von Forscher:innen und außerhalb der Wissenschaft entstanden (Wolff 2011: 246) und somit auch nicht unmittelbar für den Zweck der Analyse gespeichert. Teilnehmer:innen sind sich zur Zeit der Erstellung und des Nachrichtenaustauschs in der Regel nicht bewusst, dass diese später analysiert werden könnten. Dies wirft auch forschungsethische Fragen auf, die ich in einem späteren Abschnitt dieses Beitrags diskutiere. Im Fallbeispiel protokolliert das *Stall-Catchers*-Team den Chatverlauf automatisiert und speichert diesen in einer Datenbank, zu welcher mir Zugang gegeben wurde. Jeder Eintrag besteht dabei unter anderem aus Sender:innen-ID, Datum, Uhrzeit der Nachricht und dem Nachrichteninhalte. Mit Erlaubnis des *Human Computation Institutes* wur-

den die Chatnachrichten, die bis zum Zeitpunkt des Abrufens gesendet wurden und ungefähr 1,5 Jahre umfassten, als CSV-Datei aus der Datenbank extrahiert. Eine CSV-Datei (*comma-separated-values*-Datei) ist eine Textdatei, in der die Werte durch Kommas getrennt werden. Diese enthielt in meinem Beispiel mehr als 17.000 Einträge bzw. Chatnachrichten. Diese Datenmenge überschreitet vielleicht noch nicht die Schwelle dessen, was als Big Data bezeichnet wird. Die manuelle Analyse jedes einzelnen Eintrags erwies sich jedoch dennoch als nicht realisierbar und, wie im Folgenden gezeigt wird, aufgrund vieler für meine Forschung uninteressanter Einträge, nicht sinnvoll.

Datenbereinigung

Diese ursprünglichen Chatdaten enthielten darüber hinaus verschiedene Informationen und Metadaten, die für meine Analyse nicht relevant schienen. Um mich auf die interessanten Informationen, nämlich die Nachrichteninhalte, Sender:innen-IDs sowie Datum und Zeitpunkt der verschickten Nachricht konzentrieren zu können, musste der Datensatz zunächst von diesen gelöst und von Datenfehlern bereinigt werden. Fehler können beispielsweise beim Exportieren des Datensatzes auftreten. Diese könnten später zu Problemen in der Strukturierung der Daten führen und sollten deshalb behoben werden. Daher stellt die Datenbereinigung, trotz des Zeitaufwands und der Mühsamkeit, die damit verbunden sein kann, eine wichtige Vorarbeit dar. Vergleichbar ist dieser Schritt mit dem Verbessern falscher oder fehlender Worte in einem Interviewtranskript. Je nachdem, in welcher Form Chatdaten vorliegen, kann die Datenbereinigung unterschiedliche Anforderungen mit sich bringen. Im Fallbeispiel konnten die überflüssigen und fehlerhaften Daten leicht in Excel ausgeschlossen bzw. korrigiert werden. Die resultierende CSV-Datei wies nun keine Fehler mehr auf, bestand jedoch nach wie vor aus derselben Datenmenge. Um einen Einblick in individuelle Perspektiven von Teilnehmer:innen zu erhalten, müsste ich also linear durch die Datei gehen

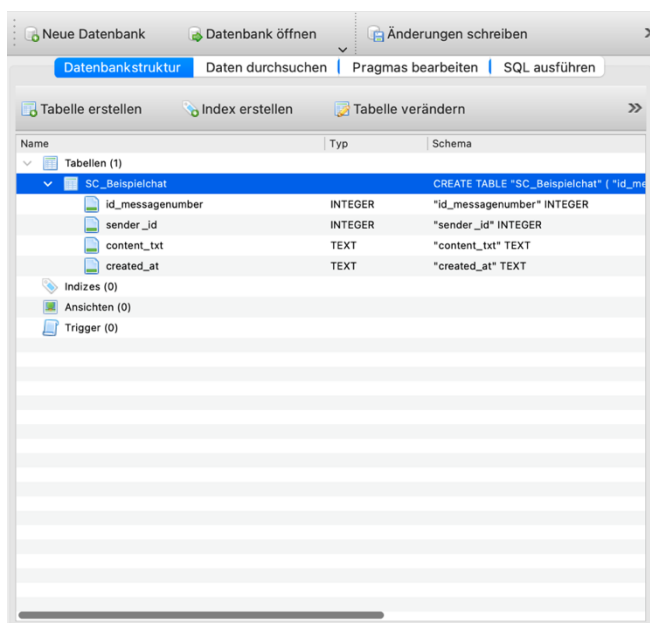


Abbildung 3: DB Browser for SQLite, Übersicht der Datenbankstruktur.

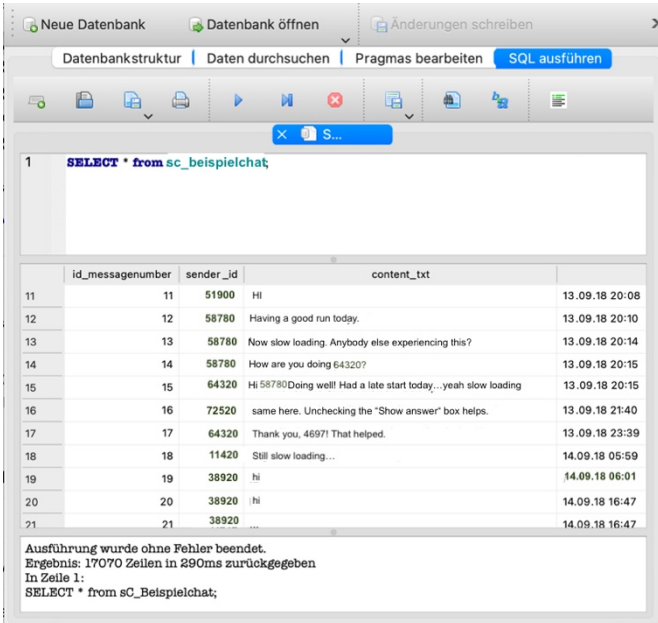
Quelle: Bildschirmaufnahme vom 25.07.2022.

und sie Zeile für Zeile lesen – bei circa 17.000 Datenpunkten eine kaum lösbare Aufgabe. Um die Chatdaten für die qualitative Analyse handhabbarer zu machen, wendete ich mich relationalen Datenbanken zu.

Erstellen einer Datenbank für die vorliegenden Daten

Ein wichtiger Schritt im ethnografischen Vorgehen ist das Sortieren und Ordnen unterschiedlichen Materials, etwa mittels eines Inventars oder Karteikästen. Bei diesem muss festgelegt werden, nach welchen Aspekten geordnet werden soll. Bei der Verwendung von Analysesoftware wie *MAXQDA* kann das Material in einer Baumstruktur hierarchisch und in verschiedenen Kategorien unterteilt werden. Der Vorteil beim digitalen Vorgehen ist, dass eine spätere Umstrukturierung leichter fällt.

Für die Arbeit mit relationalen Datenbanken wird daher als nächstes eine Datenbank erstellt, in die der vorliegende Datensatz sowie zukünftige weitere Datensätze importiert werden. Mit dem *DB Browser for SQLite* können Datenbankdateien einfach erstellt, gestaltet und bearbeitet werden. Dabei sind viele Datenbankabfragen in nutzer:innenfreundliche Menüs übersetzt worden, sodass das Programm auch für Benutzer:innen mit wenigen oder sogar keinen SQL-Kenntnissen zugänglich ist, je nach Komplexität der gewünschten Datenbankabfragen. Neue Datenbankdateien können im *DB Browser for SQLite* sowohl über SQL-Abfragen im Reiter „SQL ausführen“ oder direkt über den Button „Neue Datenbank“ erstellt werden. In die neue Datenbank können die in einer CSV-Datei vorliegenden Chatdaten dann durch Drag-and-Drop vom aktuellen Speicherort importiert werden. Falls noch keine Daten vorhanden sind, können hier auch manuell neue Tabellen erstellt werden. In Abbildung 3 ist die Beispiel-Datenbank mit der bislang einzigen Tabelle „sc_beispielchat“ zu sehen. Diese Tabelle enthält einen Chat-Datensatz aus *Stall Catchers* und besteht aus vier verschiedenen



The screenshot shows the DB Browser for SQLite interface. At the top, there are buttons for 'Neue Datenbank', 'Datenbank öffnen', and 'Änderungen schreiben'. Below that, there are tabs for 'Datenbankstruktur', 'Daten durchsuchen', 'Pragmas bearbeiten', and 'SQL ausführen'. The 'SQL ausführen' tab is active, showing a SQL query: 'SELECT * from sc_beispielchat;'. Below the query, there is a table with the following data:

	id_messagenumber	sender_id	content_txt	
11	11	51900	Hi	13.09.18 20:08
12	12	58780	Having a good run today.	13.09.18 20:10
13	13	58780	Now slow loading. Anybody else experiencing this?	13.09.18 20:14
14	14	58780	How are you doing 64320?	13.09.18 20:15
15	15	64320	Hi 58780! Doing well! Had a late start today...yeah slow loading	13.09.18 20:15
16	16	72520	same here. Unchecking the "Show answer" box helps.	13.09.18 21:40
17	17	64320	Thank you, 4697! That helped.	13.09.18 23:39
18	18	11420	Still slow loading...	14.09.18 05:59
19	19	38920	hi	14.09.18 06:01
20	20	38920	hi	14.09.18 16:47
21	21	38920	...	14.09.18 16:47

Below the table, there is a status bar that reads: 'Ausführung wurde ohne Fehler beendet. Ergebnis: 17070 Zeilen in 290ms zurückgegeben. In Zeile 1: SELECT * from sc_Beispielchat;'

Abbildung 4: Ausschnitt der Ausgabe der Ergebnisdaten im *DB Browser for SQLite*. Zum Schutz der Privatsphäre der Teilnehmer:innen wurden die Originaldaten mit fiktiven ersetzt.

Quelle: Bearbeitete Bildschirmaufnahme mit fiktiven Daten von Libuše Hannah Vepřek, 20.07.2023.

Spalten („id_messagenumber“, „sender_id“, „content_txt“ und „created_at“). Jeder Spalte ist ein Datentyp zugeordnet: Der Inhalt der Chatnachrichten sowie der Zeitstempel, an dem eine Nachricht erstellt wurde, sind als „TEXT“, während die eindeutige Kennung der Nachricht sowie die ID der: s Sender:in in „INT“, also ganzzahligen Werten gespeichert sind. Im Reiter „Daten durchsuchen“ werden nun die Daten in Form einer Tabelle wiedergegeben und können mittels Filter durchsucht werden. Hier können also sehr einfache Abfragen spezifischer Datenpunkte über die vorprogrammierte Suchleiste eingegeben werden. Im Reiter „SQL ausführen“ können aus dem Basisdatensatz strukturierte Datensätze für die Analyse erstellt werden.

Beispielhafte Umstrukturierung des ausgewählten Chatmaterials mit Abfragen

Im nächsten Schritt können nun Abfragen an das ausgewählte Chatmaterial gestellt werden und so bestimmte Daten abgefragt und/oder das Material umstrukturiert werden. Wenn wir uns das Chatmaterial auf Papier gedruckt vorstellen, kann das Vorgehen im weiten Sinne mit dem Auswählen und Ausschneiden interessanter Passagen oder der Nachrichten einer bestimmten Nutzer:in verstanden werden, die dann in neuer Reihenfolge beliebig aneinandergelagt werden können.

Das Interface des *DB Browser for SQLite* zum Ausführen von SQL-Abfragen ist in drei Fenster unterteilt: Im obersten können SQL-Abfragen eingegeben werden. Darunter werden anschließend die Ergebnisdaten der Abfrage ausgegeben. In der unteren Leiste sind schließlich Informationen über die Ausführung zu entnehmen, wie zum Beispiel in Abbildung 4 zu sehen ist.

Je nach Forschungsinteresse können nun sogenannte Views bzw. virtuelle Tabellen erstellt werden. Views beschreiben neue Tabellen, die aus den Ergebnissen einer Datenbankabfrage bestehen. In meinem Forschungsbeispiel wurde durch das explorative Lesen des Chatverlaufs deutlich, dass die meisten der über 29.000 registrierten Benutzer:innen (Stand 18. September 2020) den Chat nicht nutzten und nur ein kleiner Teil von ihnen Nachrichten über den Chat austauschen würden. Eine spätere Abfrage ergab, dass im März 2020 insgesamt 866 Teilnehmer:innen eine oder mehrere Nachrichten über den Chat verschickt hatten. Davon hatte wiederum nur ein Bruchteil mehr als 100 und nur wenige mehrere 1000 Nachrichten versendet. Die meisten Nachrichten von Benutzer:innen, die insgesamt nur eine Nachricht verschickt hatten, bestanden lediglich aus Grußformeln wie „Hi“. Zudem kam es auch bei *Stall Catchers*, wie in vielen öffentlichen Chats, trotz Moderationsbemühungen und in den AGBs und Datenschutzbestimmungen festgehaltenen Regeln hin und wieder zu Spam-Nachrichten, die aus zufällig zusammengewürfelten Zeichen o.Ä. bestehen. Da diese Nachrichten regelmäßig im Datensatz auftraten, war es für die spätere Analyse (mit der gegebenen Forschungsfrage) sinnvoll und übersichtlicher, diese auszuschließen. Mit einer ersten Abfrage stellte ich daher zunächst fest, welche Teilnehmer:innen wie viele Nachrichten verschickt hatten. Diese Informationen können mit folgender Abfrage abgerufen werden (in grün werden die einzelnen Schritte jeweils als Kommentare erläutert):

```

/* Mit SELECT werden die Spalten ausgewählt, die ausgegeben werden sollen.
In diesem Fall soll die "sender_id", da eine Zeile eine Nachricht eine*r
bestimmten Sender*in darstellt, und mit COUNT(*) die Gesamtzahl der Einträge
zu dieser „sender_id“ ausgegeben werden. */
SELECT sender_id,
       Count(*)

/* Mit FROM wird wie bereits oben erläutert die Tabelle angegeben,
auf der die Abfrage ausgeführt werden soll. */
FROM   sc_beispielchat

/* Mit GROUP BY definiere ich, dass das Ergebnis nach Sender*innen gruppiert
werden soll. */
GROUP BY sender_id

/* und schließlich mit ORDER BY, dass die Ergebnisdaten in absteigender Folge
sortiert werden sollen. */
ORDER BY Count(*) DESC;

```

Mit dieser Abfrage erhalten wir eine Tabelle, aus der wir ablesen können, welche Benutzer:innen sich wie viel in den Chat eingebracht haben. Um ein tieferes Verständnis der Perspektiven einzelner Teilnehmer:innen zu erhalten, können nun Konversationen aus dem Datensatz gefiltert werden, an der zum Beispiel ein:e bestimmte:r Teilnehmer:in teilgenommen hat. Interessant waren hierbei in meiner Forschung besonders jene Teilnehmer:innen, die den Chat regelmäßig nutzten. Es wäre aber genauso denkbar nach bestimmten Themen zu filtern oder eine Gruppe von Teilnehmer:innen in den Fokus zu setzen. Der Nachvollziehbarkeit halber soll im Folgenden ein Beispiel anhand eine:s bestimmten Teilnehmer:in mit der fiktiven ID 50001 durchgespielt werden. Dazu soll folgende Abfrage durchgeführt werden, die alle Nachrichten, die von Sender:in 50001 verschickt wurden sowie zusätzlich jeweils zwei Nachrichten vor und nach einer Nachricht von 50001, enthält.

```

/* Zunächst erstellen wir aus der Tabelle der Chatdaten eine
gefilterte Kopie mit dem Namen parent, die alle Nachrichtennummern (IDs)
der*des Sender*in mit der ID 50001 enthält. */
WITH parent
  AS (SELECT id_messagenumber
        FROM   sc_beispielchat
        WHERE  sender_id = 50001)

/* Wir wählen nun alle Spalten der sc_beispielchat-Datei zur Ausgabe aus. */
SELECT sc_beispielchat.*

/* Wir verbinden die beiden Tabellen, indem wir jene Einträge auswählen,
bei denen die +/- 2 Nachrichten-IDs der vollständigen sc_beispielchat-Tabelle
vor bzw. hinter einer Nachrichten-ID von 50001 auftreten. */
FROM   sc_beispielchat
JOIN   parent
      ON ( sc_beispielchat.id_messagenumber BETWEEN
          parent.id_messagenumber - 2 AND parent.id_messagenumber + 2 )

/* Zuletzt wird nach der Nummer der Nachrichten sortiert,
sodass eine chronologische Nachrichtenabfolge ausgegeben wird. */
ORDER BY id_messagenumber;

```

Indem zwei Nachrichten jeweils hinzugenommen werden, werden die Nachrichten von Sender:in 50001 kontextualisiert und ganze Unterhaltungen, in denen diese:r beteiligt war, extrahiert. Es wäre ebenso denkbar, den Rahmen auf drei oder vier Nachrichten zu erweitern, je nachdem, wie viele Teilnehmer:innen sich üblicherweise an Unterhaltungen beteiligen. Diese Rahmen kann jedoch ebenso später zu jeder Zeit vergrößert werden, sollten Lücken in Unterhaltungen auftreten. Daraus resultierende Unterhaltungen sehen dann beispielsweise wie in Abbildung 5 dargestellt aus.

Der Vorteil dieser Darstellung besteht darin, dass jene eben beschriebenen wiederkehrenden Spam-Nachrichten oder einzelne Grußworte, die für die Analyse im Rahmen meiner Forschung weniger relevant waren, herausgefiltert wurden. Im Fallbeispiel fokussierte ich auf einzelne Akteur:innen und ihre Perspektiven. Genauso könnte das Verfahren aber genutzt werden, um auf konkrete Themen oder Gespräche, die zu einer bestimmten Uhrzeit etc. geführt wurden, zu fokussieren. Um dies zu erreichen, müssten lediglich die Datenbankabfragen entsprechend angepasst werden. Was als interessant gilt, ist dabei immer je nach Forschungsprojekt und -frage individuell zu beurteilen. Während die Struktur der Kommunikation mit diesem Vorgehen durch die Auswahl und Umorganisation der Daten verändert wird, wird die Gesprächsdynamik selbst beibehalten. Zu beachten und reflektieren ist, dass digitale Datenbankabfragen einer binären Logik und somit klaren logischen Konditionen folgen, wodurch zwischen den Ergebnissen und dem Originaldatensatz klare Grenzen liegen und implizite Bezüge in den Chatunterhaltungen verloren gehen könnten. Dies könnte zum Beispiel der Fall sein, wenn es einen parallelen Gesprächsverlauf zum Analysierten gibt, bei dem es um Nachrichten von User:in 50001 geht, diese:r jedoch nicht auf die Unterhaltung Bezug nimmt. Da die Ergebnisse einer Abfrage nicht den Datenbestand selbst verändern, sondern virtuelle Tabellen erstellt und dann exportiert werden können, ist es zudem jederzeit möglich, zum Datensatz zurückzukehren und andere Abfragen zu stellen. Besonders für ethnografische und induktive Forschungen ist diese Flexibilität wichtig, da sich Feldforschungsphasen mit Analyse und Reflexion abwechseln und keinem strikten linearen Schema folgen. Die Ergebnistabellen können dann zum Beispiel direkt als CSV-Dateien exportiert und anschließend in Software für qualitative Datenanalyse importiert werden.

Analyse der Chatdaten

Schließlich können die umstrukturierten Daten kodiert werden, vergleichbar mit der qualitativen Analyse von anderem empirischen Material. Meine ethnografische Forschung und Analyse orientierte sich insgesamt am Vorgehen der Grounded Theory (Glaser & Strauss

2	50001	Hi 58780! How are you? What time is it in Australia?	13.09.19 10:44
3	58780	Hey 50001! Thanks, doing well. A quarter to nine. Just started catching.	13.09.19 10:45
4	50001	Great. I try to hit 400.000 points today.	13.09.19 10:46
5	58780	Another 124.500 to go then.	13.09.19 10:48
6	58780	Good luck!	13.09.19 10:48
7	50001	Thanks!	13.09.19 10:49
8	58780	I don't have much time this morning. Have to run some errands and work in the garden.	13.09.19 10:49
9	50001	Every bit helps! Yeah, lots to do here as well.	13.09.19 10:51
10	58780	You're right. I wish I had more time though.	13.09.19 10:52
11	50001	Next time!	13.09.19 10:53

Abbildung 5: Beispiel einer Ergebnistabelle nach der Umstrukturierung der Chatdaten.

Quelle: Libuše Hannah Vepřek, 20.07.2023.

1971 [1967]; Charmaz 2014), d.h. an einem induktiven, systematischen und dennoch flexiblen Forschungsansatz, der zwischen Datenanalyse und -erhebung iteriert und die Entwicklung von Theorien mittlerer Reichweite ermöglicht, die aus dem empirischen Material hervorgehen (Charmaz 2012: 2). Nachdem ich die zu analysierenden Chatfragmente beispielsweise nach Gesprächen ausgewählter Teilnehmer:innen strukturiert hatte, ließ sich mit einem akteurszentrierten Kodieren im Sinne der Grounded Theory anschließen. Vor allem bleiben Gesprächseinheiten erhalten, während Spam oder etwa einzelne kurze Begrüßungsformeln herausgefiltert werden können. Insgesamt ist ein generell für ethnografische Forschung wichtiges Hin- und Herwechseln zwischen neustrukturierten und initialen Daten jederzeit möglich. Dadurch wird ein flexibles und offenes Vorgehen unterstützt. Wenn dieses Vorgehen der Umstrukturierung von Chatdaten in der frühen explorativen Forschungsphase durchgeführt wird, können sich neue Forschungsfragen oder Themen ergeben. In meiner Forschung zu *Stall Catchers* wurde durch die Analyse des Chats unter anderem auch die Untersuchung des Umgangs der Teilnehmer:innen miteinander und mit dem Team möglich. *Stall Catchers*, so wurde deutlich, stellt für einige Teilnehmer:innen neben dem Citizen-Science-Ansatz auch einen geteilten Care-Raum dar, in denen sie ihre Alltagserfahrungen mit der Alzheimer-Krankheit teilen können.

Potenziale und Grenzen des Einbezugs relationaler Datenbanken zur Datenstrukturierung

Nach diesem Überblick über die konkreten Schritte, die eine ethnografische Arbeit mit relationalen Datenbanken umfassen kann, gehe ich abschließend auf Potenziale und Grenzen dieses Vorgehens für ethnografische Forschungen ein. Zunächst wende ich mich drei konkreten Potenzialen zu. Ein wichtiges Potenzial liegt in der bereits konstatierten Flexibilität. Forscher:innen können je nach Fragestellung selbst beschließen, welche Kommunikationselemente und Nachrichten für das Forschungsvorhaben relevant sind und welche durch eine spezifische Datenbankabfrage in der Analyse des Materials weggelassen werden können. Würde meine Fragestellung beispielsweise auf die Kommunikationsebene fokussieren, könnten gerade diese Interaktionen interessant sein und in Datenbankabfragen münden, die weniger den Inhalt der Nachrichten als vielmehr die Kommunikationsrhythmen in den Blick nehmen. Diese Flexibilität macht es gleichzeitig schwierig, in diesem Artikel ‚eine‘ konkrete Methodik zu präsentieren, da diese immer von der Fragestellung und dem konkreten Forschungsgegenstand geleitet wird. Ein weiteres Potenzial dieses Vorgehens liegt in der Möglichkeit, es nicht nur auf schriftliche digitale Chatdaten, sondern ebenso auf Tweets und Kommentare in Foren anzuwenden. Voraussetzung hierfür ist die strukturierte Speicherung der Daten. Um die Möglichkeiten für die eigene Forschung zu erschließen, empfiehlt es sich, sich explorativ mit den Daten in einer Datenbank auseinanderzusetzen und unterschiedliche Abfragen durchzuspielen. Dabei sollten die einzelnen Schritte und Entscheidungen, wie oben erwähnt, dokumentiert werden. Schließlich kann eine Vertrautheit mit der Funktionsweise von Datenbanken und Datenbankabfragesprachen auch allgemein in der Feldforschung von Vorteil sein. Wenn Zugang zu Prozessdaten vorhanden ist, ermöglicht dies Ethnograf:innen etwa, selbst weitere Informationen über Praktiken herauszufinden, die möglicherweise von den Plattformbetreiber:innen und Entwickler:innen nicht extrahiert werden, weil sie keinen direkten Nutzen für die Instandhaltung der Plattform zu haben scheinen. In meinem Beispiel konnten so über die Chatanalyse hinaus weitere Informationen über die

Spielweisen der Teilnehmer:innen einbezogen werden. Wichtig ist hier, dass diese Daten ebenfalls in den Kontext weiterer Quellen und Informationen gestellt werden müssen, um erkenntnisbringend zu sein.

Neben den genannten Potenzialen weist das Vorgehen Grenzen auf, auf die ich im Folgenden eingehen werde. Gerade die beschriebene Flexibilität kann dazu führen, dass interessante und wichtige Aspekte möglicherweise unentdeckt bleiben, wenn sie unbewusst aus den abgefragten Datentabellen ausgeschlossen wurden. Dies stellt jedoch ein Begleitproblem dar, das die meisten methodischen Zugänge besitzen, schließlich ist eine Eingrenzung des Forschungsfeldes und das Setzen von „Schnitten“ (Barad 2015: 160) ein unumgänglicher Bestandteil ethnografischen Forschens, bei dem das Forschungsfeld selbst erst kreierte wird. Zudem ist die Präsenz der Nutzer:innen sowie die Kommunikation in schriftlichen Chats immer zunächst auf schriftliche und digital darstellbare Zeichen beschränkt. Para- und nonverbale Elemente der Kommunikation fehlen genauso wie die Körperlichkeit der Nutzer:innen. Carsten G. Ullrich und Daniela Schiek (2015) haben auf diese Einschränkung in Bezug auf digitale Foren und online Gruppendiskussionen hingewiesen und hervorgehoben, dass dies Auswirkungen auf das qualitative Interpretationsverfahren haben müsse. Gleichzeitig bleibt jedoch umstritten, wie sich online Gruppendiskussionen aufgrund der fehlenden Körperlichkeit von jenen im physischen Raum unterscheiden. Denn die fehlende Körperlichkeit kann sowohl zur Herabsetzung der Hemmschwellen in der Kommunikation also auch zur Abnahme der Interaktionsdichte führen (ebd.: 134). Chatnachrichten sind darüber hinaus nicht direkt repräsentativ für alle Teilnehmer:innen einer Plattform. Im Fallbeispiel etwa beteiligte sich nur ein kleiner Prozentteil der Teilnehmer:innen am Chat. Hier wird es besonders wichtig, weitere Methoden einzubeziehen. In meinen Interviews berichteten einige Teilnehmer:innen, dass sie den Chat aktiv meiden würden, da dieser sie nur von ihrer Arbeit abhalten würde. Dass nicht alle Benutzer:innen digitaler Plattformen in digitalen Chatnachrichten repräsentiert sind, ist daher immer mitzudenken. Zudem stellt sich bei digitalen Chatdaten, wie allgemein beim Einbezug von Daten, die nicht von dem:der Ethnograf:in selbst gebildet wurden, die forschungsethische Frage nach der Privatsphäre der Chatteilnehmer:innen und ihrer Daten. Diskussionen in Internetforen, Tweets auf Twitter oder Chatdaten sind oftmals ‚öffentlich‘ in dem Sinne, dass Nutzer:innen (meist) wissen, dass ihre Nachrichten von einer breiten Leser:innenschaft gelesen werden können (siehe dazu auch [Bareither & Schramm in diesem Band](#)). Im Falle von *Stall Catchers*, bei dem Teilnehmer:innen mit dem Akzeptieren der AGBs und Datenschutzbestimmungen zustimmen, sich über die Öffentlichkeit der Chatnachrichten bewusst zu sein, ist der Chat nach einer kurzen Registrierung für alle Interessierten zugänglich. Dennoch können digitale Chats durch eine „öffentliche Intimität“ (Pink et al. 2016: 110) charakterisiert werden, wie Sarah Pink et al. es im Kontext von Onlineforen beschreiben:

„Because of the narrow-cast, quasi-oral nature of online threads, participants may feel as if they are sharing a conversation with an intimate group of conversational partners. At the same time, forumers are aware that potentially anyone in the World Wide Web could be lurking in the shadows.“ (ebd.)

In *Stall Catchers* teilen einige Teilnehmer:innen im Chat sehr persönliche Informationen, zum Beispiel über Gesundheitsprobleme oder ihnen nahestehende Menschen, die sie (an Alzheimer) verloren haben. Aus diesem Grund ist ein reflektierter Umgang mit diesen sensiblen Daten notwendig, der über die Unterscheidung öffentlicher versus privater Daten hinaus-

geht. Gleichzeitig müssen digitale Chatdaten deshalb nicht grundsätzlich aus dem ethnografischen Datenmaterial ausgeschlossen werden. An dieser Stelle kann das vielschichtige Problem nicht in all seinen Dimensionen diskutiert werden. Dennoch möchte ich mein eigenes Vorgehen als Angebot zum Nachdenken offenlegen. Im geschilderten Fallbeispiel haben diese Überlegungen dazu geführt, dass ich die Chatdaten in erster Linie als indirekte Quelle verstehe und keine direkten Zitate aus dem Chat in Publikationen aufnehme. Dadurch war es nicht weniger nützlich, schließlich konnte ich so Einblicke in die Bedeutungen von *Stall Catchers* für Teilnehmer:innen erhalten und besser nachvollziehen, was während des Spiels passierte. Gleichzeitig vermied ich es, die Teilnehmer:innen, die in den Chatdaten auftauchen, mit den interviewten Teilnehmer:innen oder anderen Daten der gleichen Teilnehmer:innen zu verknüpfen. Außerdem nutzte ich Pseudonyme und veränderte teilweise Geschlechterzuordnungen, um Profile möglichst zu verschleiern und die Privatsphäre meiner Forschungspartner:innen zu schützen. Darüber hinaus können ethnografische Porträts (Bourdieu & Accardo 1998; Gutekunst & Rau 2017) erstellt werden, die aus einer idealtypischen Kombination von Identitäten und Perspektiven verschiedener Teilnehmer:innen bestehen.

Die Frage der Analyse großer Datenmengen wirft schließlich auch die Frage auf, inwiefern gerade die Quantität empirischer Daten ausschlaggebend für das Erkenntnisinteresse einer Forschung sein kann. Ist mehr immer erkenntnisreicher? In Forschungssettings wie dem meinigen kann der Einbezug des Chatdatensatzes einen analytischen und ethnografischen Mehrwert bringen. Diese Frage lässt sich jedoch nur für den ethnografischen Einzelfall beantworten und ist immer von der Forschungsfrage, den Feldzugängen und den Forschungsentscheidungen abhängig. Wichtig ist hierbei immer die Frage, welche Daten für die Beantwortung der eigenen Forschungsfrage sinnvoll sind (vgl. Bischoff & Oehme-Jüngling 2014). Schließlich hängt dies auch von den jeweiligen Chatdaten oder großen Datenmengen insgesamt ab. Denn während die einen Daten Dichte aufweisen können, können andere für die eigene Analyse oberflächlich wirken.

Sollten große Datenmengen in einer ethnografischen Forschung vorliegen und potenziell neue Erkenntnisse oder Zugänge erlauben, können viele der beschriebenen Grenzen durch ein überlegtes Vorgehen sowie die Kombination verschiedener Methoden im ethnografischen Forschen überwunden werden. So kann die qualitative Analyse großer Chatdatensmengen erkenntnisbringend in das eigene Vorgehen integriert werden.

Zusammenfassung: „Ein Gefühl für die Daten entwickeln“ – Zwischen Datenstrukturierung und -analyse

In diesem Beitrag habe ich ein Vorgehen präsentiert, das einen erkenntnisbringenden Einbezug großer Chatdaten in ethnografische Forschungen erlaubt, ohne dabei den Fokus von der qualitativen Analyse verlieren zu müssen. Zwar mögen beim hier präsentierten Vorgehen die Materialitäten, Geräusche und Atmosphären anders als jene sein, die Michaela Fenske in ihren Archiverfahrungen beschreibt (2006: 161), dennoch können digitale Daten eine Recherchelust entfachen, der mithilfe relationaler Datenbanken nachgegangen werden kann. Die explorative Auseinandersetzung mit digitalen Chatdaten ermöglicht es besonders am Anfang des Forschungsprozesses, ein Gefühl für die vorliegenden Daten zu entwickeln, Nähe zum Feld herzustellen und in dieses einzutauchen. Als besonders wichtig gilt dabei, die Vorgehensentscheidungen Schritt für Schritt festzuhalten. Dies ermöglicht nicht nur eine spätere

Rekonstruktion, sondern das kritische kulturwissenschaftliche Reflektieren des eigenen Vorgehens, das nicht nur zwischen Strukturierung und Analyse, sondern auch zwischen mathematisch-digitalen und ethnografischen Logiken abläuft.

Anmerkungen

Ich danke meinen Forschungspartner:innen für ihre Zeit und die wertvollen Einblicke, die ich gewinnen durfte. Die dem Kapitel zugrunde liegende Forschung ist Teil des Forschungsprojekts „Spielend in the Loop: Neue Mensch-Software Relationen in Human Computation Systemen und deren Auswirkungen auf Sphären des Alltags“, gefördert durch die Deutsche Forschungsgemeinschaft (DFG) – 464513114.

Literatur

- Barad, Karen (2015): *Verschänkungen*. Berlin: Merve Verlag.
- Bareither, Christoph (2020): *Affordanz*. In: Timo Heimerdinger & Markus Tauschek (Hgs.), *Kulturtheoretisch argumentieren. Ein Arbeitsbuch*. Münster, New York: Waxmann, 32–55.
- Bauernschmidt, Stefan (2014): *Kulturwissenschaftliche Inhaltsanalyse prozessgenerierter Daten*. In: Christine Bischoff, Karoline Oehme-Jüngling & Walter Leimgruber (Hgs.), *Methoden der Kulturanthropologie*. Bern: Haupt, 415–430.
- Beißwenger, Michael (2000): *Kommunikation in virtuellen Welten: Sprache, Text und Wirklichkeit. Eine Untersuchung zur Konzeptionalität von Kommunikationsvollzügen und zur textuellen Konstruktion von Welt in synchroner Internet-Kommunikation, exemplifiziert am Beispiel eines Webchats*. Stuttgart: Ibidem-Verlag.
- Bischoff, Christine & Karoline Oehme-Jüngling (2014): *Fragestellungen entwickeln*. In: Christine Bischoff, Karoline Oehme-Jüngling & Walter Leimgruber (Hgs.), *Methoden der Kulturanthropologie*. Bern: Haupt, 32–52.
- Bourdieu, Pierre & Alain Accardo (1997): *Das Elend der Welt. Zeugnisse und Diagnosen alltäglichen Leidens an der Gesellschaft*. Konstanz: UVK.
- Burkhardt, Marcus (2015): *Digitale Datenbanken: eine Medientheorie im Zeitalter von Big Data*. Bielefeld: Transcript. <https://www.transcript-verlag.de/shopMedia/openaccess/pdf/oa9783839430286.pdf>. Letzter Zugriff: 18.08.23.
- Charmaz, Kathy (2012): *The Power and Potential of Grounded Theory*. In: *Medical Sociology Online* 6/3, 2–15. <https://epicpeople.org/wp-content/uploads/2019/12/Charmaz-2012.pdf>. Letzter Zugriff 18.08.23.
- Charmaz, Kathy (2014): *Constructing grounded theory. Introducing qualitative methods*. Los Angeles u.a.: Sage.
- Fenske, Michaela (2006): *Mikro, Makro, Agency: Historische Anthropologie als kulturanthropologische Praxis*. In: *Zeitschrift für Volkskunde* 102/2, 151–177.
- Franken, Lina (2023): *Digitale Methoden für qualitative Forschung: computationelle Daten und Verfahren*. Münster, New York: Waxmann. <https://doi.org/10.36198/9783838559476>.
- Gibson, James J. (1982): *Wahrnehmung und Umwelt. Der ökologische Ansatz in der visuellen Wahrnehmung*. München: Urban & Schwarzenberg.
- Glaser, Barney G. & Anselm L. Strauss (1971) [1967]: *The Discovery of Grounded Theory. Strategies for Qualitative Research*. 4. Aufl. Chicago, New York: Aldine.
- Gutekunst, Miriam & Alex Rau (2017): *Das ethnographische Porträt. Ein Plädoyer für eine didaktische Auseinandersetzung mit dem Schreiben als Lernprozess*. In: Alex Rau & Irene

- Götz (Hgs.), Facetten des Alter(n)s. Ethnografische Porträts über Vulnerabilitäten und Kämpfe älterer Frauen. Münchner Ethnografische Schriften 25. München: Herbert Utz Verlag, 119–128.
- Knecht, Michi (2013): Nach Writing Culture, mit Actor-Network: Ethnografie/ Praxeografie in der Wissenschafts-, Medizin- und Technikforschung. In: Sabine Hess, Johannes Moser & Maria Schwertl (Hgs.), *Europäisch-ethnologisch Forschen. Neue Methoden und Konzepte*. Berlin: Reimer, 79–106.
- Koch, Gertraud & Lina Franken (2020): Filtern als digitales Verfahren in der wissenssoziologischen Diskursanalyse: Potenziale und Herausforderungen der Automatisierung im Kontext der Grounded Theory. In: Samuel Breidenbach, Peter Klimczak & Christer Petersen (Hgs.), *Soziale Medien. Interdisziplinäre Zugänge zur Onlinekommunikation*. Wiesbaden: Springer Vieweg, 121–38. https://doi.org/10.1007/978-3-658-30702-8_6.
- Hengartner, Thomas (2007): Volkskundliches Forschen im, mit dem und über das Internet. In: Silke Göttisch & Albrecht Lehmann (Hgs.), *Methoden der Volkskunde. Positionen, Quellen, Arbeitsweisen der Europäischen Ethnologie*. 2. Aufl. Berlin: Dietrich Reimer Verlag, 189–218.
- Hine, Christine (2006): Databases as Scientific Instruments and Their Role in the Ordering of Scientific Work. In: *Social Studies of Science* 36/2, 269–298. <https://doi.org/10.1177/0306312706054047>.
- Lemke, Matthias & Gregor Wiedemann (Hgs.) (2016): *Text Mining in den Sozialwissenschaften. Grundlagen und Anwendungen zwischen qualitativer und quantitativer Diskursanalyse*. Wiesbaden: Springer Fachmedien. <https://doi.org/10.1007/978-3-658-07224-7>.
- Lemke, Matthias (2017): Blended Reading. In: *Sozialwissenschaftliche Methodenberatung. Blog mit Beiträgen zu qualitativen sozialwissenschaftlichen Methoden*. <https://sozmethode.hypotheses.org/139>. Letzter Zugriff: 20.07.2023.
- Litovski, Ivan, Richard Maynard & James Aley (2010): *Inside the Symbian SQL: A Mobile Developer's Guide to SQLite*. Hoboken: Wiley.
- Manderscheid, Katharina (2019): Text Mining. In: Nina Baur & Jörg Blasius (Hgs.), *Handbuch Methoden der empirischen Sozialforschung*. Wiesbaden: Springer VS, 1103–1116. https://doi.org/10.1007/978-3-658-21308-4_79.
- Moretti, Franco (2016): *Distant Reading*. Konstanz: Konstanz University Press.
- Müller, Martin (o.J.): Scalable Reading. <https://sites.northwestern.edu/scalablereading/scalable-reading/>.
- Nam, Sang-Hui (2019): Qualitative Analyse von Chats und anderer usergenerierter Kommunikation. In: Nina Baur & Jörg Blasius (Hgs.), *Handbuch Methoden der empirischen Sozialforschung*. Wiesbaden: Springer VS, 1041–1051. https://doi.org/10.1007/978-3-658-21308-4_74.
- Passoth, Jan-Hendrik & Josef Wehner (2018): Listen, Daten, Algorithmen. Ordnungsformen des Digitalen. In: Thorben Mämecke, Jan-Hendrik Passoth, & Josef Wehner (Hgs.): *Bedeutende Daten*. Wiesbaden: Springer Fachmedien, 51–68. https://doi.org/10.1007/978-3-658-11781-8_4.
- Pink, Sarah, Heather A. Horst, John Postill, Larissa Hjorth, Tania Lewis & Jo Tacchi (Hgs.) (2016): *Digital ethnography: principles and practice*. Los Angeles: SAGE.
- Schirmer, Dominique, Nadine Sander & Andreas Wenninger (Hgs.) (2015): *Die qualitative Analyse internetbasierter Daten: methodische Herausforderungen und Potenziale von*

- Online-Medien. Soziologische Entdeckungen. Wiesbaden: Springer VS.
<https://doi.org/10.1007/978-3-658-06296-5>.
- Schönberger, Klaus (2015): Persistenz und Rekombination Digitale Kommunikation und soziokultureller Wandel. In: Zeitschrift für Volkskunde 111/ 2, 201–213.
https://www.waxmann.com/index.php?eID=download&id_artikel=ART101761&uid=frei.
Letzter Zugriff: 18.08.23.
- Schubert, Matthias (2007): Datenbanken. Theorie, Entwurf und Programmierung relationaler Datenbanken. Wiesbaden: Vieweg Verlag. <https://doi.org/10.1007/978-3-8351-9108-2>.
- Studer, Thomas (2019): Relationale Datenbanken: von den theoretischen Grundlagen zu Anwendungen mit PostgreSQL. Lehrbuch. Berlin, Heidelberg: Springer Vieweg.
<https://doi.org/10.1007/978-3-662-58976-2>.
- Ullrich, Carsten G. & Daniela Schiek (2015): Forumdiskussionen im Internet als reaktives Instrument der Datenerhebung. In: Dominique Schirmer, Nadine Sander & Andreas Wenninger (Hgs.), Die qualitative Analyse internetbasierter Daten. Methodische Herausforderungen und Potenziale von Online-Medien. Wiesbaden: Springer Fachmedien, 133–159. https://doi.org/10.1007/978-3-658-06296-5_5.
- Quamen, Harvey & Jon Bath (2016): Databases. In: Constance Crompton, Richard J. Lane & Ray Siemens (Hgs.), Doing Digital Humanities: Practice, Training, Research. London: Routledge, 145–162.
- Walz, Markus (2014): Prosopografie: Rekonstruktion und Analyse historischer Biografien. In: Christine Bischoff, Karoline Oehme-Jüngling & Walter Leimgruber (Hgs.), Methoden der Kulturanthropologie. Bern: Haupt, 195–209.
- Wolff, Stephan (2011): Textanalyse. In: Ruth Ayaß & Jörg Bergmann (Hgs.), Qualitative Methoden der Medienforschung. Mannheim: Verlag für Gesprächsforschung, 245–273.

Autor:inneninformation

Libuše Hannah Vepřek studierte Empirische Kulturwissenschaft sowie Informatik in Regensburg und München. Sie promoviert am Institut für Empirische Kulturwissenschaft und Europäische Ethnologie der LMU München und ist wissenschaftliche Mitarbeiterin im DFG-Projekt „Spielend in the Loop: Neue Mensch-Software Relationen in Human Computation Systemen und deren Auswirkungen auf Sphären des Alltags“. Zu ihren Forschungsinteressen zählen unter anderem die digitale und Technikanthropologie, Science and Technology Studies, Moralanthropologie, Critical Code Studies, Digitale Methoden und Erinnerungskulturen.