

Bots — Nicht-menschliche Mitglieder der Wikipedia-Gemeinschaft

Robin D. Fink / Tobias Liboschik
Arbeitspapier Nr. 28 (Dezember 2010)

ISSN 1612-5355

Herausgeber:

Prof. Dr. Hartmut Hirsch-Kreinsen
Lehrstuhl Wirtschafts- und Industriesoziologie
is@wiso.tu-dortmund.de
www.wiso.tu-dortmund.de/IS

Prof. Dr. Johannes Weyer
Fachgebiet Techniksoziologie
johannes.weyer@tu-dortmund.de
www.wiso.tu-dortmund.de/TS

Wirtschafts- und Sozialwissenschaftliche Fakultät
Technische Universität Dortmund
D-44221 Dortmund

Ansprechpartnerin:

Dipl.-Päd. Martina Höffmann, e-mail: martina.hoeffmann@tu-dortmund.de

Die Soziologischen Arbeitspapiere erscheinen in loser Folge. Mit ihnen werden Aufsätze (oft als Preprint), sowie Projektberichte und Vorträge publiziert. Die Arbeitspapiere sind daher nicht unbedingt endgültig abgeschlossene wissenschaftliche Beiträge. Sie unterliegen jedoch in jedem Fall einem internen Verfahren der Qualitätskontrolle. Die Reihe hat das Ziel, der Fachöffentlichkeit soziologische Arbeiten aus der Wirtschafts- und Sozialwissenschaftlichen Fakultät der Technischen Universität Dortmund vorzustellen. Anregungen und kritische Kommentare sind nicht nur willkommen, sondern ausdrücklich erwünscht.

Inhaltsverzeichnis

Abbildungsverzeichnis	8
1 Einführung	9
1.1 Was sind Bots?	10
1.2 Was macht Bots interessant?	10
1.3 Kurzfassung des Argumentationsgangs	12
1.4 Datengrundlage und Auswertungsmethodik	13
1.5 Zentrale Begriffe	14
2 Editierverhalten	15
2.1 Entwicklung des Editierverhaltens	15
2.2 Vergleich von Bots mit menschlichen Nutzern	17
2.3 Tätigkeitsfelder der Bots	18
3 Bots und Hybridität	21
3.1 Beispiel I: Bots als Initatoren hybrider Interaktionsketten	21
3.2 Beispiel II: Hybride Konzeption eines Bots	23
3.3 Plädoyer für eine hybride Sicht auf Bots	24
4 Weisheit der Wenigen	25
4.1 Exkurs: Kontroverse zur 'wisdom of crowds' in der Wikipedia	25
4.2 Entwicklung von Editieranteilen	26
4.3 Ungleichverteilung der Editierungen	26
5 Fazit und Forschungsperspektiven	29
Literaturverzeichnis	31

Abbildungsverzeichnis

1	Selbstbeschreibungsseite eines Bots	11
2	Editierverhalten	15
3	Zeitliche Entwicklung des Editierverhaltens	16
4	Vergleich von Bots und menschlichen Nutzern	17
5	Aktivität nach Namensraum	18
6	Versionshistorie eines Bildes	21
7	IWLC: Beispiel für die hybride Konzeption eines Bots	23
8	Zeitliche Entwicklung von Editieranteilen	27
9	Ungleichverteilung der Editierungen	28

1 Einführung

Mit dem Kursverfall des Dow-Jones-Index um fast 1000 Punkte am 06.06.2005 (Spiegel Online 2010), der durch einen Tippfehler in Verbindung mit automatisierten Verkäufen sogenannter 'Autonomous Trading Agents' verursacht wurde, ist die Handlungswirksamkeit autonomer Softwareagenten bzw. Bots und ihre Verschränkung mit menschlichem Handeln einmal mehr in das öffentliche Bewusstsein gerückt. Aktivitäten autonomer Softwareagenten finden ihren Vollzug nicht ausschließlich im virtuellen Raum, sondern sind innerhalb hybrider Akteurskonstellationen mit Aktivitäten menschlicher Akteure gekoppelt und damit Gegenstand der Soziologie. Dabei wird ihnen nicht nur die Rolle passiver technischer Artefakte zuteil, sondern sie sind bereits als aktive Mitspieler einer hybriden Gemeinschaft, in unserem Fall der Wikipedia-Gemeinschaft, konzipiert.

Die Idee der Hybrid-Akteure stammt ursprünglich aus der Actor Network Theory (ANT), wenngleich es bei einem ihrer Hauptvertreter, Bruno Latour (1998), um sehr simple bzw. konventionelle Technik geht. Die Bürgerwaffe etwa, die mit ihrem Besitzer zu einem Hybrid-Akteur und damit zu einer möglichen Tötungsmaschine verschmilzt, ist eines der vielfach angeführten Beispiele. Wir möchten uns zwar den ontologischen Postulaten der ANT, etwa der generalisierten Symmetrie, nicht anschließen, sehen allerdings den dort aufgeworfenen Hybriditätsaspekt als wichtigen Impulsgeber für ein modernes Verständnis des Zusammenhangs von Mensch und autonomer Technik.

Rammert u. Schulz-Schaeffer (2002) nehmen die zunehmende Eigenaktivität nicht-menschlicher Akteure, etwa von Software-Agenten, zum Anlass, hybride Konstellationen in ihrem „graduellen Modell verteilter Handlungsträgerschaft“ zu konzeptualisieren. Sie fassen Handlungsfähigkeit (*agency*) als Attributionsphänomen auf und stellen die ergebnisoffene Forschungsfrage, ob in einem spezifischem Handlungszusammenhang Mensch und Technik gleich zu behandeln seien. Neuere Forschungsarbeiten (Fink u. Weyer 2011) operationalisieren die Interaktion zwischen Mensch und autonomer Technik mit Hilfe eines hybriden Handlungsmodells, dem Modell soziologischer Erklärung hybrider Systeme, und fassen autonome Technik als „Herausforderung der soziologischen Handlungstheorie“ auf.

Das vorliegende Forschungspapier möchte keinen theoretischen Beitrag zur Diskussion um autonome Technik und hybride System leisten, sondern anhand des Fallbeispiels der Wikipedia-Bots zeigen, in welcher Weise autonome Technik aktiv wird und in hybride Interaktionszusammenhänge eingebettet ist. Im Gegensatz zu Untersuchungen zur Mensch-Maschine-Interaktion, etwa von Braun-Thürmann (2007), geht es bei der hier betrachteten autonomen Technik nicht um verkörperte bzw. in ihrer Erscheinung anthropomorphe künstliche Intelligenz, sondern um

Software-Agenten, die kein menschenähnliches Äußeres aufweisen. Die Software-Agenten, Bots in der Sprechweise der Wikipedia, und ihr Beitrag zur Wikipedia-Gemeinschaft werden hinsichtlich ihrer sichtbaren Handlungsvollzüge auf der Plattform und ihres Beitrages zur gemeinsamen Lösung von Problemen evaluiert.

Bisherige (soziologische) Forschungsarbeiten zur Wikipedia konzentrieren sich insbesondere auf die Frage nach der Entstehung eines eigentlich „unmöglichen“ öffentlichen Gut[s]“ (Stegbauer 2006: 228) sowie auf die Entstehung von positionalen Strukturen innerhalb der Wikipedia-Gemeinschaft. Gegenstand des vorliegenden empirisch ausgerichteten soziologischen Arbeitspapiers sind hingegen Bots, nicht-menschliche Mitglieder der Wikipedia-Gemeinschaft, und die empirische Frage nach ihrem Beitrag für das Gesamtsystem Wikipedia.

Im folgenden Abschnitt wird zunächst die Frage geklärt, was Bots sind und was sie im Falle der Wikipedia als Forschungsgegenstand interessant macht.

1.1 Was sind Bots?

Unter Bots versteht man nicht, wie die Begriffsnähe zu 'Robot' und Darstellungen von Bots, etwa in Filmen wie 'Krieg der Sterne' andeuten könnten, humanoide Roboter, sondern vielmehr Software, die im Dienste ihrer Programmierer insbesondere repetitive Tätigkeiten (teil-)autonom erledigt. Im Falle der Wikipedia nehmen Bots zumeist kleinere Editierungen vor, geben Hinweise auf formale Mängel und Inkonsistenzen in Artikeln. Die Wikipedia-Gemeinde charakterisiert Bots als „[...] Computerprogramme oder Skripte, die ihren Betreibern stupide, langweilige und häufig auftretende Aufgaben abnehmen“¹.

Bots tauchen in der Wikipedia, ebenso wie menschliche Nutzer, als angemeldete Benutzer auf. Sie sind allerdings durch ihren Namen und durch eine spezielle Kennzeichnung als Bots erkennbar. In der Regel lassen sich Bots über ihre Selbstbeschreibungsseite (vgl. Abbildung 1) eindeutig einem Programmierer zuordnen. Neben dieser Zuordnung finden sich auf der Selbstbeschreibungsseite in der Regel Angaben zur Funktionsweise des Bots und in manchen Fällen auch der Quelltext. Einige Bots sehen über ihre Selbstbeschreibungsseite eine Sperrfunktion vor, die bei Bot-Fehlfunktion (z. B. massenhaft falsche Editierungen) von menschliche Nutzer genutzt werden kann.

1.2 Was macht Bots interessant?

Paradoxerweise halten wir Bots gerade deswegen für beachtenswert, weil sie im überwiegenden Teil der uns bekannten Wikipedia-Studien entweder herausgefiltert wurden (z. B. Voss 2005; Ortega u. Gonzalez Barahona 2007; Javanmardi u. a. 2009), ihr Einfluss als marginal angesehen (Kittur u. a. 2007) oder nicht gesondert betrachtet wurde (Stegbauer 2009). Angesichts der keinesfalls schlechteren Datenlage über Bots ist fraglich, ob ein solcher Umgang berechtigt ist.

¹Quelle: <http://de.wikipedia.org/wiki/Wikipedia:Bots> (eingesehen am 15.10.2010)

Benutzer:PortalBot

Hallo Leute!

Ich bin ein reiner Pywikipediabot. Gesteuert werde ich vom
Benutzer:Augiasstallputzer.

+	Dieser Benutzer ist ein Bot .
Wichtige Hinweise für andere Benutzer	
Wer möchte, dass ein Bot etwas für ihn erledigt, der soll hier anfragen.	
Wenn der Bot läuft, dann bin ich grundsätzlich unter Benutzer Diskussion:Augiasstallputzer zu erreichen.	
Wenn nach einer Verschiebung das Redirect gelöscht werden soll, dann fixiere ich gelegentlich auch Links auf Benutzerseiten, denn es ist für einen Benutzer in aller Regel besser, wenn ein Bot einen Link ändert, als dass dieser Link plötzlich tot ist und der Benutzer nicht weiß, ob und welchen Ersatz es gibt. Wer das trotzdem nicht möchte, der kann auf seiner Benutzerseite, am Besten ganz oben, den Kommentar: <code><!--NOBOTLINKFIX--></code> einfügen. Mein Bot überspringt diese Seite dann.	

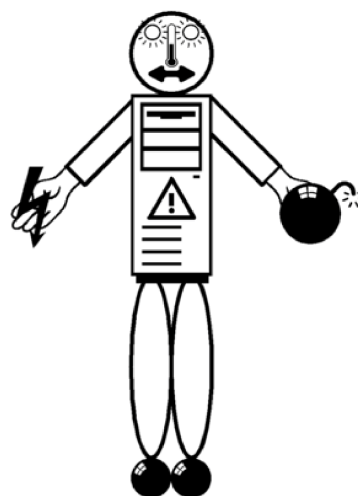


Abbildung 1: Selbstbeschreibungssseite eines Bots

Ebenso sträflich vernachlässigt werden die unregistrierten Nutzer, wenngleich aus technischen Gründen die Zuordnung zu unterschiedlichen Personen schwierig ist und deshalb hier die Datenlage tatsächlich schlechter ist.

Ein erster quantitativer Überblick zeigt, dass pro Woche durchschnittlich etwa 60 Bots aktiv sind, die etwa 10% des gesamten Editiervolumens in der Wikipedia ausmachen. Ihre Bedeutung ist somit auf ähnlich hohem Niveau wie das Editiervolumen der unregistrierten Nutzer, das ebenfalls bei etwa 10% liegt.

Neben rein quantitativen Kenngrößen ist an dem grundsätzlichen Konzept von Bots ihre dezentrale Architektur besonders interessant. Aufgaben, die im Prinzip auch zentral auf Wikipedia-Servern erledigt werden könnten, werden dezentral, ganz im Sinne des Wiki-Prinzips (dezentrale, kollaborative Wissensgenerierung), von (teil-)autonomen Softwareagenten durchgeführt. Diese Beobachtung untermauert die von Hannah Mormann stammende These, *Software als Datenquelle*² anzusehen. Folgt man Mormann, die hauptsächlich betriebswirtschaftliche Software für Unternehmen untersucht hat, finden sich in den softwaretechnischen Umsetzungen Daten, die Aussagen über die organisationalen Strukturen eines Unternehmens ermöglichen.

Im Falle der Wikipedia werden Aufgaben wie z.B. Tippfehlerkorrekturen oder Kategorisierungen, die durchaus zentral organisiert werden könnten, durch Bots dezentral gelöst. Teilweise werden Bots auf den Computern der Programmierer und damit unabhängig von der Wikipedia-Infrastruktur, ausgeführt. In vielen Fällen erfolgt auch die Erstellung eines Bots nach dem Wiki-

²Vortrag auf dem DGS-Kongress 2010 im Rahmen der Ad-hoc-Gruppe „Auf dem Weg in die Datengesellschaft“, Download: http://www.uni-due.de/imperia/md/content/soziologie/schulz-schaeffer/mormann_realit__tdatenverarbeitung.pdf

Prinzip. Eine Idee für einen Bot wird gemeinsam diskutiert und nach gründlicher Diskussion dann softwaretechnisch umgesetzt³. Zusammenfassend kann argumentiert werden, dass sich die Idee der Wikipedia durchaus in der Softwarearchitektur der Bots, ihrer Ausführungsumgebung und in ihrer Entstehungsgeschichte widerspiegelt.

1.3 Kurzfassung des Argumentationsgangs

Wir möchten uns dem Phänomen Bots aus unterschiedlichen Perspektiven nähern. Die hier vorgestellte Arbeit präsentiert einen Zwischenstand, der zwar einige Fragen hinsichtlich der Aktivitäten und des Editierverhaltens der nicht-menschlichen Mitspieler der Wikipedia klären wird, der aber mindestens genauso viele offene Fragen aufwirft und weitere Forschungsperspektiven eröffnet.

Zunächst betrachten wir in Abschnitt 2 das Editierverhalten der Bots und weisen insbesondere nach, dass eine sehr kleine Anzahl von Bots einen beachtenswerten Anteil am Gesamteditervolumen in der Wikipedia produziert. In einer näheren Analyse der Aktivitäten der Bots, die u.a. Tätigkeitsfelder benennt, in denen Bots besonders aktiv sind, stellt sich heraus, dass ihre Editierungen eher „lektorativen“ Charakter haben, was zunächst vielleicht wenig überraschend ist. Die zur Zeit aktive Bot-Generation ist noch weit davon entfernt, autonom längere, inhaltlich gehaltvolle Textbeiträge zu verfassen. Die Handlungswirksamkeit von Bots wird allerdings dann interessant, wenn man sie im Kontext hybrider Interaktionsketten, also beim Zusammenwirken von Bots und menschlichen Nutzern, betrachtet. Ein Beispiel wird dabei anschaulich zeigen, wie Bots Initiatoren von Interaktionsketten werden und damit handlungswirksam sind. Ein weiteres Beispiel wird zeigen, dass Bots z.T. bereits hybrid konzipiert, d. h. auf die Zusammenarbeit von menschlichen und nicht-menschlichen Akteuren ausgelegt sind.

Ausgehend davon, dass sehr wenige Bots sehr viele Editierungen durchführen, Bots nahtlos in Interaktionsketten eingebettet sind und ihre Programmierung spezielle Kenntnisse erfordert, stellen Bots das Prinzip der 'wisdom of crowds', der Weisheit der Vielen, in Frage, das davon ausgeht, dass Projekte wie die Wikipedia insbesondere von vielen getragen werden, die wenig beitragen. Neben den Analysen, die sich auf Bots fokussieren, fanden wir im vorliegenden Datenmaterial starke Indizien für eine Zunahme an Ungleichheit (wer trägt wie viel bei?) in der Wikipedia. Dieses Phänomen zunehmender Ungleichheit ist zwar grundsätzlich nicht neu, wenngleich es dazu widersprüchliche Forschungsergebnisse gibt. Unser Ansatz geht dabei methodisch über bisherige Analysen hinaus und berücksichtigt insbesondere auch unregistrierte Nutzer. Uns erscheint es vor dem Hintergrund der analysierten Daten äußerst fraglich, weiterhin von einer 'wisdom of crowds' zu sprechen.

³Für ein Beispiel einer Diskussion über Bot-Architektur und -Funktionsweise s. http://de.wikipedia.org/w/index.php?title=Vorlage_Diskussion:Infobox_Gemeinde_in_Frankreich#Botlauf_M.C3.A4rz_2010 (eingesehen am 22.11.2010)

1.4 Datengrundlage und Auswertungsmethodik

Der Zugriff auf die Daten erfolgte über den Wikimedia Toolserver⁴. Über den Toolserver ist es möglich, replizierten Echtzeitzugriff auf sämtliche Wikipedia-Datenbanken (ohne Artikelinhalte) zu erhalten und mit der Datenbanksprache SQL komplexe Abfragen an diese zu stellen. Damit unterscheidet sich diese Zugriffsart von einer anderen häufig verwendeten, die auf sogenannten „Dumps“⁵ basiert, darin, dass auf einfache Weise sehr viel komplexere und differenziertere Abfragen durchgeführt werden können und prinzipiell sogar Echtzeitauswertungen möglich sind. Eine dritte Möglichkeit, Daten aus der Wikipedia maschinell zu prozessieren, ist die Nutzung der Mediawiki-API⁶, welche hinsichtlich ihres Funktionsumfangs im Vergleich zu SQL-Abfragen stark eingeschränkt ist.

Unser Betrachtungszeitraum endet am 30.04.2010, um im Verlauf der Auswertungen auf eine konsistente Datenbasis zugreifen zu können. Für die spätere Auswertung wurde eine eigene Datenbank mit zwei zentralen Tabellen angelegt, in denen Informationen aus verschiedenen Tabellen der Wikipedia-Datenbanken zusammengeführt wurden. Wesentlich waren dabei vor allem die Tabellen *revision*, welche Metainformationen zu allen in der Wikipedia durchgeführten Editierungen enthält, *user*, welche Informationen zu allen registrierten Benutzern enthält und *page*, welche Metainformationen zu allen in der Wikipedia geführten Seiten enthält. Die erste der beiden selbst generierten Tabellen (vgl. Tabelle 1) enthält Informationen zum Editierverhalten jedes angemeldeten Benutzers im Zeitraum vom 01.02.2010 bis zum 30.04.2010 und dient der Analyse aktueller Entwicklungen des Editierverhaltens. Die zweite selbst generierte Tabelle (vgl. Tabelle 2) enthält Informationen zum Editierverhalten in jeder Woche von der Gründung der deutschsprachigen Wikipedia bis zum 30.04.2010 und dient der Analyse zeitlicher Entwicklungen. Die zur Befüllung dieser beiden Tabellen notwendigen iterierten SQL-Abfragen wurden mit der Programmiersprache Perl skriptgesteuert umgesetzt.

Die Auswertungen erfolgten mit dem frei verfügbaren statistischen Programmpaket R⁷, mit dem ein direkter Zugriff auf die selbst generierte Datenbank möglich war.

Benutzer	Namensraum	# Editierungen	Ø Revisionslänge
BenutzerX	Artikel	730	10047.5014
BenutzerX	Artikel_Diskussion	5	8479.2000
BenutzerY	Artikel	7	8513.0000
...

Tabelle 1: Struktur der selbst generierten Tabelle zum Editierverhalten einzelner Nutzer, differenziert nach Namensraum

⁴<http://www.toolserver.org>

⁵Dumps umfassen i.d.R. vollständige Datenbankabbilder. Sie sind Momentaufnahmen, die die vollständige Datenbasis zum Zeitpunkt der Erstellung des Dumps umfassen. Wikipedia-Datenbank-Dumps sind downloadbar unter: <http://download.wikimedia.org>

⁶<http://www.mediawiki.org/wiki/API>

⁷<http://www.r-project.org>

Woche	Namensraum	Benutzer- gruppe	# Editierungen	Ø Revisions- länge	# aktiver Benutzer
...
Woche 424	Artikel	Bots	32961	9298.1888	70
...
Woche 424	Artikel_Diskussion	Registrierte	1805	23803.8205	846
...

Tabelle 2: Struktur der selbst generierten Tabelle zur wochenbasierten Auswertung, differenziert nach Namensraum und Benutzergruppe

1.5 Zentrale Begriffe

Im weiteren Verlauf der Arbeit werden einige Begriffe immer wieder verwendet. In diesem Abschnitt werden die zentralen Begriffe erläutert. Dabei wird begrifflich unterschieden zwischen *Maßzahlen* und *Merkmalen*, nach denen in der Auswertung differenziert wurde.

Die zentrale Maßzahl ist die Anzahl der vorgenommenen Editierungen. Als Editierung zählt dabei eine an einer beliebigen Seite in der Wikipedia vorgenommene Änderung beliebigen Ausmaßes, welche in einem abgeschlossenen Editiervorgang durchgeführt wird.

Um den Umfang einer solchen Editierung beurteilen zu können, wird als weitere Maßzahl die Revisionslänge herangezogen. Diese gibt dabei die Anzahl der in einer Editierung geänderten Zeichen in der Einheit Byte an. Ein Benutzer gilt als „aktiv“, wenn er mindestens eine Editierung vorgenommen hat.

Das zentrale Merkmal, nach dem in den folgenden Auswertungen unterschieden wird, ist die Benutzergruppe. Dabei werden drei Benutzergruppen betrachtet: *Registrierte*, *Unregistrierte* und *Bots*. Die erste Benutzergruppe umfasst nicht die als Bot gekennzeichneten Benutzer, da diese als eigene Benutzergruppe geführt werden. Editierungen unregistrierter Benutzer werden, soweit möglich, über die gespeicherte IP-Adresse identifiziert. Wenn von „Benutzern“ gesprochen wird, sind stets Autoren gemeint, nicht Leser der Wikipedia.

Ein weiteres Merkmal ist der Namensraum, in dem eine Editierung vorgenommen wurde. Namensräume (engl.: namespaces) sind in der Wikipedia ein Konstrukt, um unterschiedliche „Abteilungen“ voneinander zu trennen. So finden z. B. Diskussionen zu Artikeln in einem anderen Namensraum statt als inhaltliche Arbeiten an Artikeln. Die Berücksichtigung des Namensraums als Merkmal gibt Aufschluss darüber, in welchen Bereichen editiert wird bzw. um welche Art der Editierung es sich handelt.

2 Editierverhalten

Um einen ersten Eindruck über das Editierverhalten in der Wikipedia zu bekommen, betrachten wir zunächst das Editierverhalten in den Monaten Februar bis April 2010. In diesem Zeitraum waren insgesamt 50.000 Nutzer in allen Namensräumen aktiv, davon 112 Bots. Über die Anzahl aktiver unregistrierter Benutzer kann hier keine Aussage gemacht werden, da wir deren Aktivitäten über den gesamten Zeitraum betrachtet haben. Über einen längeren Zeitraum ist es nicht möglich, einzelne Benutzer auf Grundlage der gespeicherten IP-Adresse zu identifizieren. Im Vergleich von Abbildungen 2a und 2b ist zu erkennen, dass Bots nur einen verschwindend geringen Anteil aller aktiven Benutzer ausmachen, aber mit über 15% einen erheblichen Anteil an der Anzahl aller Editierungen haben.

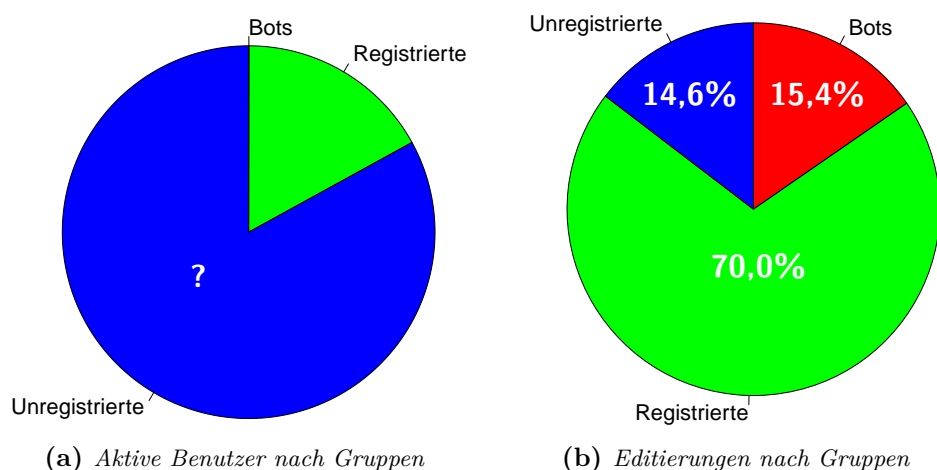


Abbildung 2: Editierverhalten im Zeitraum 01.02.2010–30.04.2010

2.1 Entwicklung des Editierverhaltens

Während in Abbildung 2 über die Zeit aggregierte Daten dargestellt sind, lassen sich in den beiden zeitlichen Darstellungen in Abbildung 3 längere Entwicklungstendenzen ablesen. Beide wurden mit einer Fensterbreite von 53 Wochen geglättet¹.

¹Als geglätteter Wert einer Woche wurde das arithmetische Mittel über 26 Wochen vor bis 26 Wochen nach dieser Woche verwendet, also insgesamt über 53 Beobachtungswerte gemittelt. Für die ersten sowie die letzten 26 Wochen wurde deshalb kein geglätteter Wert bestimmt. Die Glättung über ungefähr die Dauer eines Jahres führt dazu, dass jährliche Saisoneffekte herausgemittelt werden.

Betrachtet man die Anzahl aktiver Bots, wir definieren sie als aktiv, wenn sie in der jeweiligen Woche mindestens eine Editierung durchgeführt haben, so ist festzustellen, dass in den Jahren 2006 bis 2008 die Aktivität stark zunimmt, von 2008 bis 2009 leicht rückläufig ist und sich aktuell offenbar auf einem Niveau von etwa 60 aktiven Bots bewegt. Diese Entwicklung ist in Abbildung 3a wegen der großen Skala allerdings nicht abzulesen. Die Tatsache, dass die Anzahl der durch Bots durchgeführten Editierungen zunimmt, während die Anzahl aktiver Bots stagniert, könnte darauf hindeuten, dass die aktiven Bots immer intelligenter werden und Änderungen an immer mehr Artikeln durchführen.

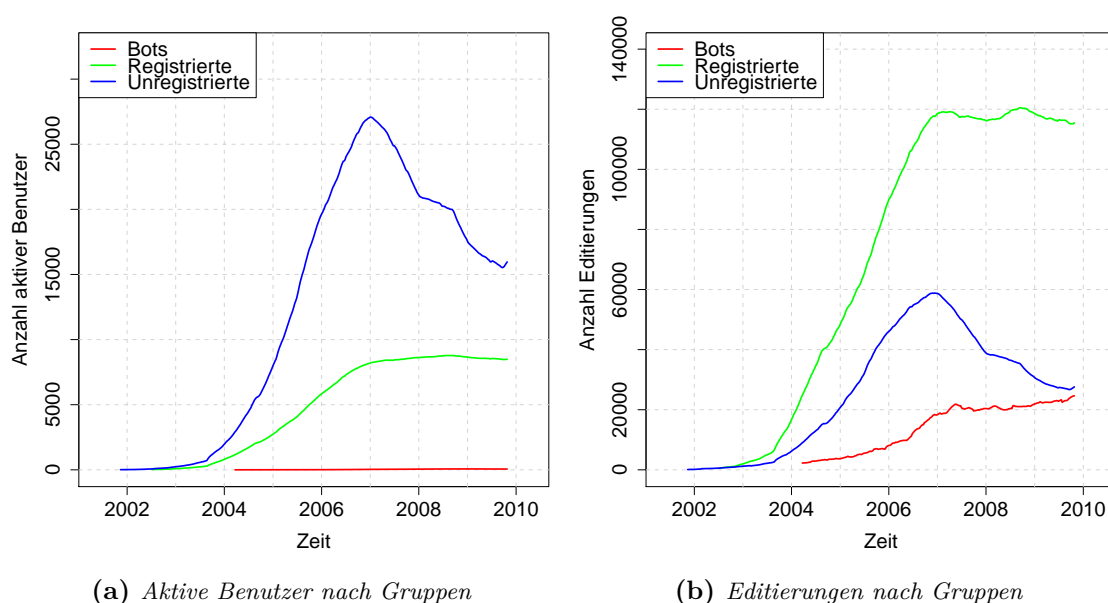


Abbildung 3: Zeitliche Entwicklung des Editierverhaltens für den Zeitraum 01.02.2010–30.04.2010 (geglättete Darstellung, Fensterbreite: 53 Wochen)

Schaut man sich die Entwicklung des Editierverhaltens der drei Benutzergruppen Registrierte, Unregistrierte und Bots über einen längeren Zeitraum an (Abbildung 3b), sind relativ klare Entwicklungsphasen und Tendenzen erkennbar: Bots gibt es erst seit Mitte 2003, wegen der geglätteten Darstellung beginnt der zugehörige Graph sogar erst etwas später. Bis zum Jahr 2007 steigt in allen Benutzergruppen die Anzahl der Editierungen. Im Jahr 2007 gibt es offenbar einen Wendepunkt, ab dem die Anzahl der Editierungen bei den registrierten Benutzern stagniert. Die Anzahl der Editierungen von unregistrierten Benutzern nimmt seitdem stark ab, während die Anzahl der von Bots durchgeführten Editierungen zunimmt. Unter anderem wurden Anfang 2008 sogenannte Artikelsichtungen eingeführt, bei denen jede Änderung eines anonymen Benutzers zunächst gesichtet werden muss, bevor diese öffentlich einsehbar ist. Der starke Rückgang der Anzahl aktiver unregistrierter Benutzer könnte somit auch als erfolgreiche Vandalismusbekämpfung angesehen werden. In der Studie von Anthony u. a. (2005) finden sich allerdings auch Hinweise darauf, dass gerade die Beiträge nicht registrierter Benutzer äußerst wichtig sind. Es

wird hervorgehoben, dass Beiträge von höchster Qualität häufig von nicht angemeldeten Nutzern, sogenannten 'Good Samaritians' stammen.

Im Verhältnis der Anzahl registrierter und unregistrierter Benutzer sowie Bots zu der jeweiligen Anzahl von Editierungen finden sich bereits Anhaltspunkte für eine zunehmende Ungleichheit in Bezug auf die Editierungen: Immer weniger Nutzer tragen immer mehr bei. Auf diesen Zusammenhang werden wir in Abschnitt 4 noch zurückkommen.

2.2 Vergleich von Bots mit menschlichen Nutzern

Beim Vergleich von Bots und menschlichen Nutzern anhand von Boxplots² ist ein stark voneinander abweichendes Editierverhalten festzustellen (Abbildung 4).

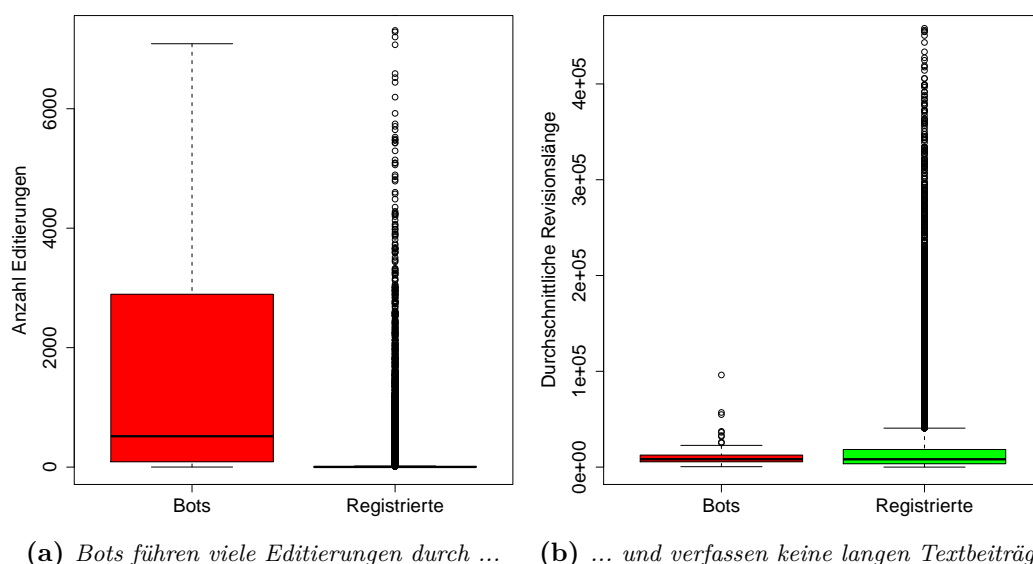


Abbildung 4: Vergleich von Bots und menschlichen Nutzern in Bezug auf die Anzahl der Editierungen (a) und der Revisionslänge (b) im Zeitraum 01.02.2010–30.04.2010

Mehr als 75% der Bots haben mehr Editierungen durchgeführt als der Durchschnitt der registrierten Nutzer (Abbildung 4a). Es gibt, erkennbar an den zahlreichen Ausreißern, sehr aktive menschliche Nutzer. Bots hingegen sind fast alle sehr aktiv, gemessen an der Anzahl ihrer Editierungen.

Bei Bots gibt es praktisch keine längeren Editierungen, während bei den registrierten Benutzern solche zahlreich vorkommen, da diese z.T. sehr textlastige Artikel-Editierungen vornehmen (Abbildung 4).

²Ein Boxplot charakterisiert grafisch die Verteilung eines Beobachtungsmerkmals. Jeweils 25% der Beobachtungswerte sind größer (bzw. kleiner) als der Wert an der Oberkante (bzw. Unterkante) der eingezeichneten Box. Als waagerechte Linie innerhalb der Box ist der Median der Beobachtungswerte eingezeichnet.

Bots editieren durch ihr automatisiertes Vorgehen zwar viel, schreiben aber keine längeren Textbeiträge. Diese Charakterisierung deckt sich weitgehend mit der zu Anfang präsentierten Selbstbeschreibung der Bots.

2.3 Tätigkeitsfelder der Bots

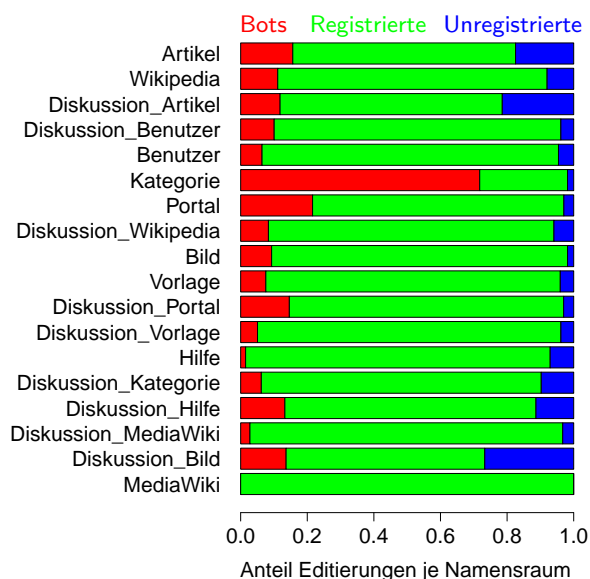


Abbildung 5: Aktivität nach Namensräumen im Zeitraum 01.02.2010–30.04.2010

Namensraum	# Editierungen	Gesamtanteil (%)
Artikel	2 301 800	68.88
Wikipedia	252 244	7.55
Diskussion_Artikel	207 694	6.21
Diskussion_Benutzer	183 893	5.50
Benutzer	169 347	5.07
Kategorie	70 782	2.12
Portal	51 539	1.54
Diskussion_Wikipedia	33 453	1.00
Bild	31 616	0.95
Vorlage	21 657	0.65
Diskussion_Portal	12 097	0.36
Diskussion_Vorlage	2 135	0.06
Hilfe	1 076	0.03
Diskussion_Kategorie	903	0.03
Diskussion_Hilfe	768	0.02
Diskussion_MediaWiki	394	0.01
Diskussion_Bild	337	0.01
MediaWiki	208	0.01
Σ	3 341 943	100

Tabelle 3: Anzahl der Editierungen differenziert nach Namensräumen im Zeitraum 01.02.2010–30.04.2010: Die Editierungen im Artikelnamensraum dominieren mit einem Anteil von fast 69% an allen Editierungen

Eine rein quantitative Analyse der Bot-Aktivitäten nach Namensräumen (Abbildung 5) zeigt, dass Bots in allen Namensräumen aktiv sind. Besonders auffällig ist der hohe Anteil der Bot-Editierungen im Namensraum *Kategorie*. Dieser ist damit zu erklären, dass die Kategorisierung von Artikeln eine z. T. mechanische Arbeit ist. Einige Bots, die sich mit Kategorien beschäftigen, rekategorisieren en bloque direkt mehrere hundert Artikel und führen somit zahlreiche Editierungen in diesem Namensraum durch. Eine weitere interessante Beobachtung ergibt sich für die unregistrierten Nutzer im Namensraum *Diskussion_Bild*: Da es innerhalb der Wikipedia oft zu Diskussionen über Urheberrechte an Bildern kommt, sind unregistrierte Nutzer an Diskussionen über Bilder überdurchschnittlich häufig beteiligt, da sie vermutlich nur ungenau über die rechtlichen Problematiken Bescheid wissen, die beim Einstellen von Bildern in der Wikipedia entstehen.

Bei der Interpretation der Grafik ist jedoch Vorsicht geboten: Wie Tabelle 3 zeigt, haben die meisten Editierungen (fast 69 Prozent) im Artikel-Namensraum stattgefunden. Bezogen auf

Tätigkeitsfeld	Anzahl
Interwiki-Links setzen/ändern/entfernen	86
Administrative Aufgaben	10
Links korrigieren	9
Kategorien verschieben/umbenennen/löschen	6
Vorlagen aktualisieren (z.T. mit externen Daten)	5
Rechtschreib- und Typographiefehler korrigieren	4
Bilder auf korrekte Auszeichnung überprüfen	3
Listen pflegen	2
Vandalismus erkennen	1
<i>Artikel verfassen</i>	<i>0</i>

Tabelle 4: Tätigkeitsfelder der Bots (Zeitraum: 01.02.2010–30.04.2010)

die Bots bedeutet das, dass sie relativ zu den anderen Benutzergruppen in diesem zwar weniger editiert haben als beispielsweise im Namensraum *Kategorie*, in absoluten Zahlen haben sie jedoch die meisten Editierungen im Artikelnamensraum durchgeführt.

Nachdem gezeigt wurde, dass Bots an Bedeutung zunehmen und in allen Namensräumen und damit Bereichen der Wikipedia aktiv sind, ist die Frage von zentraler Bedeutung, ob die Bot-Editierungen inhaltlich relevant sind.

Tabelle 4 zeigt das Ergebnis einer qualitativen Analyse der 112 in den letzten drei betrachteten Monaten, Februar bis April 2010, aktiven Bots. Eine erste Durchsicht der Bot-Editierungen und einschlägiger Informationen aus der Wikipedia³ führte zu einer Liste potenzieller Bot-Tätigkeitsfelder (1. Tabellenspalte). Betrachtet wurden dann alle 112 Bots, die aktiv waren, also mindestens eine Editierung vorgenommen haben. Neben den Selbstbeschreibungsseiten der Bots, die allerdings hinsichtlich der durchgeführten Tätigkeiten z.T. übertreiben, wurden vor allem die Versionshistorien herangezogen, die sämtliche Änderungs-Details enthalten. Auf Basis dieser Informationen lassen sich die Bots einem oder mehreren der in der Tabelle aufgeführten Tätigkeitsfelder zuordnen. Die Information in der zweiten Tabellenspalte gibt an, wie viele Bots in dem jeweiligen Tätigkeitsfeld aktiv waren. Wir haben dabei bewusst nicht auf die formalen Namensräume als Differenzierungsmerkmal zurückgegriffen, da die Angabe, dass ein Bot beispielsweise ausschließlich im Artikel-Namensraum editiert hat, keine Auskunft darüber gibt, *welche Tätigkeit* er dort genau durchgeführt hat.

Auffällig an der Tabelle ist zunächst, dass mit 86 von 112 registrierten Bots eine Vielzahl dieser *Interwiki-Bots* sind. Interwiki-Bots kümmern sich um die Verknüpfung von thematisch deckungsgleichen Artikeln in unterschiedlichen Sprachversionen der Wikipedia. Der Grund für die Vielzahl von Interwiki-Bots liegt zum einen darin, dass ein standardisiertes Framework zur Erstellung von Interwiki-Bots existiert, womit die Hürde zur Programmierung eines solchen Bots relativ niedrig ist. Zum anderen sind angemeldete Interwiki-Bots häufig keine nativen Bots, d.h.

³Liste der Bots, Link: http://de.wikipedia.org/wiki/Wikipedia:Bots/Liste_der_Bots (eingesehen am 22.12.2010)

Helfer von vornehmlich in der deutschsprachigen Wikipedia aktiven Programmierern, sondern entstammen den zahlreichen fremdsprachlichen Versionen.

Weitere Tätigkeitsfelder sind *administrative Aufgaben*, die sich mit internen Wikipedia-Abläufen befassen, z. B. Archivierung, Artikelumbenennungen und Organisation und Zusammenfassung umfangreicher Diskussionen. Wie in Tabelle 4 genannt, korrigieren neun Bots fehlerhafte *Links*. Dies ist z. B. notwendig, wenn sich interne Artikel- oder Bildernamen ändern oder wenn externe Web-Ressourcen nicht mehr verfügbar sind. Wikipedia verfügt über ein umfangreiches *Kategorien*system, das von aktuell sechs Bots unterstützend gepflegt wird. *Vorlagen*, die in Artikeln verwendet werden, vereinheitlichen Artikel (z. B. für tabellarische Angaben zu Städten). Werden Vorlagen allerdings verändert, müssen zahlreiche Artikel auf die neuen Vorlagenformate angepasst werden – eine typische Tätigkeit für einen der fünf dort aktiven Bots. Immer wiederkehrende *Rechtschreib- und Typographiefehler* werden von vier Bots korrigiert. Weitere drei Bots kümmern sich um die *korrekte Auszeichnung von Bildern* und überprüfen beispielsweise, ob Lizenzinformationen korrekt vermerkt sind. *Listen*, die z. B. auf Themenportalen zu finden sind, müssen regelmäßig gepflegt werden, woran zwei Bots mitwirken. Die Erkennung und Meldung von *Vandalismus*, der z. B. durch massenhafte oder sinnlose Editierungen böswilliger Nutzer entsteht, wurde im Betrachtungszeitraum durch einen Bot durchgeführt.

Wie der Tabelle zu entnehmen ist, stellen Bots, die sich um Interwiki-Verlinkungen kümmern, quantitativ die größte Tätigkeitsgruppe dar. Auf den ersten Blick scheinen Interwiki-Bots keine besonders intelligente Arbeit zu verrichten, sie verlinken lediglich thematisch identische Artikel aus unterschiedlichen Sprachversionen. Betrachtet man die Arbeit der Bots allerdings genauer, wie wir es anhand zweier Bots durchgeführt haben, erscheint dieser erste Eindruck wesentlich differenzierter, insbesondere in Hinblick auf das Vermögen der Bots, Folgehandlungen anderer Nutzer zu initiieren. Auf diesen Aspekt werden wir in Abschnitt 3 weiter eingehen.

Die Bots der verschiedener Kategorien sind hinsichtlich ihres Verhaltens unterschiedlich weit fortgeschritten. Besonders avancierte Bots nehmen sogar Editierungen auf Grundlage selbst beschaffter externer Daten vor (PatDi BOT)⁴. Deutlich geworden ist aber auch, dass Bots (bisher) Artikel nicht eigenständig verfassen.

Insgesamt lässt sich konstatieren, dass die einzelnen von Bots durchgeführten Aktivitäten der Selbstcharakterisierung der Wikipedia sehr nahe kommen, in der es heißt, dass sie „ihren Betreibern stupide, langweilige und häufig auftretende Aufgaben abnehmen“. Aus rein inhaltlichen Gesichtspunkten betrachtet, kann Bots nur eine untergeordnete Bedeutung zugemessen werden, wenngleich erste Tendenzen einer intelligenteren Bot-Generation erkennbar sind.

⁴Koordinaten für französischen Gemeinden wurden externen, im Web verfügbaren, Tabellen entnommen.

3 Bots und Hybridität

Nach dem (etwas enttäuschenden) Ergebnis der Untersuchung ihres Editierverhaltens stellt sich die Frage, ob Bots tatsächlich nicht weiter beachtenswert sind, da die einzeln betrachteten Tätigkeiten offenbar keine größere inhaltliche Bedeutung und ihre Editierungen im Wesentlichen „lektorativen“ Charakter haben. In den folgenden Beispielen wird sich allerdings zeigen, dass Bots hochgradig handlungswirksam sein können. Zur Verdeutlichung ihrer Handlungswirksamkeit genügt es nicht, die Aktivitäten der Bots isoliert zu betrachten. Vielmehr wird ihre Wirksamkeit deutlich, wenn die Einzelaktivitäten im Kontext ihrer zugehörigen Interaktionsketten betrachtet werden.

3.1 Beispiel I: Bots als Initiatoren hybrider Interaktionsketten

In Abbildung 6 ist die Versionshistorie für eine Grafikdatei, die in der Wikipedia eingestellt ist, dargestellt. Für jeden Wikipedia-Artikel und jede Datei (z. B. eine Grafik) existiert eine solche Historie, in der sämtliche Bearbeitungsschritte chronologisch aufgeführt sind. Die Protokollierung umfasst neben der genauen Änderungsuhrzeit auch den Benutzernamen desjenigen, der die Änderung durchgeführt hat, die Größe der Änderung, gemessen in Bytes und einen optionalen Kommentar. Mit einer Vergleichsfunktion ist es auch möglich, Änderungen zwischen zwei Artikel- oder Dateiversionen anzuzeigen.

In dem hier diskutierten Beispiel geht es um die Klärung einer Lizenzfrage bzgl. eines eingescannten 10.000 Dong-Geldscheins (vietnamesische Währung). In Tabelle 5 sind die Änderungen der Nutzer ausformuliert, so dass deutlich wird, wie die hier betrachtete Interaktion zustande kommt.

15:41, 21. Jun. 2009	Sa-se (Diskussion Beiträge)	K (153 Bytes)	(Informationsmängel beseitigt)	(entfernen)	[automatisch gesichtet]
00:20, 18. Nov. 2007	Augiasstallputzer (Diskussion Beiträge)	K (156 Bytes)	(entfernen)	[gesichtet von Bondom]	
01:45, 26. Jul. 2007	Forrester-Bot (Diskussion Beiträge)	K (138 Bytes)	(ersetze Vorlage: Bild-UN durch Vorlage: Bild-frei)	(entfernen)	
10:47, 23. Jul. 2007	ChristianBier (Diskussion Beiträge)	(136 Bytes)	(geldscheine sind gemeinfrei)	(entfernen)	
21:12, 16. Jul. 2007	ChristianBier (Diskussion Beiträge)	(155 Bytes)	(entfernen)		
23:31, 5. Jul. 2007	BuschBot (Diskussion Beiträge)	K (148 Bytes)	(Bot: Füge Vorlage:Dateüberprüfung ein. Bei Fragen: WP:DÜP.)	(entfernen)	
16:46, 3. Feb. 2007	Ar-ras (Diskussion Beiträge)	K (125 Bytes)	(entfernen)		
20:50, 3. Aug. 2006	Sebbot (Diskussion Beiträge)	K (177 Bytes)	(clean up)	(entfernen)	
20:38, 19. Dez. 2005	Crux (Diskussion Beiträge)	K (129 Bytes)	{{{Bild-PD}}} (Scannen stellt keine Schöpfungshöhe dar)	(entfernen)	
14:07, 13. Dez. 2005	Senegur (Diskussion Beiträge)	(131 Bytes)	(entfernen)		
13:59, 13. Dez. 2005	Senegur (Diskussion Beiträge)	(45 Bytes)	(Eigenes Digitalfoto 10000 Đồng Rückseite)		

Abbildung 6: Versionshistorie eines Bildes: Die Einzelaktivitäten von Bots und menschlichen Nutzern sind aufeinander bezogen

Bearbeitungszeitpunkt	Benutzer	Bot?	Beschreibung der Bearbeitung
13:59, 13. Dez 2005	Senegur	✗	Benutzer stellt erste Version eines eingescannten Geldscheins ein.
14:07, 13. Dez 2005	Senegur	✗	Die Lizenz des Geldscheins wird auf „GFDL“ (Copyleft-Lizenz mit Pflicht zur Nennung des Autors) eingestellt.
20:38, 19. Dez. 2005	Crux	✗	Lizenzinformation wird auf 'Public Domain' (gemeinfrei) geändert, da Benutzer der Auffassung ist, dass Einscannen nicht dazu berechtigt eigene Rechte an der Grafik anzumelden, wie z. B. die Nennung von Benutzer Senegur als Schöpfer.
20:50, 3. Aug. 2006	Sebbot	✓	Anpassen der Bildbeschreibung an neues Format.
16:45, 3. Feb. 2007	Ar-ras	✗	Entfernen der Lizenzinformation, da Benutzer nicht die Meinung teilt, dass es sich um eine gemeinfreie Grafik handelt.
23:31, 5. Jul. 2007	BuschBot	✓	Benutzer stellt fest, dass wichtige Informationen zu der Grafik (u.a. die Lizenz) fehlen und kennzeichnet die Grafik zur Überprüfung.
21:12, 16. Jul. 2007	ChristianBier	✗	Benutzer überprüft die Grafik und stellt fest, dass das Problem eine fehlende Lizenzinformation ist.
10:47, 23. Jul. 2007	ChristianBier	✗	Nach einigen Tagen findet der Benutzer heraus, dass Geldscheine gemeinfrei sind und kennzeichnet die Grafik entsprechend.
01:45, 26. Jul. 2007	Forrester-Bot	✓	Einfache Vorlagenersetzung für gemeinfreie Bilder in der deutschsprachigen Wikipedia.

Tabelle 5: Kommentierte Versionshistorie, die die Interaktion zwischen Bots und menschlichen Nutzern illustriert

Entscheidend für die Interaktion zwischen Bots und menschlichen Nutzern ist die Editierung des Benutzers *Ar-ras*, der die Lizenzinformation zu der Grafik entfernt, da er diese nicht für gemeinfrei hält. *Ar-ras* unterbreitet allerdings keinen Gegenvorschlag. Auf die fehlende Lizenzinformation wird der Bot *BuschBot* aufmerksam und kennzeichnet die Grafik als zu überprüfend. Dieser Aufforderung kommt der Benutzer *ChristianBier* wenige Tage später nach und präzisiert das Problem, das der Bot bezüglich der Grafik festgestellt hat. Es handelt sich offenbar um ein Lizenzproblem, das der Benutzer *ChristianBier* nicht direkt klären kann. Einige Tage später stellt er fest, dass Geldscheine gemeinfrei sind und kennzeichnet die betroffene Datei entsprechend. Der Bot *Forrester-Bot* ersetzt wenige Tage später eine vom Benutzer *ChristianBier* verwendete englischsprachige Vorlage zur Grafikbeschreibung durch eine deutschsprachige Version für gemeinfreie Bilder.

Insgesamt führte die Interaktion zwischen verschiedenen Bots und den menschlichen Nutzern schließlich dazu, dass ein Lizenzproblem, nämlich die Frage, unter welcher Lizenz eingescannte Geldscheine stehen, geklärt wurde. Beachtenswert ist, dass ohne die Kennzeichnung durch

den Bot *BuschBot*, die auf ein Problem mit der Grafik hindeutete, der Benutzer *ChristianBier* höchstwahrscheinlich nicht darauf aufmerksam geworden wäre und die Lizenzfrage ungeklärt geblieben wäre. Die folgende Editierung durch den Bot *Forrester-Bot* wäre zudem nicht ohne die vorherige Editierung des Benutzers *ChristianBier* erfolgt.

3.2 Beispiel II: Hybride Konzeption eines Bots

Das vorherige Beispiel hat gezeigt, dass Bots Initiatoren hybrider Interaktionsketten werden können. Das folgende Beispiel geht noch einen Schritt weiter und verdeutlicht, dass Bots z. T. bereits *hybrid konzipiert* sind.

Der sogenannte Interwiki Link Checker (IWLC)¹ ist dabei ein Bot (genauer: ein aus mehreren Bots bestehendes Tool), der auf die „Mitarbeit“ menschlicher Nutzer angewiesen ist. Wie bereits in Abschnitt 2.3 angesprochen, ist die Verlinkung von Artikeln unterschiedlicher Sprachversionen ein wichtiges Anliegen der Wikipedia. Insbesondere für Sprachen, die auf Grund ihrer geringen Verbreitung nicht auf eine derart große (potenzielle) Autorenschaft wie beispielsweise die englischsprachige Wikipedia zurückgreifen können, ist die Methode der Interwiki-Verlinkung ein probates Mittel, um zu einem Artikel in der eigenen Sprache weiterführende Informationen zu erhalten.

Die Oberfläche des IWLC ist in Abbildung 7 dargestellt. Unter Zuhilfenahme von Bots und des Betreibers des IWLC werden Listen von Artikeln generiert, die jeweils in zwei unterschiedlichen Sprachversionen ein vermeintlich identisches Thema behandeln.

Susie

出典：フリー百科事典『ウィキペディア (Wikipedia)』
 移動：ナビゲーション, 検索

Susie	
作者	竹村嘉人 (たけちん)
初版	0.01 (1994年5月25日)
最新版	0.47b (2002年10月20日)
最新評価版	0.50 beta2a (2010年3月7日)
対応OS	Microsoft Windows

Susie

Fra Wikipedia, den frie encyklopedi
 Gå til: [navigasjon](#), [søk](#)

Susie er et
 kvinnenavn og en variant av [Susanne](#)

Susie	
Betydning:	av hebraisk, "Lilje"
Navnedag:	Ingen (Norge) Ingen (Sverige)
Varianter og popularitet i Norge (2005)	
Susie:	< 200 kvinner

Last Change	Statistics	At least one of the pages is ...	Both pages deal with the same subject	Change databases
same	not the same	4 ... a redirect	1 Yes	3 Don't know
for bot-operator	6 ... a disambiguation		2 No	7 Article(s) deleted
				Cancel

Abbildung 7: Der Interwiki Link Checker (IWLC): Beispiel für die hybride Konzeption eines Bots. Menschliche Nutzer werden zum „Mit-Handeln“ ermutigt.

¹<http://toolserver.org/~flacus/IWLC/>

Im hier präsentierten Beispiel werden einem menschlichen Nutzer zwei Artikel mit dem Titel *Susie* auf einer Webseite nebeneinander dargestellt. Der Interwiki Link Checker ruft die thematisch potenziell identischen Artikel selbstständig ab und präsentiert sie einem menschlichen Nutzer, der beide Artikelsprachen (hier: Japanisch und Norwegisch) beherrscht. Dieser kann leicht beurteilen, ob die Artikel ein identisches Thema behandeln. Mit einem Mausklick auf die vom IWLC bereitgestellten Buttons beantwortet er diese Frage mit 'Yes', 'No', 'Don't know' oder 'Article(s) deleted'. Für einen Bot wäre es hier nur schwer zu entscheiden, ob die Artikel das gleiche Thema behandeln. Aus dem Grund unterstützt der IWLC menschliche Nutzer bei der Klassifizierungsaufgabe durch das Abrufen der Artikel und eine möglichst übersichtliche Gegenüberstellung.

Nachdem die Artikel von sprachkundigen Nutzern entsprechend klassifiziert wurden, kümmert sich täglich ein Bot um das Einpflegen der Verlinkung. Dabei verlässt sich der Bot nicht auf die Entscheidung eines einzelnen Nutzers, sondern verlinkt zwei Artikel erst dann miteinander, wenn mindestens fünf Benutzer diese als identisch klassifiziert haben.

3.3 Plädoyer für eine hybride Sicht auf Bots

Anhand des ersten Beispiels haben wir gesehen, dass zwar die isoliert betrachtete Aktivität eines Bots eine relativ geringe Bedeutung hat, diese jedoch Folgehandlungen von menschlichen Nutzern initiiert, die ihrerseits weitere Folgehandlungen hervorrufen, entweder von Bots oder wiederum von menschlichen Nutzern. Wir möchten im Falle dieser aufeinanderfolgenden Handlungsstränge von hybriden Interaktionsketten sprechen. Hybrid deswegen, weil sich die Gesamthandlung, zum Beispiel die Lösung eines Copyright-Problems in einem „Handlungsstrom“ (Rammert u. Schulz-Schaeffer 2002) vollzieht, in dem die einzelnen Aktivitäten auf menschliche und nicht-menschliche Nutzer verteilt sind und die Handlungen aufeinander bezogen sind.

Im zweiten Beispiel haben wir gesehen, dass der Interwiki Link Checker stellvertretend für einen Bot-Typus steht, der bereits hybrid konzipiert ist und auf die Mitarbeit von menschlichen Nutzern angewiesen ist bzw. sie zur Mitarbeit ermutigt.

Wie in der Einleitung schon angesprochen, lassen sich derartige hybride Zusammenhänge theoretisch zum Beispiel mit dem graduellen Modell verteilter Handlungsträgerschaft von Rammert u. Schulz-Schaeffer (2002) fassen. Die Dortmunder Techniksoziologen arbeiten derzeit an einem Modell, das von Rammert u. Schulz-Schaeffer inspiriert ist und versucht, den Aspekt der Hybridität in das Esser'sche Modell soziologischer Erklärung zu integrieren (Fink u. Weyer 2011). Zur Anwendung des Modells ist es erforderlich, die situativen Bedingungen der beteiligten menschlichen und nicht-menschlichen Akteure zu modellieren und ihre subjektiven Intentionen herauszufinden. Im Rahmen der vorliegenden Forschungsarbeit haben wir auf die Übertragung des Modells auf den Fall Wikipedia-Bots verzichtet, sehen aber perspektivisch Anwendungsmöglichkeiten.

4 Weisheit der Wenigen

Das Interaktionsbeispiel hat gezeigt, dass bestimmte Akteure innerhalb der Wikipedia, ganz im Sinne von Stegbauer, bestimmte Positionen einnehmen und diese durch ihre Tätigkeiten ausfüllen. Stegbauer hat auch gezeigt, und dies trifft auf Bots schon wegen ihrer zumeist großen Anzahl von Editierungen zu, dass sich um gewisse Benutzer Machtzentren herausbilden (müssen). Bots, ebenso wie besonders aktive menschliche Nutzer, tragen gemessen an ihrer geringen Anzahl überproportional viel zur Wikipedia bei.

Eine insbesondere in der amerikanischen Literatur vielfach diskutierte These ist die Frage nach dem Potenzial der 'wisdom of crowds' (Surowiecki 2004), der Weisheit der Vielen. Im Kern geht es um die Frage, ob es in einem System, bestehend aus sehr vielen, dezentralen, mittelmäßig gut informierten und unabhängigen Individualakteuren, möglich ist, in der Aggregation eine Entscheidung herbeizuführen, die hinsichtlich ihrer Qualität das Urteil einiger weniger Experten übertreffen kann. Surowiecki beantwortet diese Frage eindeutig mit einem Ja, sofern in einem System ein funktionierender Aggregationsmechanismus angewendet wird, der dezentrale Entscheidungen in „intelligenter“ Weise zusammenführt.

Ausgehend von den Fragen nach Aktivitätsungleichheit und der 'wisdom of crowds' drängt es sich geradezu auf, diese Fragen, zumindest in Teilen, anhand des vorliegenden Datenmaterials zu diskutieren. Bereits Stegbauer (2009: 76) nahm die Ungleichheit der Beiträge zum Anlass, die (positionale) Struktur in der Wikipedia vornehmlich netzwerkanalytisch näher zu untersuchen. Wir betrachten im Folgenden nicht die Struktur, sondern gehen noch einmal einen Schritt zurück und untersuchen im Detail, wie ungleich die Beiträge verteilt sind. Dabei berücksichtigen wir durch ein wochenbasiertes Zeitfenster insbesondere auch die anonymen Nutzer, die erstens einen wichtigen Bestandteil der 'wisdom of crowds' darstellen und zweitens in bisherigen Untersuchungen weitgehend unberücksichtigt blieben.

4.1 Exkurs: Kontroverse zur 'wisdom of crowds' in der Wikipedia

Die Thesen von Surowiecki sind nicht unumstritten und wurden insbesondere von Andrew Keen (2007) scharf kritisiert, der Surowiecki und allgemein dem gesellschaftlich-medialen Diskurs ein 'cult of the amateur' vorwirft.

In Bezug auf die Wikipedia wird die Frage nach der 'wisdom of crowds' ebenfalls höchst kontrovers diskutiert. Kittur u. a. (2007) sind auf Grundlage ihrer Untersuchungen der englischsprachigen Wikipedia zu dem Ergebnis gekommen, dass normale Nutzer mit wenigen Editierungen

die tragenden Säulen der Wikipedia seien: „The results suggest that although Wikipedia was driven by the influence of 'elite' users early on, more recently there has been a dramatic shift in workload to the 'common user'“. Zu einem ähnlichen Ergebnis kommen Ortega u. Gonzalez Barahona (2007): Den größten Beitrag zur Wikipedia leisten Autoren mit einer niedrigen Aktivität, wobei Aktivität auf Grundlage der Anzahl der Editierungen gemessen wird.

Umso bemerkenswerter ist es, dass die beiden Forscher nur ein Jahr später im Rahmen einer Forschergruppe (Ortega u. a. 2008), an der sie selbst beteiligt sind, zum exakt gegenteiligen Ergebnis kommen. Sie argumentieren, dass sich in den untersuchten Sprachversionen eine große Ungleichheit zwischen Beitragsanteil und Autorenkreis zeigt: 10% aller Autoren sind für 90% der Inhalte in der Wikipedia verantwortlich. Eine derartige Ungleichheit konnten sie auch in einer historischen Perspektive nachweisen.

Ein weiteres Jahr später fällt das Urteil einer Forschergruppe um Javanmardi u. a. (2009), der es um den Zusammenhang zwischen registrierten und anonymen Nutzern geht, noch eindeutiger gegen die 'wisdom of crowds' aus: „Although anonymous users comprise 83% of users in Wikipedia, their submission ratio is only 33%“.

Vor dem Hintergrund der uneindeutigen Literaturlage und des allgemeinen (medialen) Hypes um die Weisheit der Vielen bietet es sich an, die Fragestellung nach den tragenden Säulen auf Grundlage der uns zur Verfügung stehenden Daten, zumindest für die deutsche Sprachversion der Wikipedia, näher zu untersuchen.

4.2 Entwicklung von Editieranteilen

Die wochenbasierte Betrachtung von Editieranteilen gibt einen Hinweis auf die zunehmende Ungleichheit. Die 'wisdom of crowds', die insbesondere durch die unregistrierten Benutzer repräsentiert wird, trägt, gemessen an der Anzahl ihrer Editierungen, im Verhältnis zu den registrierten Benutzern und Bots immer weniger bei. Bereits anhand von Abbildung 8 kann man sich fragen, ob spätestens seit 2008 überhaupt noch von einer Weisheit der Vielen gesprochen werden kann oder ob sich die Wikipedia, zusammen mit ihren Power-Usern und Bots, zu einem elitären Wiki einer Minderheit entwickelt.

4.3 Ungleichverteilung der Editierungen

Ein Maß, Ungleichheit genauer zu angeben, ist der Gini-Koeffizient. Er ist 0 bei völliger Gleichverteilung und 1 bei völliger Ungleichverteilung. Ungleichheit in Bezug auf die Anzahl Editierungen wurde bereits von Stegbauer (2009) betrachtet, allerdings ohne diese mittels Gini-Koeffizient näher zu quantifizieren. Eine ähnliche Fragestellung findet sich auch bei Ortega u. a. (2008). Dort werden bereits Lorenz-Kurven benutzt, allerdings wird nur monatsbasiert gescannt, mit der Folge, dass anonyme Nutzer nicht berücksichtigt werden können.

Um die besonders spannende Frage nach dem Beitrag der 'crowd' zu beantworten, ist es wichtig,

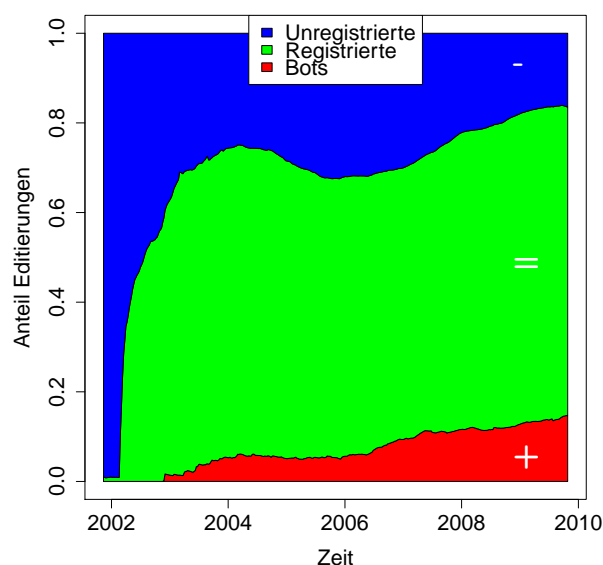
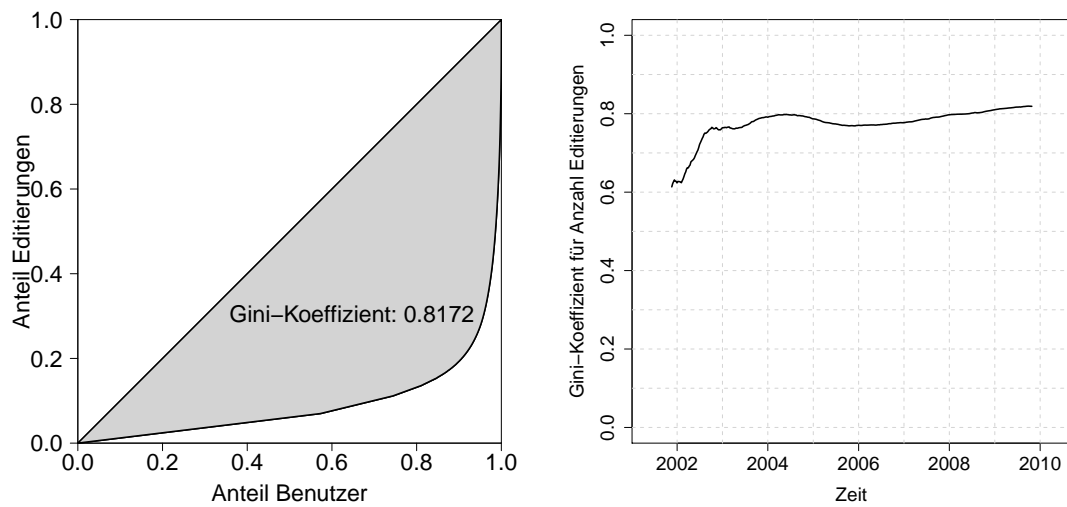


Abbildung 8: Zeitliche Entwicklung von Editieranteilen (geglättete Darstellung, Fensterbreite: 53 Wochen): Die zeitliche Entwicklung von Editieranteilen deutet auf ein zunehmendes Maß an Ungleichheit hin.

unregistrierte Nutzer mit zu berücksichtigen. Durch einen wochenbasierten Scan ist die Annahme berechtigt und macht zumindest nur geringe Fehler, eine IP-Adresse als einen Nutzer anzusehen. Betrachtet man die letzte Woche des Untersuchungszeitraums (17. KW 2010), die hier beispielhaft herausgegriffen ist, zeigt sich anhand der Lorenz-Kurve¹ zur Anzahl der Editierungen in Abbildung 9a eine große Ungleichheit: 80% der Nutzer haben weniger als 20% der Editierungen beigetragen. Für diese Woche ergibt sich ein Gini-Koeffizient² von 0.82, der von einer großen Ungleichheit zeugt. Die für diese Woche dargestellte Lorenzkurve hat auch für die anderen Wochen des Untersuchungszeitraums eine sehr ähnliche Form. Sie dient hier vor allem der Illustration, um die Darstellung der zeitlichen Entwicklung des Gini-Koeffizienten besser zu verstehen: In einer längeren zeitlichen Perspektive (Abbildung 9b) zeigt sich, dass eine derartige Ungleichheit, außer ggf. in der Anfangszeit der Wikipedia, was allerdings auch mit unvollständigen Daten zu tun haben kann, schon immer existiert hat. Wenn überhaupt kann nur von einer geringen Zunahme von Ungleichheit seit 2006 gesprochen werden. Das Ergebnis deckt sich mit dem von Ortega u. a. (2008). Dort wurde ebenfalls mit Hilfe des Gini-Koeffizienten nachgewiesen, dass 90% der Inhalte von nur 10% der Nutzer produziert werden. Vor dem Hintergrund

¹Die Lorenz-Kurve ist ein grafisches Hilfsmittel um die Ungleichheit der Verteilung eines Beobachtungsmerkmals darzustellen. Je weiter sich die Lorenz-Kurve von der eingezeichneten Winkelhalbierenden entfernt, desto ungleicher verteilt sich die Summe aller Beobachtungswerte auf die Beobachtungen.

²Der Gini-Koeffizient ist ein Maß für die Ungleichheit der Verteilung eines Beobachtungsmerkmals und kann Werte zwischen Null und Eins annehmen. Bei einem Wert von Null ist die Summe aller Beobachtungswerte gleichmäßig auf die Beobachtungen verteilt, es herrscht größtmögliche Gleichheit. Bei einem Wert von Eins vereinigt nur eine Beobachtung die gesamte Summe aller Beobachtungswerte auf sich, es herrscht größtmögliche Ungleichheit. Der Gini-Koeffizient kann grafisch als Flächeninhalt der grau eingefärbten Fläche zwischen Lorenz-Kurve und Winkelhalbierender abgelesen werden.



(a) *Lorenz-Kurve für die letzte betrachtete Woche (17. KW 2010)* (b) *Gini-Koeffizient über die Zeit (geglättete Darstellung, Fensterbreite: 53 Wochen)*

Abbildung 9: Ungleichverteilung der Editierungen. (a) zeigt anhand der Lorenz-Kurve und des Gini-Koeffizienten ein hohes Maß an Ungleichheit für die letzte betrachtete Woche, (b) zeigt, dass ein ähnlich hohes Maß an Ungleichheit bereits seit 2004 besteht.

der Auswertung der 'Contributions' sowohl bezogen auf die letzte betrachtete Woche als auch auf die Berechnung der Ungleichheit über einen längeren Zeitraum, erscheint die These von einer Weisheit der Vielen nicht haltbar. Bots sind hierbei ein besonders deutlicher Ausdruck der festgestellten Ungleichheit.

5 Fazit und Forschungsperspektiven

Wir haben gesehen, dass Bots einen erheblichen Anteil am Gesamteditervolumen der deutschsprachigen Wikipedia ausmachen und dass ihre Aktivität zunimmt. Ausgehend von den 10% aller Editierungen, die von Bots durchgeführt werden, dürfen Bots in Studien zur Wikipedia keinesfalls unberücksichtigt bleiben oder ausgefiltert werden. Somit lautet die erste Feststellung: Bots sind wichtig! Bei einer näheren Betrachtung der Bot-Aktivitäten stellte sich heraus, dass diese zwar in allen Bereichen der Wikipedia aktiv sind, eine isolierte Betrachtung ihrer Änderungen zu der (ernüchternden) Erkenntnis führt, dass diese Editierungen allenfalls „lektorativen“ Charakter haben. Davon, dass Bots autonom Artikel z. B. auf Grundlage selbst beschaffter, externer Daten verfassen, sind wir (noch) weit entfernt. Ein drittes Resultat ist, dass Bots durch ihre Aktivitäten als Initiatoren hybrider Interaktionsketten auftreten und somit in höchstem Maße handlungswirksam sind. Teilweise sind Bots bereits hybrid konzipiert. Ihr Mangel an der Fähigkeit, inhaltliche Urteile selbstständig zu fällen, wird dadurch kompensiert, dass sie vorbereitende Arbeiten durchführen und ein Gesamturteil letztlich per Mehrheitsbeschluss von menschlichen Nutzern gefällt wird.

Als viertes Resultat zeigte sich eine generelle Tendenz zunehmender Ungleichheit, die insbesondere die gängige These der Weisheit der Vielen widerlegt. Bots sehen wir dabei als einen weiteren Aspekt zur Widerlegung dieser These an, da ihre Anzahl aktuell nicht wesentlich wächst, sie aber, ebenso wie weitere 'Power User', immer mehr zur Wikipedia beitragen.

Weitere Forschungsperspektiven

Bei unserer Untersuchung hat sich gezeigt, dass *einzelne* Bots (soziologisch) nicht besonders beachtenswert sind. Sie werden aber dann interessant, wenn man sie im Zusammenhang hybrider Interaktionsketten betrachtet. Dabei sehen wir insbesondere weiteren Forschungsbedarf in der Analyse dieser Interaktionsketten. Eine derartige Analyse könnte sich an den vornehmlich eher quantitativ orientierten netzwerkanalytischen Methoden von Stegbauer orientieren. In dem Zusammenhang wären Fragen nach der positionalen Struktur und nach Machtzentren in der Wikipedia unter Berücksichtigung von Bots von besonderer Relevanz. Ein eher qualitativ orientierter Ansatz, z. B. die Durchführung einer Konversationsanalyse, könnte einzelne hybride Interaktionsketten in den Mittelpunkt rücken und dabei charakteristische Hybriditätsaspekte herausstellen.

Eine weitere Möglichkeit zur Analyse derartiger Interaktionsketten bietet das in Dortmund entwickelte Modell soziologischer Erklärung hybrider Systeme (Fink u. Weyer 2011). In Interviews wäre dann zu klären, wie Bot-Betreiber und -Programmierer sowie menschliche Interaktionspartner die Zusammenarbeit oder das Zusammenhandeln mit Bots beurteilen. Als mögliches Ergebnis wäre denkbar, dass menschliche Interaktionspartner Bots ganz andere Intentionen zuschreiben als ursprünglich vom Bot-Programmierer vorgesehen. Dies könnte zu Konflikten und Missverständnissen führen.

Des Weiteren bietet es sich an, die Analyse auf unterschiedliche Sprachversionen auszuweiten. Für eine vergleichende Analyse erscheinen insbesondere die Bots interessant, welche sich nicht um die Verknüpfung von Wikipedia-Artikeln in verschiedenen Sprachversionen bemühen (Interwiki-Bots), da diese ohnehin „international orientiert“ sind, sondern vielmehr die Bots, die andere Aufgaben erfüllen. Zu erwarten wären möglicherweise länderspezifische Praktiken, Wissen zu strukturieren oder Aufgaben zu automatisieren. Ferner wäre die Frage interessant, ob es länderspezifische Unterschiede hinsichtlich des Vertrauens in automome Technik gibt, z.B. eine besondere Skepsis oder eine besondere Offenheit gegenüber nicht-menschlichen Mitgliedern der Wikipedia-Gemeinschaft.

Literaturverzeichnis

Anthony u. a. 2005

ANTHONY, D. ; SMITH, S. W. ; WILLIAMSON, T.: Explaining quality in Internet collective goods: Zealots and good Samaritans in the case of Wikipedia. In: *Hanover: Dartmouth College* (2005)

Braun-Thürmann 2007

BRAUN-THÜRMANN, Holger: Künstliche Interaktion. In: RAMMERT, Werner (Hrsg.): *Technik - Handeln - Wissen*. Wiesbaden : VS Verl. für Sozialwiss., 2007, S. 221–243

Fink u. Weyer 2011

FINK, Robin D. ; WEYER, Johannes: Autonome technische Systeme als Herausforderung der soziologischen Handlungstheorie: in Ersch. In: *Zeitschrift für Soziologie* 40 (2011), Nr. 2

Javanmardi u. a. 2009

JAVANMARDI, S. ; GANJISAFFAR, Y. ; LOPES, C. ; BALDI, P.: User contribution and trust in Wikipedia. In: *Collaborative Computing: Networking, Applications and Worksharing, 2009. CollaborateCom 2009. 5th International Conference on Collaborative Computing*, 2009, 1-6

Keen 2007

KEEN, Andrew: *How blogs, MySpace, YouTube, and the rest of today's user-generated media are destroying our economy, our culture, and our values*. Reprint. Random House Inc., 2007

Kittur u. a. 2007

KITTUR, Aniket ; CHI, Ed ; PENDLETON, Bryan A. ; SUH, Bongwon ; MYTKOWICZ, Todd: Power of the Few vs. Wisdom of the Crowd: Wikipedia and the Rise of the Bourgeoisie. In: *Proceedings of the 25th Annual ACM Conference on Human Factors in Computing Systems (CHI 2007)*. 2007

Latour 1998

LATOUR, Bruno: Über technische Vermittlung. Philosophie. Soziologie, Genealogie. In: RAMMERT, Werner (Hrsg.): *Technik und Sozialtheorie*. Campus Verlag, 1998, S. 29–81

Ortega u. a. 2008

ORTEGA, F. ; GONZALEZ BARAHONA, J. M. ; ROBLES, G.: On the Inequality of Contributions to Wikipedia. In: *Hawaii International Conference on System Sciences, Proceedings of the 41st Annual*, 2008, 1-7

Ortega u. Gonzalez Barahona 2007

ORTEGA, Felipe ; GONZALEZ BARAHONA, Jesus M.: Quantitative analysis of the wikipedia community of users. In: *Proceedings of the 2007 international symposium on Wikis*. Montreal, Quebec, Canada : ACM, 2007, S. 75–86

Rammert u. Schulz-Schaeffer 2002

RAMMERT, Werner (Hrsg.) ; SCHULZ-SCHAEFFER, Ingo (Hrsg.): *Können Maschinen handeln? Soziologische Beiträge zum Verhältnis von Mensch und Technik*. Frankfurt/M. : Campus, 2002

Spiegel Online 2010

SPIEGEL ONLINE: *Tippfehler verursachte möglicherweise Wall-Street-Chaos*. www.spiegel.de/wirtschaft/unternehmen/0,1518,693495,00.html. Version: 2010

Stegbauer 2006

STEGBAUER, Christian: Wikipedia: Die Erstellung einer Online-Enzyklopädie als Herausforderung für die Erklärung von Kooperation. In: STEGBAUER, Christian (Hrsg.) ; RAUSCH, Alexander (Hrsg.): *Strukturalistische Internetforschung*. Wiesbaden : VS Verl. für Sozialwiss., 2006, S. 221–244

Stegbauer 2009

STEGBAUER, Christian: *Netzwerkforschung*. Bd. 2: *Wikipedia*. 1. Aufl. Wiesbaden : VS Verl. für Sozialwiss., 2009

Surowiecki 2004

SUROWIECKI, James: *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies, and nations*. 1st ed. New York : Doubleday, 2004

Voss 2005

VOSS, Jakob: Measuring Wikipedia. In: *Proceeding of the ISSI 2005*. 2005, S. 221–231

Bereits erschienene Soziologische Arbeitspapiere

- | | |
|---------|--|
| 1/2003 | Hartmut Hirsch-Kreinsen, David Jacobsen, Staffan Laestadius, Keith Smith
Low-Tech Industries and the Knowledge Economy: State of the Art and Research Challenges
(August 2003) |
| 2/2004 | Hartmut Hirsch-Kreinsen
"Low-Technology": Ein innovationspolitisch vergessener Sektor
(Februar 2004) |
| 3/2004 | Johannes Weyer
Innovationen fördern – aber wie? Zur Rolle des Staates in der Innovationspolitik
(März 2004) |
| 4/2004 | Konstanze Senge
Der Fall Wal-Mart: Institutionelle Grenzen ökonomischer Globalisierung
(Juli 2004) |
| 5/2004 | Tabea Bromberg
New Forms of Company Co-operation and Effects on Industrial Relations
(Juli 2004) |
| 6/2004 | Gerd Bender
Innovation in Low-tech – Considerations based on a few case studies in eleven European countries
(September 2004) |
| 7/2004 | Johannes Weyer
Creating Order in Hybrid Systems. Reflexions on the Interaction of Man and Smart Machines
(Oktober 2004) |
| 8/2004 | Hartmut Hirsch-Kreinsen
Koordination und Rationalität
(Oktober 2004) |
| 9/2005 | Jörg Abel
Vom Kollektiv zum Individuum? Zum Verhältnis von Selbstvertretung und kollektiver Interessenvertretung in Neue Medien-Unternehmen (Juli 2005) |
| 10/2005 | Johannes Weyer
Die Raumfahrtspolitik des Bundesforschungsministeriums
(Oktober 2005) |
| 11/2005 | Horst Steg
Transnationalisierung nationaler Innovationssysteme
(Dezember 2005) |
| 12/2006 | Tobias Haertel
UsersAward: Ein Beitrag zur optimalen Gestaltung von Mensch-Maschine-Systemen in der Logistik
(Februar 2006) |
| 13/2006 | Doris Blutner, Stephan Cramer, Tobias Haertel
Der Mensch in der Logistik: Planer, Operateur und Problemlöser
(März 2006) |
| 14/2006 | Johannes Weyer
Die Zukunft des Autos – das Auto der Zukunft. Wird der Computer den Menschen ersetzen?
(März 2006) |

15/2006	Simone Reineke Boundary Spanner als Promotoren des Wissensmanagementprozesses (Juli 2006)
16/2006	Johannes Weyer Die Kooperation menschlicher Akteure und nicht-menschlicher Agenten. Ansatzpunkte einer Soziologie hybrider Systeme (Juli 2006)
17/2006	Jörg Abel/Sebastian Campagna/Hartmut Hirsch-Kreinsen (Hg.) Skalierbare Organisation - Überlegungen zum Ausgleich von Auftragsschwankungen - (August 2006)
18/2007	Tabea Bromberg Engineering-Dienstleistungen in der Automobilindustrie: Verbreitung, Kooperationsformen und arbeitspolitische Konsequenzen (Mai 2007)
19/2007	Hartmut Hirsch-Kreinsen Lohnarbeit (September 2007)
20/2008	Katrin Hahn Der Lissabon-Prozess: Das Innovationskonzept und die Auswirkungen auf die Politikgestaltung (März 2008)
21/2008	Anja J. Lorenz/ Johannes Weyer (Hrsg.) Fahrerassistenzsysteme und intelligente Verkehrssteuerung. Soziologische Analysen hoch automatisierter Verkehrssysteme (Juni 2008)
22/2008	Hartmut Hirsch-Kreinsen Innovationspolitik: Die Hightech-Obsession (August 2008)
23/2008	Hartmut Hirsch-Kreinsen Multinationale Unternehmen (September 2008)
24/2009	Jörg Abel/ Hartmut Hirsch-Kreinsen/ Peter Ittermann Einfacharbeit in der Industrie. Status quo und Entwicklungsperspektiven (Mai 2009)
25/2009	Robin D. Fink Attributionsprozesse in hybriden Systemen. Experimentelle Untersuchung des Zusammenspiels von Mensch und autonomer Technik (Juli 2009)
26/2009	Hartmut Hirsch-Kreinsen Innovative Arbeitspolitik im Maschinenbau? (September 2009)
27/2010	Hartmut Hirsch-Kreinsen Technological Innovation and Finance (Oktober 2010)