

HiER|2013

Griesbaum, Heuwing, Ruppenhofer, Werner (Hrsg.)

HiER 2013

Proceedings des 8. Hildesheimer Evaluierungs-
und Retrievalworkshop

Hildesheim, 25.–26. April 2013

J. Griesbaum, B. Heuwing, J. Ruppenhofer, K. Werner (Hrsg.):
HIER 2013 - Proceedings des 8. Hildesheimer Evaluierungs-
und Retrievalworkshop, Hildesheim 2013

© Institut für Informationswissenschaft und Sprachtechnologie,
Universität Hildesheim, 2013.

Was bedeutet 11.882 Bearbeitungen?

Die qualitative Inhaltsanalyse zur Untersuchung des Entstehungsprozesses von Wikipedia-Artikeln

Clemens Roth, Joachim Griesbaum, Ralph Kölle

Stiftung Universität Hildesheim

Marienburger Platz 22

31141 Hildesheim

c.roth@fastmail.fm, {griesbau | koelle}@uni-hildesheim.de

Zusammenfassung

Der Wikipedia-Artikel „Deutschland“ ist trotz 11.882 Bearbeitungen kein als exzellent oder lesenswert ausgezeichnete(r) Artikel. Man kann dementsprechend davon ausgehen, dass er im Sinne der Wikipedia-Qualitätskriterien nicht besonders hochwertig ist. Dieses Beispiel dient als Aufhänger für die Frage, ob und inwiefern Zusammenhänge zwischen der Anzahl und Art von Bearbeitungen und der Qualität von Beiträgen existieren. Was bedeutet eine Bearbeitung eines Wikipedia-Artikels hinsichtlich der Qualität des Beitrags und welche Aspekte sind bei der Untersuchung dieses Zusammenhangs wichtig? Um diese Frage zu adressieren, wurde eine explorative Studie in Form einer qualitativen Inhaltsanalyse von sechs Wikipedia-Artikeln durchgeführt. Der nachfolgende Artikel stellt den methodischen Ansatz dieser Studie sowie deren Ergebnisse dar.

Abstract

What does high edit count mean and which aspects are important when analyzing the evolution of Wikipedia articles? This paper discusses methods and approaches to investigate relations between the evolution of Wikipedia articles and their quality. Results of a first exploratory analysis which focuses on a qualitative in-depth categorization of an article's history are presented.

Einleitung

Die Wikipedia ist wohl die größte (Online-) Enzyklopädie (am 29.03.2013 weist die deutschsprachige Version über 1,5 Mio Einträge auf) und gehört zu den meistbesuchten Seiten im Internet (vgl. Alexa 2013). Die Artikel werden von freiwilligen, teils anonymen Benutzern geschrieben, können von jedem bearbeitet werden und sind frei verfügbar. (vgl. Wikipedia 2013a)

Die große Popularität und die freie Verfügbarkeit aller Inhalte stellen Gründe dafür dar, dass die Wikipedia in vielfältiger Hinsicht erforscht wird. Einen Schwerpunkt bildet dabei die Frage der Qualität der Inhalte. Neben Untersuchungen, die einen Schwerpunkt auf artefaktzustandsbezogene Attribute wie etwa Beitragslänge, Schreibstil, Objektivität oder Anzahl vorhandener Links legen, z.B. Lewandowski & Spree (2011), Yaari et al. (2011), existieren auch Ansätze, welche dabei den Entstehungsprozess der Beiträge fokussieren, z.B. Liu & Ram (2011), Kane (2011), Stein & Hess (2008). In diesen Bereich ordnet sich die hier dargestellte Untersuchung mit ein. Die Kernfrage ist, zu eruieren, welche Arten von Bearbeitungen an Wikipedia-Artikeln durchgeführt werden, wie häufig diese vorkommen und ob sich diesbezüglich Unterschiede zwischen Artikeln verschiedener Qualitätsstufen identifizieren lassen.

Der Wikipedia-Artikel „Deutschland“ eignet sich, um das Erkenntnisinteresse zu veranschaulichen. Dieser weist zwar mit 11.882 Bearbeitungen die höchste Anzahl an Bearbeitungen in der deutschen Wikipedia auf (Wikipedia 2013b: Stand: 27.1.2013), gleichzeitig ist er nicht mehr als lesenswert oder exzellent ausgezeichnet. Er wird also aktuell, zumindest wenn man die Auszeichnungsverfahren der Wikipedia selbst (Wikipedia 2013c) als Vergleichsmaßstab nimmt, trotz der besonders vielen Bearbeitungen nicht als besonders hochwertiger Artikel betrachtet.

Anhand einer qualitativen Inhaltsanalyse von sechs Wikipedia-Artikeln verschiedener Qualitätsstufen soll in diesem Beitrag ein konkreter Ansatz zur entstehungsprozessbezogenen Analyse vorgestellt und diskutiert werden.

Hierzu wird zunächst das Konzept der Versionshistorie in der Wikipedia skizziert. Anschließend werden methodische Grundlagen der Inhaltsanalyse angeführt. Auf dieser Basis wird ein kurzer State of the Art zu quantitativen und qualitativen Untersuchungen im Themenfeld gegeben. Schließlich werden das eigene Vorgehen und zentrale Ergebnisse der Untersuchung geschil-

dert. Abschließend wird die Untersuchung eingeordnet und ein Ausblick auf mögliche weitere Untersuchungen gegeben.

Das Konzept der Versionshistorie bei der Wikipedia

Eines der wichtigsten Charakteristika der Wikipedia ist die lückenlose Speicherung aller Änderungen an Artikeln in der Versionshistorie. Sie ermöglicht ein genaues Nachvollziehen aller Bearbeitungen und so des gesamten Entstehungsprozesses.

Nach jeder Bearbeitung eines Artikels, d.h. immer dann, wenn ein Benutzer auf Speichern klickt, wird eine neue Version inklusive verschiedener Metadaten, wie Autor, Datum und Kommentar, in der Versionshistorie angelegt (vgl. Wikipedia 2013d). Welche Änderung er tatsächlich vorgenommen hat, ist nicht ersichtlich. Die Quantität der Versionen erlaubt damit zunächst keine Rückschlüsse hinsichtlich der qualitativen Entwicklung eines Artikels. Auch bei Änderungen, die den Artikel nicht inhaltlich betreffen, werden neue Versionen erzeugt. Beispiele hierfür sind Vandalismus und Revertierungen, also das vorsätzliche Verfälschen und Beschädigen von Artikeln und das anschließende Wiederherstellen von vorhergehenden Versionen sowie Wikipedia-spezifische Aktivitäten, wie das Hinzufügen von Kategorien oder Links zu anderen Sprachversionen. Des Weiteren werden von einem Autor oft mehrere Versionen hintereinander erstellt. Diese können zu einem Bearbeitungsschritt zusammengefasst werden.

Zusammenfassend lässt sich konstatieren, dass sich aus der Versionsanzahl eines Artikels zunächst nur die Quantität der Editierungen ablesen lässt. Es können keine weiteren Aussagen über die Entstehung eines Artikels und seine Qualitätsentwicklung getroffen werden. Für derartige Einschätzungen ist die Durchführung von inhaltsanalytischen Untersuchungen unabdingbar.

Inhaltsanalyse

Nach Mayring ist die Inhaltsanalyse ein sehr flexibles Werkzeug um „[...] Material, das aus irgendeiner Art Kommunikation stammt“ (Mayring 2010:11), zu analysieren. Wobei sie „[...] nicht nur Inhalte der Kommunikation zum Gegenstand hat [...]“ (Mayring 2010:13), sondern auch als eine kategoriengeleitete Textanalyse gesehen werden kann.

Die Grundlage der Inhaltsanalyse ist ein Kategoriensystem, das auf ein bestimmtes Material, zum Beispiel Foren-Diskussionen oder auch Zeitungsartikel, angewendet wird. Dabei wird zwischen deduktiver und induktiver Kategorienbildung unterschieden. Deduktiv bedeutet, dass Kategorien a priori aus der Theorie oder aus dem Forschungsinteresse abgeleitet und begründet werden. Bei der induktiven Vorgehensweise werden die Kategorien direkt aus dem zu untersuchenden Material extrahiert. (vgl. Mayring 2010:49f.) Reinhoffer (2008:127) weist daraufhin, dass deduktive Ansätze zwar eine größtmögliche Systematik und ein regelgeleitetes Vorgehen bieten, aber manchmal nicht komplett offen und nachvollziehbar sind. Zudem ist es möglich, dass bestimmte Aspekte des Materials übersehen werden, da sie vorher nicht bekannt waren. Induktive Vorgehensweisen bieten hingegen eine sehr große Offenheit und die Möglichkeit, alle Varianzen des Materials im Kategoriensystem abzubilden.

Bei der Entwicklung eines Kategoriensystems muss zudem entschieden werden, ob die Texte quantitativ oder qualitativ kategorisiert werden sollen. Des Weiteren müssen klare Definitionen und Regeln für die einzelnen Kategorien festgelegt und möglichst Beispiele zur Illustration gefunden werden.

Inhaltsanalysen von Wikipedia-Artikeln

Um den tatsächlichen Entwicklungsprozess von Wikipedia-Artikeln im Sinne der Zusammenarbeit verschiedener Benutzer analysieren zu können, ist es sinnvoll, zunächst die Versionshistorie zu filtern. Dabei werden nicht relevante Versionen entfernt und aufeinanderfolgende Versionen eines Autors zu Bearbeitungsschritten zusammengefasst. Die Anzahl der so entfernten Versionen schwankt zwischen verschiedenen Artikeln sehr stark, da insbesondere

die Anzahl der als Vandalismus einzustufenden Versionen unterschiedlich ist und auch die Anzahl der durch Zwischenspeicherungen entstandenen Versionen stark von dem individuellen Verhalten der Autoren abhängt. In der vorliegenden Untersuchung liegt die Anzahl der entfernten Versionen zwischen 219 und 599 Versionen, das entspricht einem prozentualen Anteil zwischen 50,8% und 69% an allen Editierungen.

Anschließend können die Bearbeitungsschritte anhand des vorher definierten Kategoriensystems kategorisiert werden. Dazu bietet sich die Verwendung einer Diff-Funktion an. Sie visualisiert die Änderungen der Bearbeitungsschritte eines Artikels, indem die Unterschiede zweier aufeinanderfolgender Versionen textuell gekennzeichnet werden.

Auf diese Weise lassen sich zum Beispiel Aussagen über die Häufigkeit und Verteilung verschiedener Aktionen treffen (Roth 2012), kulturelle Unterschiede im Editierverhalten untersuchen (Pfeil et al. 2006) oder Rollen und Kollaborationsmuster definieren, die auf den Aktivitäten der Autoren basieren (Liu & Ram 2011). Die Schwierigkeit bei der Entwicklung von Kategorien und auch bei der späteren Codierung liegt darin, dass meist nicht bekannt ist, welche Intention einer Änderung zugrunde lag. Bei der induktiven Entwicklung des Kategoriensystems werden bestimmte Aktionen aus den Unterschieden zwischen zwei Text-Versionen abgeleitet und in Kategorien abstrahiert. Bei dem Kategorisieren bzw. Codieren handelt es sich schließlich um die Interpretation von Unterschieden zwischen zwei Text-Versionen auf Basis festgelegter Regeln. Die Unterschiede werden anhand des Kategoriensystems codiert und repräsentieren dann Aktionen von Autoren.

Qualitative oder quantitative Datenerhebung?

Insgesamt werden im Rahmen von Inhaltsanalysen bestimmte Sachverhalte quantifiziert. Bei der Konzeption einer Inhaltsanalyse ist abzuwägen, ob die Codierung von Menschen durchgeführt werden muss oder automatisiert erfolgen kann. Dies ist davon abhängig, wie bestimmte Sachverhalte abstrahiert werden und welche Regeln und Indikatoren zur Codierung herangezogen werden.

Die folgende Darstellung von Untersuchungen im Themenfeld verdeutlicht die Unterschiede beider Vorgehensweisen.

Übersicht zu inhaltsanalytischen Untersuchungen

Die nachfolgende Darstellung erhebt keinen Anspruch auf Vollständigkeit. Sie dient primär dazu, das Vorgehen der eigenen Untersuchung einzuordnen.

In qualitativen Studien zur Versionshistorie von Wikipedia-Artikeln werden die verschiedenen Aktionen der Benutzer von Menschen interpretiert bzw. codiert. Dadurch kann meist auf die Intention und somit auf die tatsächlich durchgeführte Aktion geschlossen werden.

Pfeil et al. (2006) zeigen in einer qualitativen Untersuchung mehrerer Sprachversionen eines Artikels, dass sich kulturelle Unterschiede im Editierverhalten von Wikipedia-Autoren beobachten lassen. Dabei setzen sie ein Kategoriensystem ein, das auch zwischen inhaltlichen und formalen Änderungen unterscheidet. Kane (2011) führt zunächst eine qualitative Untersuchung der Versionshistorie eines Artikels durch und identifiziert viele verschiedene Arten von Editierungen. Insbesondere das sogenannte Content Shaping, sinngemäß das Formatieren, Umstrukturieren und Überarbeiten von fremdem Inhalt ist von großer Bedeutung. Weitere qualitative Untersuchungen der Versionshistorien liefern Pentzold (2007), Kallass (2008) und Swarts (2009). Pentzold (2007) weist in einer Diskursanalyse von zwei Artikeln diskursive Regime im Sinne Michel Foucaults nach, während Kallass (2008) Aspekte des kollaborativen Schreibens anhand der Entwicklung von zwei Wikipedia-Artikeln untersucht. Swarts (2009) erforscht wiederum die kollaborative Konstruktion von Fakten und kommt zu dem Schluss, dass sie wie alle wissenschaftlichen Fakten in einem Prozess entstehen.

Diese qualitativen Kategoriensysteme haben den Vorteil, dass detailliert unterschieden werden kann, ob ein Bearbeitungsschritt den Inhalt verändert hat oder nicht. So kann der Entwicklungsprozess von Wikipedia-Artikeln sehr genau analysiert und nachvollzogen werden. Die Bearbeitungsschritte müssen von Menschen und unter Einbezug verschiedener Kontext-Informationen intellektuell interpretiert und codiert werden. Ein zentraler methodischer Aspekt ist dabei die Frage der Sicherstellung der intersubjektivität der intellektuellen Codierung. Weitergehend stellt die mangelnde Automatisierbarkeit einen zentralen Nachteil dar. Dadurch wird der Aufwand bei der Datenerhebung meist sehr hoch und eine große Stichprobe kann nur mit erheblichem Ressourcenaufwand untersucht werden.

Quantitative Untersuchungen

Bei quantitativen Untersuchungen von Wikipedia-Artikeln und deren Versionshistorie werden die Benutzeraktivitäten auf Basis formaler Regeln abstrahiert. Das heißt, die Änderungen werden mit einer Diff-Software automatisch erkannt und direkt kategorisiert.

Stein & Hess (2008) untersuchen die Entwicklung exzellenter Artikel anhand von Editierungsquantitäten. Sie verfolgen die Frage, ob die Anzahl der Autoren oder eher die Qualität einzelner Autoren bei der Entstehung exzellenter Artikel entscheidend ist. Dafür bewerten sie die Fähigkeit eines Autors anhand der Anzahl von ihm getätigter Editierungen exzellenter Artikel und der Anzahl exzellenter Artikel, die von ihm editiert wurden. Es gilt die Grundannahme: Je höher der jeweilige Anteil an seinen gesamten Editierungen bzw. insgesamt editierten Artikeln, als desto fähiger wird der Autor eingestuft. Es wird dabei nicht berücksichtigt, welche inhaltlichen Aktionen er vollzieht oder wie umfassend diese sind. Anhand des Anteils der fähigen Autoren, die an einem Artikel mitwirken, wird auf die Qualität des Artikels geschlossen. Die Autoren kommen zu dem Schluss, dass nicht die reine Anzahl der Autoren entscheidend für die Entstehung exzellenter und lesenswerter Artikel ist, sondern die Fähigkeit einzelner Autoren.

Ebenfalls einem quantitativen Ansatz folgend, analysieren Liu & Ram (2011) 1.600 Artikel der englischen Wikipedia. Dabei wird jede Änderung eines Benutzers automatisch erfasst und ausgezählt. Es wird zwischen den drei Aktivitäten Einfügen, Verändern und Löschen unterschieden. Diese Aktivitäten werden dann auf die Einheiten Satz, Link und Referenz bezogen. Zusätzlich gibt es die Aktion Revertierung. Es werden also alle Sätze von zwei Versionen formal miteinander verglichen und für jeden Satz abgeleitet, ob er neu hinzugefügt, verändert, gelöscht oder nicht verändert wurde. In verschiedenen weiteren Analysen werden anschließend Kollaborationsmuster bzw. -rollen, wie *All-round Contributor*, *Watchdog* und *Content Justifier*, entwickelt, um die Artikelqualität vorherzusagen. Die Autoren kommen zu dem Schluss, dass die von ihnen entwickelten Kollaborationsmuster eine gute Möglichkeit bieten, Artikelqualität vorherzusagen und dass die Artikelqualität von den Autoren und deren Rolle abhängt.

Der Vorteil quantitativer Inhaltsanalysen ist, dass sehr große und somit repräsentative Stichproben untersucht werden können. Gleichzeitig können durch die große Abstraktion die Aktionen nicht so detailliert erfasst werden wie in qualitativen Erhebungen. Es können keine inhaltlichen Aspekte be-

rücksichtigt werden. Die Aktion kann letztlich neuen Inhalt hinzufügen, eine inhaltliche Ergänzung, eine Korrektur auf inhaltlicher oder formaler Basis oder auch eine sprachlich-stilistische Änderung sein.

Untersuchungsdesign und Durchführung

Da das Ziel der hier dargestellten Untersuchung zunächst darin bestand, einen explorativen Einblick bezüglich der Anzahl und auch der Art der Bearbeitungen zu gewinnen, wurde ein qualitatives Untersuchungskonzept gewählt. Dabei werden im Rahmen der qualitativen Inhaltsanalyse die Versionshistorien von sechs Artikeln verschiedener Qualität zunächst wie oben beschrieben gefiltert, so dass nur noch inhaltlich relevante Bearbeitungsschritte enthalten sind. Anschließend werden diese anhand eines Kategoriensystems von mehreren Menschen codiert. Das Ziel der Analyse ist es, die kollaborative Entwicklung der Inhalte der Wikipedia-Artikel möglichst umfassend und detailliert zu beschreiben und zu prüfen, ob sich Unterschiede in der Entstehung von Artikeln verschiedener Qualitätsstufen identifizieren lassen. Ein besonderer Fokus liegt dabei auf der Unterscheidung zwischen inhaltlichen und nicht-inhaltlichen Änderungen.

Das Kategoriensystem wurde zunächst deduktiv aus der Theorie der Wissenskonstruktion in Wikis nach Cress & Kimmerle (2008), den Ausführungen von Schmalz (2007) zu Organisationsprinzipien in Wikis und der Untersuchung zu kulturellen Unterschieden im Editierverhalten von Wikipedia-Autoren von Pfeil et al. (2006) abgeleitet. Anschließend wurde es induktiv am Material getestet und weiterentwickelt. Es besteht aus drei Oberkategorien:

- A. Inhaltliche Änderungen
- B. Formale, nicht-inhaltliche Änderungen
- C. Sonstiges

Die Oberkategorie *Inhaltliche Änderungen* fasst alle Aktionen, die inhaltliche Aussagen bzw. den semantischen Gehalt eines Artikels verändern. Dazu gehört das Hinzufügen von neuem Inhalt, das Ergänzen und Spezifizieren

sowie das Korrigieren und Aktualisieren von bestehendem Inhalt sowie das gezielte, begründete Löschen von Inhalt.

Die zweite Oberkategorie, *Formale, nicht-inhaltliche Änderungen*, beinhaltet alle Aktionen, die Rechtschreibung, Sprache und Stil, Struktur sowie die Formatierung und das Layout betreffen. Entscheidend ist, dass inhaltliche Aspekte des Artikels nicht verändert werden.

Die dritte Oberkategorie *Sonstiges* enthält alle weiteren Aktionen, die den Artikel und seine Form nicht beeinflussen. Dazu gehört beispielsweise Vandalismus, das Wiederherstellen von Versionen oder das Hinzufügen von Kategorien oder Links zu anderen Sprachversionen.

Dieses Kategoriensystem ermöglicht es, die Häufigkeit und Verteilung der einzelnen Aktionen, die bei der Bearbeitung von Wikipedia-Artikeln auftreten, detailliert zu analysieren.

Für die Kategorisierung der einzelnen Bearbeitungsschritte wurde ein web-basiertes Tool entwickelt, das dem Codierer die Bearbeitungsschritte nacheinander anzeigt und es gestattet, die Kategorien zu jedem Schritt direkt über ein Formular einzutragen. Alle Eingaben werden in einer Datenbank gespeichert. Die nachfolgende Abbildung zeigt das Werkzeug in der Übersicht.



Abbildung 1: Das Web-Interface für die Kategorisierung.

Von den über 1,5 Million Artikeln der deutschsprachigen Wikipedia sind 2195 als exzellent (vgl. Wikipedia 2013e) und 3539 als lesenswert (vgl. Wikipedia 2013f) ausgezeichnet (Stand: 15.4.2013). Die Verwendung dieser Qualitätsstufen hat den Vorteil, dass bei ausgezeichneten Artikeln von einem tatsächlichen qualitativen und formalen Mindeststandard ausgegangen werden kann. Für die vorliegende Untersuchung wurde eine Stichprobe von sechs Wikipedia-Artikeln verschiedener Qualität gewählt (ohne Auszeichnung: Kobe Bryant, Curaçao; Lesenswert: Chaosforschung, Lee Harvey Oswald; Exzellent: Tempel von Abu Simbel, Gesundheitssystem der Vereinigten Staaten). Die Stichprobe besteht aus zwei exzellenten, zwei lesenswerten und zwei nicht-ausgezeichneten Artikeln, die zufällig ausgewählt wurden und jeweils mindestens 400 Versionen aufweisen. Sie wurden am 7.7.2012 heruntergeladen.

Die Versionshistorien der Artikel wurden gefiltert und anschließend mit Hilfe des Web-Interfaces codiert. Um die Intersubjektivität zu gewährleisten, wurde jeder Artikel von zwei Codierern, die vorab geschult wurden, kategorisiert. Die Codierungen der verschiedenen Codierer wurden anschließend anhand des Reliabilitätskoeffizienten von Holsti (vgl. Früh 2007:190) verglichen. Die Werte reichen von 0,7 (Curaçao) über 0,74 (Lee Harvey Oswald) und 0,75 (Gesundheitssystem der Vereinigten Staaten) bis hin zu 0,76 (Kobe Bryant, Chaosforschung und Tempel von Abu Simbel). Um die Kategorisierungen zu konsolidieren, wurde anschließend gemeinsam mit den Codierern eine eindeutige Einordnung für die unterschiedlichen Kategorisierungen festgelegt. Somit konnte sichergestellt werden, dass die Kategorien nachvollziehbar, verlässlich und reliabel sind. Nachfolgend werden die Ergebnisse der Untersuchung zusammengefasst. Eine ausführliche Darstellung des Untersuchungsdesigns und der Ergebnisse findet sich in Roth (2012).

Ergebnisse

Die Anzahl der Codierungen der einzelnen Artikel schwankt zwischen 266 und 533 Codierungen. Dabei überwiegen in den beiden nicht-ausgezeichneten Artikeln die Kategorien der Oberkategorie *Inhaltliche Änderungen* mit 48,3% und 48,1% gegenüber 38,5% und 30,1% der Oberkategorie *Formale, nicht-inhaltliche Änderungen*. In den vier ausgezeichneten

Artikeln zeigt sich ein umgekehrtes Bild. Hier überwiegen die Kategorien der Oberkategorie *Formale, nicht-inhaltliche Änderungen* mit durchschnittlich 46,1% gegenüber durchschnittlich 32,8% der Oberkategorie *Inhaltliche Änderungen*.

Tabelle 1: Übersicht der Codierungen ohne die Oberkategorie 'Sonstiges'.

| | Kobe Bryant | Curaçao | Chaosforschung | Lee Harvey Oswald | Tempel von Abu Simbel | Gesundheitssystem der Vereinigten Staaten |
|--|--------------|--------------|----------------|-------------------|-----------------------|---|
| Gesamt | 418 | 266 | 418 | 365 | 293 | 533 |
| Oberkategorie <i>Inhaltliche Änderungen</i> | | | | | | |
| Neuer, selbstständiger Inhalt | 35 8,4% | 26 9,8% | 23 5,5% | 31 8,5% | 28 9,6% | 42 7,9% |
| Ergänzung, Spezifizierung | 76 18,2% | 53 19,9% | 91 21,8% | 49 13,4% | 50 17,1% | 70 13,1% |
| Inhaltliche Korrektur, Aktualisierung | 71 17 % | 32 12% | 9 2,2% | 10 2,7% | 10 3,4% | 25 4,7% |
| Löschung | 20 4,8% | 17 6,4% | 24 5,7% | 22 6% | 12 4,1% | 31 5,8% |
| Oberkategorie Gesamt | 202 48,3% | 128 48,1% | 147 35,2% | 112 30,7% | 100 34,1% | 168 31,5% |
| Oberkategorie <i>Formale, nicht-inhaltliche Änderungen</i> | | | | | | |
| Sprachlich-stilistische Überarbeitung | 49 11,7% | 27 10,2% | 52 12,4% | 56 15,3% | 29 9,9% | 89 16,7% |
| Tippfehler, Rechtschreibung, Interpunktion, Grammatik | 60 14,4% | 32 12,0% | 86 20,6% | 76 20,8 | 24 8,2% | 102 19,1 |
| Struktur | 8 1,9% | 2 0,8% | 12 2,9% | 11 3,0% | 12 4,1% | 28 5,3% |
| Formatierung, Layout | 44 10,5% | 19 7,1% | 35 8,4% | 31 8,5% | 46 15,7% | 71 13,3% |
| Oberkategorie Gesamt | 161 38,5% | 80 30,1% | 185 44,3% | 174 47,7% | 111 37,9% | 290 54,4% |

Die Verteilungen der einzelnen Kategorien der Oberkategorie *Inhaltliche Änderungen* sind in den meisten Fällen relativ ähnlich. Die Kategorien *Neuer, selbstständiger Inhalt*, *Ergänzung*, *Spezifizierung* und *Löschung* weisen anteilige Häufigkeiten von durchschnittlich 8,3% (*Neuer, selbstständiger Inhalt*), 17,2% (*Ergänzung, Spezifizierung*) und 5,5% (*Löschung*) auf. Lediglich die Kategorie *Inhaltliche Korrektur, Aktualisierung* weist stark schwankende anteilige Werte von 2,2% bis 17% auf.

In der Oberkategorie *Formale, nicht-inhaltliche Änderungen* sind hingegen größere Unterschiede zu beobachten. Lediglich die Kategorie *Sprachlich-stilistische Überarbeitung* ist ausgeglichen und weist durchschnittlich 12,7% auf. Die weiteren Kategorien schwanken zwischen 8,2% und 20,8% (Kategorie *Tippfehler, Rechtschreibung, Interpunktion, Grammatik*), 0,8% und 5,3% (Kategorie *Struktur*) sowie 7,1% und 15,7% (Kategorie *Formatierung, Layout*).

Aus diesen unterschiedlichen Verteilungen lassen sich die folgenden in Verbindung zur Artikelqualität setzen. Drei der vier ausgezeichneten Artikel weisen in der Kategorie *Tippfehler, Rechtschreibung, Interpunktion, Grammatik* mit 19,1%, 20,6% und 20,8% einen höheren Wert als die nicht-ausgezeichneten Artikel mit 12% und 14,4% auf.

Auch in der Kategorie *Struktur* zeigt sich eine steigende anteilige Häufigkeit von 1,9% und 0,8% in den nicht-ausgezeichneten Artikeln über 2,9% und 3% in den Lesenswerten bis zu 4,1% und 5,3% in den exzellenten Artikeln.

Die anteiligen Häufigkeiten der Kategorie *Formatierung, Layout* sind insbesondere in den beiden exzellenten Artikeln mit 15,7% und 13,3% recht hoch. Die Artikel der Qualitätsstufe Lesenswert liegen bei 8,4% und 8,5% und die nicht-ausgezeichneten Artikel bei 7,5% und 10,5%.

Die Ergebnisse sind zwar nicht verallgemeinerbar, weisen aber darauf hin, dass eher ein hoher Anteil an nicht-inhaltlichen Modifikationen ein Indikator für als qualitativ hochwertig gekennzeichnete Artikel darstellen könnte. Dies geht konform mit den Ergebnissen von Kane (2011) und stellt ein eher kontaintuitives Resultat dar. Das Ergebnis ist wohl auch von der Artikelauswahl der Untersuchung abhängig, dennoch ist es für die weitere Forschung nicht uninteressant, da sich gerade Modifikationen hinsichtlich Rechtschreibung, Struktur, Formatierung und Layout automatisch analysieren lassen.

Des Weiteren lässt sich beobachten, dass insbesondere die ausgezeichneten Artikel eine höhere Autorenaktivität aufweisen. Das heißt, dass einzelne Autoren einen Artikel mehrmals bearbeitet haben. In den nicht ausgezeichneten

ten Artikeln nehmen die Autoren hingegen tendenziell nur sehr wenige Änderungen vor. Dies entspricht den Ergebnissen von Stein & Hess (2008) bzw. Liu & Ram (2011), dass eher einzelne Autoren ausschlaggebend für die Qualität sind.

Fazit und Ausblick

Die Ausführungen haben gezeigt, dass es nicht sinnvoll ist, alleine aus der Versionsanzahl von Wikipedia-Artikeln belastbare Aussagen über die qualitative Entwicklung von Wikipedia-Artikeln zu treffen. 11.882 Bearbeitungen bedeuten nur, dass der Artikel Deutschland 11.882 Mal bearbeitet wurde. Welche Änderungen vorgenommen wurden, ist nicht ersichtlich. Um die Entstehung von Wikipedia-Artikeln zu untersuchen, müssen demnach detailliertere Analysen herangezogen werden. Dabei muss genau geprüft werden, welche Teile des Materials interessant sind und mit welcher Methodik stichhaltige Daten erhoben werden können.

Die Inhaltsanalyse ist ein Werkzeug, das sich aufgrund seiner Flexibilität gut eignet, um sowohl qualitative als auch quantitative Analysen der Versionshistorie von Wikipedia-Artikeln durchzuführen. Es ist jedoch von besonderer Bedeutung, dass im Rahmen der Konzeption der Forschungsarbeit genau geplant wird, wie die Änderungen erhoben bzw. codiert werden und wie abstrakt das Kategoriensystem gestaltet wird.

Für die vorliegende Untersuchung lässt sich diesbezüglich folgende Einschätzung vornehmen. Durch die zweifache Codierung durch verschiedene Personen und die anschließende Angleichung ist davon auszugehen, dass die erhobenen Daten im Sinne der Intercoderreliabilität reliabel sind. Aussagen über die Validität der Ergebnisse lassen sich hingegen nur schwer treffen. Legt man jedoch das Außenkriterium an, also einen Vergleich mit den Ergebnissen ähnlicher Studien, kann auch die Validität angenommen werden. Insbesondere Liu & Ram (2011) und Stein & Hess (2008) kommen zu vergleichbaren Ergebnissen.

Wie verschiedene andere Forschungsarbeiten bedient sich die vorliegende Arbeit qualitativer Methoden zur Datenerhebung. Durch den explorativen Charakter der Untersuchung und der tentativen Art der Ergebnisse ist sie teilweise als bestätigend, z.B. in Bezug auf die Unterscheidung zwischen

paralleler und nicht-paralleler Entwicklung von Artikel und Artikelgegenstand von Kallass (2008), aber vor allem als Vorarbeit, sozusagen als Anstoß für verschiedene weitere Forschungsansätze zu sehen.

Es konnten beispielsweise mehrere Unterschiede in der Entwicklung von Artikeln verschiedener Qualitätsstufen identifiziert werden. Ein möglicher weiterer Forschungsansatz wäre demnach zu prüfen, ob sich die Ergebnisse mit repräsentativen Stichproben reproduzieren lassen und somit als Indikator für eine automatische Qualitätserkennung verwendet werden können.

Zusätzlich wäre es denkbar, die erhobenen Daten noch tiefergehend zu analysieren. Ein möglicher Ansatz wäre beispielsweise Frage, ob sich Bearbeitungen verschiedener Kategorien auf einzelne Autoren konzentrieren und ob sich so differenzierte Rollen identifizieren lassen. Ein weiterer Ansatz könnte die zeitliche Entwicklung von Wikipedia-Artikeln sein und die Frage, ob beispielsweise Bearbeitungsphasen auftreten oder bestimmte Bearbeitungen auf andere folgen. Zudem wäre in diesem Kontext interessant, den Zeitpunkt der Auszeichnung als exzellent oder lesenswerter Artikel einzubeziehen und beispielsweise zu untersuchen, wie sich Artikel vorher und nachher entwickeln.

Um zu prüfen, ob es sich die Ergebnisse auf andere Kontexte als Wikipedia übertragen lassen, wäre es interessant, die Untersuchung anhand anderer Wikis durchzuführen.

Aus pragmatischer Sicht könnte die Erkenntnis, dass qualitativ hochwertige Artikel offenbar von einzelnen Autoren häufiger bearbeitet wurden, dazu verwendet werden, die Qualität von Wikipedia-Artikeln zu steigern, indem die Autoren aufgerufen werden, einzelne Artikel häufiger zu bearbeiten. Ob eine solche Initiative funktioniert, bleibt natürlich offen.

Des Weiteren könnte das entwickelte Kategorisierungs-Tool für das Crowdsourcing von Kategorisierungsarbeiten verwendet werden. Auf diese Weise könnten auch qualitative Inhaltsanalysen repräsentative Stichproben erreichen und beispielsweise die Ergebnisse der vorliegenden Untersuchung quantitativ überprüft werden.

References/Literaturverzeichnis

- Alexa (2013): Top Sites. The top 500 sites on the web. Retrieved March 31, 2013 from <http://www.alexacom/topsites>
- Cress, Ulrike; Kimmerle, Joachim (2008): A systemic and cognitive view on collaborative knowledge building with wikis. In: *International Journal of Computer-Supported Collaborative Learning* 3 (2), S. 105–122.
- Früh, Werner (2007): *Inhaltsanalyse*. 6., überarbeitete Auflage. Konstanz: UVK Verlagsgesellschaft mbH.
- Kallass, Kerstin (2008): Artikelentstehung in der Wikipedia. Zu Textkonstitutionsmustern und Schreibprozessen bei Wikipedia-Artikeln. In: *Berichte des DFG-Forschungsprojekts "Netzwerkkommunikation im Internet"*, Nr. 08 (3). Retrieved March 31, 2013 from <http://www.netzwerke-im-internet.de/fileadmin/downloads/Bericht2008-3-Wikipedia.pdf>
- Kane, Gerald C. (2011): A multimethod study of information quality in wiki collaboration. In: *ACM Transactions on Management Information Systems* 2 (1).
- Lewandowski, Dirk & Spree, Ulrike (2011): Ranking of Wikipedia articles in search engines revisited: Fair ranking for reasonable quality? In: *Journal of the American Society for Information Science*, 62(1):117-132.
- Liu, Jun; Ram, Sudha (2011): Who does what: Collaboration patterns in the wikipedia and their impact on article quality. *ACM Transactions on Management Information Systems* 2 (2).
- Mayring, Philipp (2010): *Qualitative Inhaltsanalyse. Grundlagen und Techniken*. 11., aktualisierte und überarbeitete Auflage. Weinheim und Basel: Beltz Verlag.
- Pentzold, Christian (2007): *Wikipedia. Diskussionsraum und Informationsspeicher im neuen Netz*. München: Verlag Reinhard Fischer (Internet research, 28).
- Pfeil, Ulrike; Zaphiris, Panayiotis; Ang, Chee Siang (2006): Cultural Differences in Collaborative Authoring of Wikipedia. *Journal of Computer-Mediated Communication* 12 (1), S. 88–113.
- Reinhoffer, Bernd (2008): Lehrkräfte geben Auskunft über ihren Unterricht. Ein systematisierender Vorschlag zur deduktiven und induktiven Kategorienbildung in der Unterrichtsforschung. In: Philipp Mayring und Michaela Gläser-Zikuda (Hg.): *Die Praxis der Qualitativen Inhaltsanalyse*. 2. Auflage. Weinheim und Basel: Beltz Verlag, S. 123–141.
- Roth, Clemens (2012): *Kollaborative Wissenskonstruktion am Beispiel von Wikipedia*. Magisterarbeit Institut für Informationswissenschaft und Sprachtechnologie, Universität Hildesheim. (Veröffentlichung auf HilDok, dem Publikationsserver

der Universität Hildesheim (<http://opus.bsz-bw.de/ubhi/index.php?la=de>), wird angestrebt.)

Schmalz, Jan Sebastian (2007): Zwischen Kooperation und Kollaboration, zwischen Hierarchie und Heterarchie. Organisationsprinzipien und Strukturen von Wikis. In: *kommunikation@gesellschaft* 8. Retrieved March 31, 2013 from http://www.soz.uni-frankfurt.de/K.G/B5_2007_Schmalz.pdf

Stein, Klaus; Hess, Monica (2008): Viele Autoren, gute Autoren? Eine Untersuchung ausgezeichneter Artikel in der deutschen Wikipedia. In: Paul Alpar und Steffen Blaschke (Hg.): *Web 2.0 - Eine empirische Bestandsaufnahme*. Wiesbaden: Vieweg+Teubner, S. 108–129.

Swarts, Jason (2009): The collaborative construction of "fact" on Wikipedia. Proceedings of the 27th ACM international conference on Design of communication. New York, NY, USA: ACM (SIGDOC '09), S. 281- 288.

Wikipedia, Die freie Enzyklopädie (Hg.) (2013a): Wikipedia:Über Wikipedia. Retrieved March 31, 2013 from http://de.wikipedia.org/w/index.php?title=Wikipedia:%C3%9Cber_Wikipedia&oldid=116077822.

Wikipedia, Die freie Enzyklopädie (Hg.) (2013b): Seiten mit den meisten Versionen. Retrieved March 31, 2013 from http://de.wikipedia.org/wiki/Spezial:Meistbearbeitete_Seiten

Wikipedia, Die freie Enzyklopädie (Hg.) (2013c): Wikipedia: Bewertungen. Retrieved March 31, 2013 from <http://de.wikipedia.org/w/index.php?title=Wikipedia:Bewertungen&oldid=115305426>

Wikipedia, Die freie Enzyklopädie (Hg.) (2013d): Hilfe:Versionen. Retrieved March 31, 2013 from

<https://de.wikipedia.org/w/index.php?title=Hilfe:Versionen&oldid=114258827>

Wikipedia, Die freie Enzyklopädie (Hg.) (2013e): Wikipedia:Exzellente Artikel. Retrieved April 15, 2013 from

http://de.wikipedia.org/w/index.php?title=Wikipedia:Exzellente_Artikel&oldid=116699124

Wikipedia, Die freie Enzyklopädie (Hg.) (2013f): Wikipedia: Lesenswerte Artikel. Retrieved April 15, 2013 from

http://de.wikipedia.org/w/index.php?title=Wikipedia:Lesenswerte_Artikel&oldid=116938655

Yaari, Eti; Baruchson-Arbib, Shifra & Bar-Ilan, Judit (2011): Information quality assessment of community generated content: A user study of Wikipedia. In: Journal of Information Science, 37(5):487-498.