

# Qualität der *Wikipedia*<sup>1</sup>

## Eine vergleichende Studie

*Rainer Hammwöhner, Karl-Peter Fuchs, Markus Kattenbeck,  
Christian Sax*

Informationswissenschaft  
Philosophische Fakultät IV  
Sprach- und Literaturwissenschaften  
Universität Regensburg

### Zusammenfassung

Die Qualität steht von Beginn an im Vordergrund der Debatte um die *Wikipedia*. In diesem Beitrag werden zunächst Studien vorgestellt, welche die Qualität der *Wikipedia* zum Gegenstand haben, sodann werden Design und Ergebnisse einer Studie präsentiert, welche eine Kollektion von Qualitätsparametern im Vergleich zwischen der deutschen *Wikipedia* und dem *Großen Brockhaus* evaluiert.

## 1. Einleitung

Prinzip und Erfolg der *Wikipedia*<sup>2</sup> sind in mancherlei Hinsicht bemerkenswert. Zunächst fällt vermutlich der rein quantitative Aspekt ins Auge: Die englische *Wikipedia* [Wikipedia 2007] hat mittlerweile (Stand 6.1.2007) mehr als 1.5 Millionen Artikel, weitere 11 in anderen Sprachen verfasste *Wikipedias* verfügen über mehr als einhunderttausend Beiträge. Beeindruckend ist auch das Wachstum der *Wikipedias*, das exponentielle Wachstumsphasen einschließt [Voss 2005, S. 27 ff.]. Der eigentliche Grund für die Aufmerksamkeit, welche die *Wikipedia* genießt, liegt jedoch in dem augenscheinlichen Konflikt zwischen dem Anspruch der *Wikipedia*, enzyklopädisches Wissen ohne Einschränkung in Qualität und Reichweite bereitzustellen, auf der einen Seite und dem Prinzip,

---

<sup>1</sup> Dieser Text ist erschienen in: Obwald, A.; Stempfhuber, M.; Wolff, C.: Open Innovation. Neue Perspektiven im Kontext von Information und Wissen. Proc. des 10. Int. Symposiums Informationswissenschaft. UVK, Konstanz, 2007, S. 77-90.



This text is published under the following Creative Commons Licence: Attribution-NonCommercial-NoDerivs 2.0 Germany (<http://creativecommons.org/licenses/by-nc-nd/2.0/de/>).

<sup>2</sup> Einen guten Überblick über die Prinzipien der Wikipedia bietet [Danowski/Voss 2005].

dieses Wissen von einer offenen, z.T. anonymen Teilnehmergruppe zusammen stellen zu lassen, auf der anderen. Auf diese Herausforderung für den traditionellen Wissens- und Wissenschaftsbetrieb mit seinen Verfahren der Qualitätssicherung erfolgten sowohl vorsichtig affirmative [Kuhlen 2005] als auch polemisch ablehnende Reaktionen [Lanier 2006], wobei letztere auch durch schwerwiegende Qualitätsprobleme in der *Wikipedia* motiviert wurden. Große Beachtung wurde einer Studie über die Qualität der *Wikipedia* zuteil, die im Auftrag von *Nature* erstellt wurde [Giles 2005]. Der *Wikipedia* wurde dort eine Qualität zugesprochen, welche jener der *Encyclopedia Britannica* gleich zu setzen sei. Diese Studie wurde von Seiten der *Britannica* angegriffen, von *Nature* aber verteidigt und mit Materialien untermauert<sup>3</sup>. Weniger als das Ergebnis dieser Debatte ist ihr Verlauf von Interesse. Es wurde deutlich, dass ein Konsens über eine angemessene Methodik zur Evaluierung der *Wikipedia* – wie auch ihrer Konkurrenzprodukte – erst noch zu erzielen ist. Hier ist ein erster Gegenstand für eine Erforschung der *Wikipedia* gegeben. Es überrascht also nicht, dass Fragen zur Qualität einen nennenswerten Umfang der bisher noch nicht sehr umfangreichen, aber durchaus qualitätsvollen Forschung zur *Wikipedia*<sup>4</sup> einnehmen.

## 2. Studien zur Qualität der *Wikipedia*

Zunächst sollen einige Studien vorgestellt werden, die sich mit Aspekten der Qualität der *Wikipedia* befassen. Diese werden jeweils hinsichtlich ihrer Methodik und Reichweite untersucht. Auf der Grundlage dieser Überlegungen wurde eine neue Studie entworfen, deren Ergebnisse in diesem Beitrag präsentiert werden.

### 2.1 Die Nature-Studie [Giles 2005]

Im Folgenden soll der Ablauf der Studie kurz skizziert werden. Zunächst wurden 50 Artikel ausgewählt, die ein breites naturwissenschaftliches Themenfeld abdecken sollten. Die Darstellungen aus der *Wikipedia* und der Online-Ausgabe der *Encyclopedia Britannica* sollten sich ungefähr im Umfang entsprechen, so dass letztendlich nur 42 Artikel zur Evaluation verblieben. Die Themen waren so ausgewählt, dass in der Redaktion Kompetenz vorhanden war. Zu jedem der Themen wurde ein Experte ausgewählt, der jeweils ein Paar aus einem *Wikipedia*- und einem *Britannica*-Artikel zu bewerten hatte, ohne die Zuordnung zu kennen. Die Gutachter sollten in den Artikeln drei Arten von Fehlern identifizieren: sachliche Fehler, kritische Auslassungen, irreführende

---

<sup>3</sup> Verweise auf diese Materialien finden sich in der Online-Fassung von [Giles 2005].

<sup>4</sup> Im Zusammenhang mit der deutschen *Wikipedia* wurde für *Wikipedia*-Forschung der Terminus *Wikipedistik* geprägt. Zusammenstellungen von Publikationen über die *Wikipedia* finden sich unter [Wikipedia 2007a, 2007b].

Formulierungen. Die Gutachten wurden dann vom Redaktionsteam gesichtet, die Fehlerangaben wurden gewichtet und summiert. Es stellte sich heraus, dass die beiden Enzyklopädien keine Unterschiede hinsichtlich gravierender Fehler aufwiesen (je 4), in Artikeln der *Wikipedia* aber deutlich mehr „kleine“ Fehler zu beobachten waren (162/123). Besagte Studie wurde durch die Redaktion der *Britannica* heftig kritisiert, da Ungereimtheiten hinsichtlich der Textgrundlage und der Bewertung von Fehlern bestünden. Die Kritik wurde von *Nature* zurückgewiesen.

Unabhängig von dieser Kritik ist jedoch zu beachten, dass die Reichweite der Studie sehr begrenzt ist. Folgende Einschränkungen sind zu bemerken:

- Die Evaluation evaluiert nur die Korrektheit von Lexikoneinträgen. Andere Qualitätsaspekte wie Lesbarkeit und Verständlichkeit, Konsistenz in der Gliederung, thematische Abdeckung etc. blieben unberücksichtigt. Die Vermutung, dass sich die Qualität redaktioneller Arbeit gerade bei diesen Qualitätsaspekten zeigen kann, ist nahe liegend.
- Die ausgewählten Artikel repräsentieren nicht die zu untersuchenden Grundgesamtheiten (Artikel der *Wikipedia* bzw. *Britannica*). Es ist davon auszugehen, dass naturwissenschaftliche Fragestellungen andere *Wikipedia*-Autoren anziehen als etwa Themen der Allgemeinbildung. Die Studie besitzt also Aussagekraft allenfalls für naturwissenschaftliche Themen.

## 2.2 Studie von Christian Schlieker [Schlieker 05]

In seiner Diplomarbeit befasst sich Herr Schlieker mit elektronischen Enzyklopädien. Neben anderen Fragestellungen, die hier nicht weiter erörtert werden sollen, nimmt eine komparative Studie bezüglich der Qualitätsparameter Abdeckung und informationelle Absicherung breiten Raum ein. Diese Studie ist vor allem deshalb von Interesse, weil sie methodisch und inhaltlich gut vergleichbar mit der in diesem Artikel neu vorgestellten Studie ist.

Untersucht werden die deutsche *Wikipedia*, *Encarta* und eine Druckfassung der Enzyklopädie *Der Große Brockhaus* (20. Auflage). Aus jeder der Enzyklopädien wird eine Zufallsstichprobe von 15 Artikeln gezogen [Schlieker 2005, S 43 ff]. Dabei werden nur Vollartikel und keine Begriffsklärungen o.Ä. berücksichtigt. Aus der *Wikipedia* wird die Stichprobe mit Hilfe der Zufallsfunktion gezogen, aus der *Encarta* mangels besserer Alternative durch „zufälliges“ Scrollen im alphabetischen Index bestimmt und aus der *Brockhaus*-Enzyklopädie durch ein mehrstufiges Zufallsverfahren (random walk), das zunächst den Band, dann Spalte, Seite und Lemma auswählt. Durch das Auswahlverfahren dürften aus dem *Brockhaus* mit leichter Präferenz längere Artikel ausgewählt werden, während bei der *Encarta* keine gleichmäßige Berücksichtigung des Alphabets zu erwarten ist. Kritischer für die Aussagekraft der Untersuchung dürfte jedoch der geringe Stichprobenumfang sein.

Die Stichproben benutzt Herr Schlieker [Schlieker 2005, S. 51 ff] zunächst, um die thematische Abdeckung der Enzyklopädien zu untersuchen:

- Er überprüft in welchem Maße die einzelnen Enzyklopädien die Gesamtmenge von 45 Lemmata abdecken. Es ergibt sich die Rangfolge *Brockhaus - Wikipedia – Encarta* (33, 29, 25 von 45).
- Es wird versucht, die 45 Lemmata zu kategorisieren und dann aus der jeweiligen Abdeckung auf thematisch Profile der Enzyklopädien zu schließen. *Brockhaus* wird ein geringerer Aktualitätsbezug unterstellt als der *Wikipedia*, während bei der *Encarta* eine Bevorzugung amerikanischer Fragestellungen beobachtet wird.

Weiterhin wird die mediale Aufbereitung der Artikel untersucht. Die Artikel der *Encarta* waren am umfassendsten mit Medienobjekten (Bilder, Audio, Video) ausgestattet, während die *Wikipedia* noch hinter dem *Brockhaus* zurückblieb. Erwartungsgemäß lag die *Wikipedia* hinsichtlich des internen Vernetzungsgrades vorne, gefolgt von *Encarta* und *Brockhaus*. Eine Absicherung der Information durch externe Literaturangaben war am häufigsten beim *Brockhaus* zu finden, gefolgt von *Encarta* und *Wikipedia*.

### **2.3 Die Studien von Stvilia, Twidale, Gasser und Smith [Stvilia et al. 2005, 2005a]**

Die vorgenannten Studien gingen jeweils von einem impliziten Qualitätsmodell aus, welches die Grundlage für eine vergleichende Bewertung der *Wikipedia* bildete. Die Studien von Stvilia, Twidale, Gasser und Smith entwickeln zunächst ein Qualitätsmodell für die *Wikipedia* [Stvilia et al 2005]. Grundlage dieses Qualitätsmodells ist die in der (englischen) *Wikipedia* vorgenommene Hervorhebung von „Featured Articles“. Grundlage sind zunächst die Qualitätskriterien der *Wikipedia*: „Comprehensive, Accurate, Verifyable, Stable, Well-written, Uncontroversial, Compliance, Appropriate Images with acceptable copyright status, Appropriate Style and Focus“. Diesem Kriterienkatalog stellen sie einen von [Crawford 2001] vorgeschlagenen gegenüber, der sich auf traditionelle Print-Enzyklopädien bezieht: „Scope, Format, Uniqueness, Authority, Accuracy, Currency, Accessibility“. Auch einen eigenen, umfangreicheren Kriterienkatalog, der hier nicht dargestellt werden soll, setzen sie zu den beiden anderen in Beziehung. Damit ist eine Grundlage für die Systematisierung von Qualitätsstudien sowie den Vergleich verschiedener Qualitätsmodelle gegeben. Im empirischen Teil der Studie stellen die Autoren sodann den Zusammenhang zwischen abstrakten Qualitätsmodellen und der Praxis der *Wikipedia* her. Dies erfolgt aufgrund einer inhaltlichen Analyse der den Lemmata zugeordneten Diskussionsseiten. Dazu wurden aus der *Wikipedia* und der Menge der Featured Articles jeweils 30 Artikel ausgewählt, die über einen hinlänglich umfangreichen Diskussionsteil verfügten. Die in der Diskussion auftretenden Argumente wurden kategorisiert und den Qualitätsdimensionen der *Wikipedia* zugeordnet. Somit wurde eine

Einschätzung der Bedeutung der einzelnen Qualitätsaspekte für die Einordnung als Featured Article möglich. Zusätzlich wurde versucht, Textparameter zu isolieren, welche als Symptome für eine hohe Qualität der Artikel anzusehen sind (Umfang und Struktur der Diskussion, Umfang und Überarbeitungsfrequenz der Artikel, Anzahl der Autoren etc.). Dieser quantitative Ansatz wurde in einer weiteren Studie vertieft [Stvilia et al 2005a], so dass ein automatischer Qualitäts-Klassifikator entwickelt werden konnte.

## **2.4 Weitere Studien mit Qualitätsbezug**

Die Studie von Andrew Lih [Lih 2004] befasst sich mit der Verlässlichkeit der *Wikipedia* als journalistische Quelle. Er entwickelt eine einfache Metrik, welche die Qualität von Artikeln erfassen soll. Sie beruht auf der Anzahl von Änderungen, die ein Artikel durchlaufen hat und der Anzahl der beteiligten Autoren. Während eine zunehmende Zahl der Bearbeitungen für eine hohe inhaltliche Konsolidierung eines Beitrags spräche, erhöhe eine große Anzahl von Autoren die Wahrscheinlichkeit von Inkonsistenzen und Argumentationsbrüchen. Die Metrik wurde an einem Testkorpus kalibriert, um dann an Artikeln, die von der Presse zitiert wurden, getestet zu werden. Diese Studie initiierte den Einzug quantitativer Methoden in die Analyse der *Wikipedia*.

Emigh und Herring [Emigh und Herring 2004] vergleichen zwei elektronische Enzyklopädien (*Wikipedia* und *Everything2*) hinsichtlich des verwendeten Sprachregisters. Sie beobachten in *Everything2* eine vergleichsweise freie Sprache, während die *Wikipedia* zunehmend eine eher förmliche Sprache verwende. Wenngleich diese Studie keinen unmittelbaren Bezug zur Frage der Qualität der *Wikipedia* hat, ist doch von Interesse, dass der Redaktionsprozess eine Vereinheitlichung der Sprache zur Folge hat. Dieser Effekt ist als relevant für Qualitätskriterien wie Lesbarkeit und sprachliche Konsistenz anzusehen.

## **2.5 Zusammenfassende Würdigung**

Bisher sind noch vergleichsweise wenige Studien zur Qualität der *Wikipedia* erschienen. Diese sind zudem methodisch nicht unangreifbar. Auffällig ist die Diskrepanz zwischen den aus Studien resultierenden Qualitätsurteilen und den zum Teil witzigen Pressemeldungen über aktuelle, zum Teil krasse Fehlleistungen der *Wikipedia* [z.B. Rühle 2006]. Dies mag z.T. an methodischen Problemen der Studien liegen, hat seinen Hauptgrund aber vermutlich darin, dass diese Studien ihren Blick auf die mittlere Qualität einer Grundgesamtheit richtet, die Presse aber immer zu Recht auf besonders problematische Einzelfälle verweist. Darin liegt kein Widerspruch. Eine (traditionelle) Enzyklopädie muss sich der Anforderung stellen, dass auch der schlechteste Artikel gewisse Qualitätsschranken nicht unterschreitet. Insofern sind die oben genannten Studien – ebenso wie die noch folgenden – nur begrenzt geeignet, Werturteile über die *Wikipedia* oder andere Enzyklopädien zu begründen. Sie

sind jedoch von Nutzen, die Effekte zu verstehen, die aus den neuen Redaktionsprozessen der *Wikipedia* resultieren. Besonders viel versprechend scheinen dabei solche Ansätze zu sein, die es erlauben, Qualitätsindikatoren automatisch für große Textmengen zu bestimmen. In diese Richtung verweisen die Arbeiten von Stvilia, Twidale, Gasser und Smith.

### **3. Die deutsche *Wikipedia* im Vergleich zu *Der Große Brockhaus***

Im Rahmen von Projektseminaren und Abschlussarbeiten wurden im Studiengang Informationswissenschaft der Universität Regensburg innerhalb des letzten Jahres einige Studien über die *Wikipedia* erarbeitet. Hier soll eine Studie vorgestellt werden, welche Qualitätsaspekte der deutschen *Wikipedia* mit solchen des *Brockhaus* vergleicht [Fuchs, Kattenbeck, Sax 2006]. Diese Studie ist sowohl hinsichtlich der Methodik als auch hinsichtlich der Fragestellungen sehr eng mit der zuvor vorgestellten Studie von Herrn Schlieker verwandt, geht aber in mancher Hinsicht über sie hinaus. Ausgewertet wurden die deutsche *Wikipedia* und *Der Große Brockhaus* (ebenfalls 20. Ausgabe). Die Stichprobe wurde nach dem gleichen Verfahren gezogen (random walk im Fall des *Brockhaus*, Zufallsartikel aus der *Wikipedia*), allerdings wurde ein Stichprobenumfang von 50 Lemmata für jede Enzyklopädie so gewählt, dass sie auch in der jeweils anderen vertreten waren. Unterschiede im Ergebnis der Studien können demnach aus dem erweiterten Stichprobenumfang sowie dem erheblichen Anwachsen der *Wikipedia* (zwischen den Stichprobenziehungen liegt ein Jahr 8.2004 – 11.2005), deren Umfang sich in diesem Zeitraum mehr als verdoppelte [Wikipedia 2007c], resultieren.

#### **3.1 Überprüfte Qualitätsaspekte**

Zunächst soll kurz dargestellt werden, welche Qualitätsaspekte in den Vergleich einbezogen werden sollen. Es sei darauf hingewiesen, dass hier keine Auffassung hinsichtlich einer absoluten Interpretation dieser Qualitätsmerkmale vertreten wird.

Nicht überprüft wird in dieser Studie die sachliche Korrektheit der zu vergleichenden Enzyklopädieartikel. Zwar ist dies ein sehr wichtiges Qualitätskriterium, eine inhaltliche Überprüfung einer großen Anzahl zufällig ausgewählter Artikel wäre im Rahmen dieses Projekts jedoch nicht zu leisten gewesen. Im Rahmen eines gleichzeitig durchgeführten Schwesterprojekts wurde jedoch auch die sachliche Korrektheit von *Wikipedia*-Artikeln untersucht. Dabei waren nicht zufällig ausgewählte Artikel Gegenstand der Untersuchung, sondern es wurde eine Vollerhebung zu einem eingeschränkten Gegenstandsbereich durchgeführt, für den Beurteilungskompetenz vorhanden war (Shakespeares Werk) [Molz, Burghardt 2006].

Überprüft wurden in dieser Studie folgende Qualitätsaspekte:

- Zunächst wurde der gegenseitige Abdeckungsgrad der beiden Enzyklopädien festgestellt. Dies kann als eine Operationalisierung des Qualitätskriteriums „Vollständigkeit“ in Bezug auf die Gesamtencyklopädie angesehen werden.
- Um festzustellen, ob der Abdeckungsgrad thematische Abhängigkeiten aufweist, wurden die untersuchten Lemmata thematischen Kategorien zugeordnet, so dass festgestellt werden konnte, ob bestimmte Themen von einer der Enzyklopädien besser repräsentiert wurden.
- Die Vollständigkeit der einzelnen Artikel wurde heuristisch gemessen, indem einerseits die Artikellänge bestimmt wurde. Diese rein quantitative Angabe wurde ergänzt durch subjektive Einschätzungen.
- Die Überprüfbarkeit der in den Artikeln erhobenen Behauptungen wurde anhand der Anzahl der dem Artikel zugeordneten Quellenangaben und deren Korrektheit festgestellt.
- Als ein Indikator für die Korrektheit von Artikeln wurde die Sorgfalt angesehen, die in seine sprachliche Ausarbeitung investiert wurde. Hier wurden Orthographie, Interpunktion als leicht operationalisierbare Qualitätskriterien untersucht.

### 3.2 Thematische Abdeckung

Um 50 Lemmata des *Brockhaus* zu finden, die auch in der *Wikipedia* vorhanden waren, mussten 87 Stichworte gezogen werden. Umgekehrt waren 205 Ziehungen aus der *Wikipedia* erforderlich, um 50 Lemmata zu erhalten, die auch im *Brockhaus* repräsentiert waren. Zur Zeit der Stichprobenentnahme verfügte die *Wikipedia* bereits über 20% mehr Einträge als der *Brockhaus*. Es war also zu erwarten, dass die *Wikipedia* eine bessere Abdeckung erzielen würde. Dass die Abdeckungsquote für den *Brockhaus* noch ungünstiger ausfällt, dürfte auf die große Anzahl von Redirects zurückzuführen sein, die nicht in die Zählung der Lemmata der *Wikipedia* eingehen – ihre Zahl hat sich im Vergleichszeitraum verdreifacht [Wikipedia 2007c].

### 3.3 Umfang und Vollständigkeit der Artikel

Neben der rein thematischen Abdeckung ist auch von Interesse, in welcher Ausführlichkeit die Themen behandelt werden. Hier hat die Online-Enzyklopädie den Vorteil, dass im Gegensatz zur Buchencyklopädie keine Einschränkungen hinsichtlich des Gesamtumfangs zu beachten sind. In der Stichprobe zeigt sich, dass 90 von 100 Artikeln des *Brockhaus* kürzer als 200 Worte sind, während 54 der *Wikipedia*-Artikel länger als 200 Worte sind. Der t-Test für verbundene Stichproben zeigt bei einem Niveau von 5%, dass die Längenunterschiede nicht zufällig sind.

Die Länge eines Artikels ist aber nur ein schwacher Indikator für seine inhaltliche Vollständigkeit. Schließlich kann ein vollständiger Artikel knapp formuliert, ein langer jedoch einfach hoch redundant sein. Als Annäherung an diese Problematik wurde für jedes Artikelpaar von 3 Personen – den drei studentischen Koautoren der Studie – ein Artikel als der vollständigere benannt. Es ist offensichtlich, dass drei Urteile für eine valide Studie zu wenig sind. Zudem waren die Juroren nicht unbefangen, da sie für jeden Artikel wussten, woher er stammte. Die Urteile wurden jedoch nur gewertet, wenn sie für einen Artikel einstimmig ausfielen. Unsicherheit besteht auch aufgrund der schwachen Definition des Begriffs „Vollständigkeit“. Letzterem Problem wurde begegnet, indem für die gefällten Urteile Cohens  $\kappa$  zur Ermittlung der Interrater-Reliabilität bestimmt wurde. Da der  $\kappa$ -Wert nur für ein Beurteilerpaar bestimmt werden kann, so wird er, nach [Roth 1984] für alle Beurteilerpaare ermittelt und der Median der Werte als durchschnittlicher Übereinstimmungsgrad für alle Rater angenommen. Der resultierende Wert von  $\kappa=0,543$  kann in diesem Kontext als Zeichen für ausreichende Überstimmung der Beurteiler hinsichtlich des zu beurteilenden Kriteriums angenommen werden. Das Urteil fiel so aus, dass in 59 Fällen der *Wikipedia*-Artikel als vollständiger angesehen wurde, in 17 Fällen der *Brockhaus*-Artikel. Für die restlichen Fälle ergab sich kein einstimmiges Urteil. Dass hier ein eigenständiges Kriterium erfasst wurde, zeigt sich darin, dass nicht immer der längere Artikel als der vollständigere eingeschätzt wurde. Der Unterschied zwischen den beiden Enzyklopädien wurde mit Hilfe des Binomialverteilungstest bei einem Signifikanzniveau von 5% als nicht zufällig eingestuft. Damit kommt diesem Test, selbst bei Berücksichtigung der oben erwähnten methodischen Probleme, ein zumindest heuristischer Wert zu.

Mit der gleichen Vorgehensweise wurde auch noch versucht, zu einer Einschätzung der Verständlichkeit der Artikel zu kommen. Damit wäre ein weiteres, zweifelsohne wichtiges Qualitätskriterium erfasst worden. Es scheint aber, dass es über Verständlichkeit nur eine schwächere Intuition gibt als über Vollständigkeit. Jedenfalls konnte in diesem Fall keine hinlängliche Interrater-Reliabilität erzielt werden, so dass die diesbezüglichen Ergebnisse verworfen werden mussten.

### 3.4 Themenverteilung

Es wurde weiterhin versucht, festzustellen, ob die unterschiedliche Abdeckung der Gesamtmenge der Lemmata der Stichprobe auf inhaltliche Präferenzen der beiden Enzyklopädien zurückzuführen sein könnte. Dazu wurden zunächst die gemeinsamen Lemmata nach einem Ad-Hoc Kategoriensystem eingeteilt. Das Ergebnis ist in Abb. 1 zu sehen. Sodann wurden auch die Lemmata kategorisiert, die sich nur in der *Wikipedia* (Abb. 2) bzw. nur im *Brockhaus* (Abb. 3) fanden. Während alle nicht in der *Wikipedia* auffindbaren Artikel unschwer in das ursprüngliche Kategoriensystem einzuordnen waren, mussten für die allein in



der *Wikipedia* gefundenen Artikel einige neue Kategorien geschaffen werden. Bezüglich der kleineren Kategorien soll an dieser Stelle keine Aussage getroffen werden. Auffällig ist aber, dass die Kategorie Persönlichkeiten in allen drei Fällen etwa ein Viertel der Lemmata ausmacht. Biographische Information scheint beiden Enzyklopädien wichtig zu sein, Dissens besteht mitunter in der Auswahl der Personen. Der *Brockhaus* scheint eine etwas bessere Abdeckung im Bereich Naturwissenschaft und Technik zu erreichen, während in der *Wikipedia* Themen mit größerem Aktualitätsbezug Einzug gehalten haben (Computertechnik, Film, Musik) bzw. solche, die für traditionelle Enzyklopädien eher nicht ausgewählt würden (Esoterik).

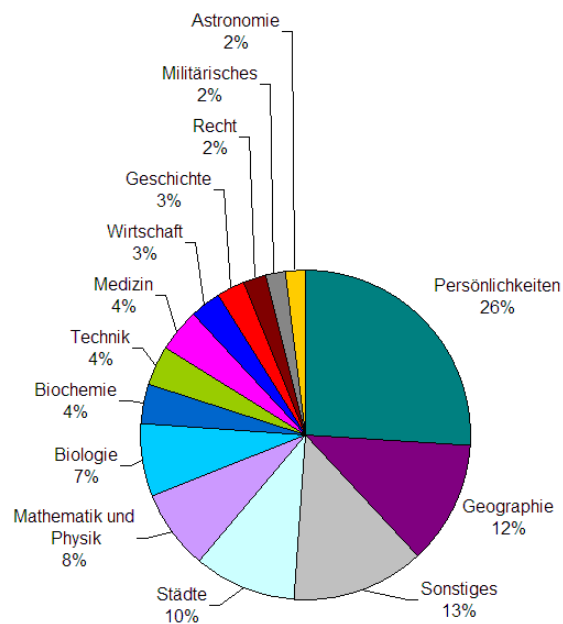


Abb. 1: Kategorien der in beiden Enzyklopädien enthaltenen Lemmata

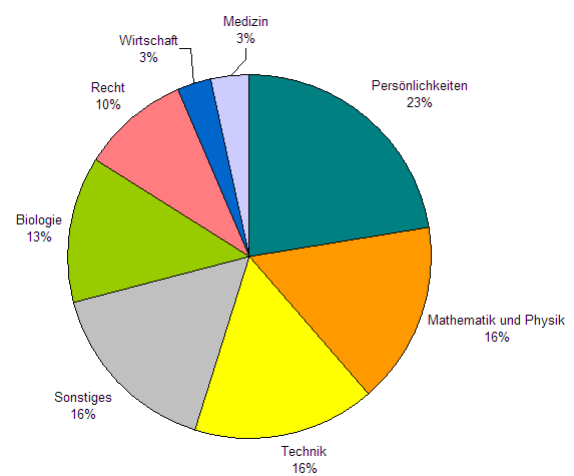


Abb. 2: Kategorien der nur im *Brockhaus* enthaltenen Lemmata

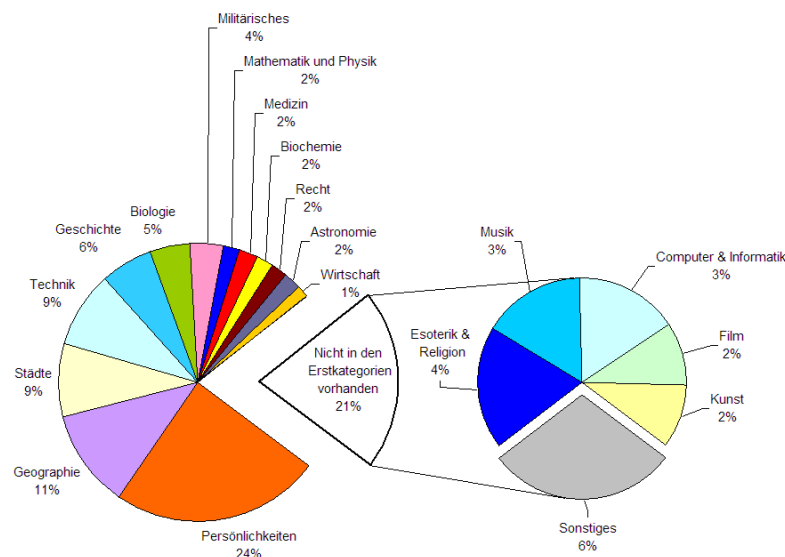


Abb. 3: Kategorien der nur in der *Wikipedia* enthaltenen Lemmata

### 3.5 Informationelle Absicherung durch Quellenangaben

Im Gegensatz zu den Beobachtungen von Schlieker sind in unserer Stichprobe die Artikel der *Wikipedia* besser durch externe Quellen abgesichert. 80 der *Brockhaus*-Artikel verfügten über keine Literaturangaben, aber nur 40 der *Wikipedia*. Zudem enthielten die *Wikipedia*-Artikel grundsätzlich mehr Quellenangaben. 344 Quellenangaben der *Wikipedia* standen 81 Angaben in den *Brockhaus*-Artikeln gegenüber. Dieser Unterschied ist auf der Basis des t-Tests für verbundene Stichproben bei einem Signifikanzniveau von 5% als nicht zufällig anzusehen. Soweit wurden bei dem Test Online-Quellen und bibliographische Angaben gleich gesetzt, da sie beide dem gleichen Zweck dienen. Ein Unterschied wurde bei der Validierung der angegebenen Quellen deutlich. Während alle – rechnet man eine kleinere Unstimmigkeit nicht – Literaturangaben des *Brockhaus*, der auf die Nutzung von Online-Quellen verzichtet, sich als valide erwiesen, waren 10 der Online-Quellen der *Wikipedia* nicht auffindbar. Die bibliographischen Angaben waren in der *Wikipedia* ebenfalls valide.

Sehr auffällig war die hohe Anzahl von Artikeln, in denen Literaturverweise innerhalb des Artikeltexts nicht aufgelöst wurden (*Brockhaus*: 39, *Wikipedia*: 49). Dies stellt für beide Enzyklopädien einen erheblichen Mangel dar, wenngleich der T-Test für verbundene Stichproben zeigt, dass der Unterschied zwischen beiden Enzyklopädien signifikant ist.

### 3.6 Korrektheit in Orthographie und Interpunktion

Grundlage der Beurteilung sind die Regeln der neuen Rechtschreibung, auf die sich beide Enzyklopädien verpflichtet haben. Berücksichtigt werden selbst kleine Ungenauigkeiten, etwa falsches Setzen von Anführungszeichen. Es erwies sich, dass alle Artikel des *Brockhaus* fehlerfrei waren, während nur 31 *Wikipedia*-Artikel frei von Rechtschreibfehlern waren und nur 21 keine Interpunktionsfehler aufwiesen. Der Median des Auftretens von Orthographie- und Interpunktionsfehlern in der *Wikipedia* beträgt 2.

## 4. Zusammenfassung und Ausblick

Im Vergleich zu der Vorgängerstudie von Schlieker hat die *Wikipedia* innerhalb eines Jahres erheblich Fortschritte in der thematischen Abdeckung und in der Qualität der informationellen Absicherung erzielt. Auch die größere Ausführlichkeit der Artikel ist eine Stärke der *Wikipedia*. Hier kann sie die Stärken des Online-Mediums ausnutzen. Die Schwächen in Interpunktion und Orthographie legen jedoch – selbst wenn es sich hier um sprachliche Sekundärtugenden handeln sollte – Schwächen im redaktionellen Prozess offen. Diese Fehler, wenn es auch nicht so viele waren, nähren den Verdacht, dass

auch dort, wo komplexe Sachzusammenhänge redaktionell überprüft werden müssen, Mängel in der Sorgfalt bestehen könnten und sei es nur, weil es keine Zuständigkeiten gibt.

Deutlich wird in dieser Untersuchung auch, dass sie ein Zwischenstadium eines unabgeschlossenen Prozesses erfasst. Zumindest die *Wikipedia* wird sich weiter verändern. Diese Veränderung wird dabei nicht nur quantitativer sondern auch qualitativer Natur sein. Dies zeigt sich schon in der hier dokumentierten neuen Rollenverteilung hinsichtlich der informationellen Absicherung enzyklopädischer Information. Auch für die traditionellen, d.h. nicht in offenen Redaktionsprozessen erstellten Enzyklopädien wird sich ein Druck zur Veränderung ergeben. Die weitere Entwicklung wird auch beeinflusst von der künftigen Rolle von Alternativprojekten zur *Wikipedia*, wie etwa das von Larry Sanger propagierte *Citizendium* [Citizendium 2007].

## Literaturangaben

- Citizendium (2007) The Citizendium Project. <http://www.citizendium.org/>, zitiert am 7.1.2007.
- Patrick Danowski und Jakob Voss (2005) Das Wissen der Welt – die Wikipedia. In: Open Source Jahrbuch. [http://www.opensourcejahrbuch.de/download/jb2005/chapter\\_06/osjb2005-06-05-danowskivoss](http://www.opensourcejahrbuch.de/download/jb2005/chapter_06/osjb2005-06-05-danowskivoss).
- Emigh, W., & Herring, S.C. (2005). In Collaborative authoring on the web: A genre analysis of online encyclopedias. Paper presented at the Proceedings of the Annual Hawaii International Conference on System Sciences. <http://ella.slis.indiana.edu/~herring/wiki.pdf>, zitiert am 6.1.2007.
- Karl-Peter Fuchs, Markus Kattenbeck, Christian Sax (2006) Brockhaus versus Wikipedia. Ausarbeitung zum Projektseminar Hypermedia im WS 2005/2006. Informationswissenschaft, Universität Regensburg.
- Jim Giles (2005) Internet encyclopaedias go head to head. *Nature* 438, S. 900-901, <http://www.nature.com/nature/journal/v438/n7070/full/438900a.html>, zitiert am 6.1.2007.
- Rainer Kuhlen (2005) Wikipedia – Offene Inhalte im kollaborativen Paradigma – eine Herausforderung auch für Fachinformation. *Forschung und Lehre*, Nr. 10, S. 546-548.
- Johannes Molz, Manuel Burghardt (2006) Shakespeares Werk in der Wikipedia. Eine bilinguale Untersuchung zu Umfang und Qualität von Wikipediaartikeln in einem abgegrenzten Themenbereich. Ausarbeitung zum Projektseminar Hypermedia im WS 2005/2006. Informationswissenschaft, Universität Regensburg.
- Jason Lanier (2006) Digitaler Maoismus. Kollektivismus im Internet, Weisheit der Massen, Fortschritt der Communities? Alles Trugschlüsse. <http://www.sueddeutsche.de/kultur/artikel/306/78228/article.html>, zitiert am 29.6.2006
- Andrew Lih (2004): *Wikipedia as Participatory Journalism: Reliable Sources? Metrics for evaluating collaborative media as a news resource*. In: Proceedings of the 5th International Symposium on Online Journalism. <http://jmsc.hku.hk/faculty/alih/publications/utaustin-2004-wikipedia-rc2.pdf>, zitiert a, 6.1.2007.
- Erwin Roth (1984) Sozialwissenschaftliche Methoden. Lehr- und Handbuch für Forschung und Praxis. München et al. 1984.

- Alex Rühle (2006) Im Daunenfedergestöber. Wikipedia-Fälschungen. süddeutsche.de, <http://www.sueddeutsche.de/kultur/artikel/631/90541/>, zitiert am 6.1.2007.
- Christian Schlieker (2005) Explorative Untersuchung von Wissen in kollektiven Hypertexten, Diplomarbeit, Fachbereich 08, Soziologie, Universität Bremen.
- Besiki Stvilia, Michael B. Twidale, Les Gasser und Linda C. Smith (2005) Information Quality Discussions in Wikipedia. In: S. Hawamdeh (Ed.), *Knowledge Management: Nurturing Culture, Innovation, and Technology - Proceedings of the 2005 International Conference on Knowledge Management*. Charlotte, NC: World Scientific Publishing Company. S. 101-113. <http://mailer.fsu.edu/~bstvilia/papers/qualWiki.pdf>, zitiert am 6.1.2007
- Besiki Stvilia, Michael B. Twidale, Linda C. Smith, Less Gasser, (2005a). Assessing information quality of a community-based encyclopedia. In: *Proceedings of the International Conference on Information Quality - ICIQ 2005*. Cambridge, MA. S. 442-454. <http://mailer.fsu.edu/~bstvilia/papers/quantWiki.pdf>, zitiert am 6.1.2007
- Jakob Voss (2005) Informetrische Untersuchungen an der Online-Enzyklopädie Wikipedia, Magisterarbeit im Fach Bibliothekswissenschaft, Institut für Bibliotheks und Informationswissenschaft, Humboldt-Universität zu Berlin.
- Wikipedia (2007) Startseite der Wikipedia. <http://www.wikipedia.org/>, zitiert am 6.1.2007.
- Wikipedia (2007a) Forschungsportal der deutschen Wikipedia. <http://de.wikipedia.org/wiki/Wikipedia:Wikipedistik>, zitiert am 6.1.2007.
- Wikipedia (2007b) Wikipedia Research Bibliography. [http://meta.wikimedia.org/wiki/Wiki\\_Research\\_Bibliography](http://meta.wikimedia.org/wiki/Wiki_Research_Bibliography), zitiert am 6.1.2007.
- Wikipedia (2007c) Wikipedia Statistik Deutsch. <http://stats.wikimedia.org/DE/TablesWikipediaDE.htm>, zitiert am 7.1.2007