

Christof Niemann und Stephan Müller

## Ein Tag sagt mehr als tausend Worte? – Kreatives Potenzial und Neotags in Tagging-Systemen

### Einleitung

Im Kontext des Information Overload (Toffler, 1990) und begünstigt durch die hohe technische Entwicklungsdynamik des Internets (pointiert ausgedrückt nicht zuletzt in der Versionierungsmetapher „2.0“) können die Strukturierung der Informationsmassen sowie die inhaltliche Erschließung digitaler Medien als größte Herausforderungen des heutigen Wissensmanagements bezeichnet werden. Neben dem „Klassiker“, nämlich der Erschließung der Ressourcen durch Wissensexperten mittels kontrollierter Vokabulare, haben sich in der vernetzten Welt zwei weitere Methoden hierzu durchgesetzt: die automatische Erschließung mittels Suchalgorithmen, wie sie in Suchmaschinen und in der computerlinguistischen Methodik zur Anwendung kommen, sowie Community-basierte Methoden, bei denen die Teilnehmer von Netzwerken selbst Annotationen vornehmen, um diese für sich und andere zur Verfügung zu stellen. Jede dieser Methoden hat ihre spezifischen Vorzüge und Mängel: Kontrollierte Vokabulare bieten Sicherheit und Struktur, sind aber bezüglich neuer Entwicklungen und aufkommender Themen träge und statisch. Automatische Verfahren bestehen durch ihre hohe Effizienz und Verfügbarkeit, können aber ebenso wenig selbst Innovationen produzieren, wie sie Ambiguitäten und inhaltliche Besonderheiten blind ignorieren. Web 2.0-Systeme sind äußerst dynamisch und bezüglich der von ihnen abgedeckten Themen ebenfalls effizient, hängen bezüglich ihres Erschließungsgrades aber stark von den Interessen ihrer Nutzer ab und weisen teilweise eklatante Mängel im Hinblick auf ihre Strukturiertheit auf. Idealerweise müssten folglich alle drei Methoden kombiniert und/oder verknüpft werden, um dem genannten Overload effektiv zu begegnen. Die Erschließung von Inhalten würde dann z.B. wie im dargestellten „Annotationsdreieck“ (vgl. Abbildung 1) ablaufen.

Die Vokabulare dienen in dem Modell als Basis der automatischen Verschlagwortung, deren Ergebnisse zur Anreicherung der Tagging-Systeme benutzt werden. So können Ressourcen mit jenen relevanten Schlagworten versehen werden, die durch die Tagger noch nicht zugewiesen wurden. Die eigenen Tags der Nutzer, die aktuelle Bedürfnisse und Entwicklungen beim Wissensmanagement am besten repräsentieren, werden dagegen als Inspirationsquelle und zur Ermittlung inhaltlicher Tendenzen bei der Pflege der kontrollierten Vokabulare

herangezogen. Auf diese Weise schließt sich dann der Kreislauf der Annotation, wobei auch Bereinigungsmechanismen zur Entfernung veralteter Schlagworte usw. implementiert werden sollten. Konkrete Prozesse und „Schnittstellen“ einer solchen Zusammenarbeit bei der Erschließung von Ressourcen müssen allerdings immer in Abhängigkeit von den jeweiligen Domänen und Bedürfnissen entwickelt werden.



Abbildung 1: Annotationsdreieck zur Erschließung von Ressourcen

In diesem Beitrag liegt nun der Schwerpunkt auf der Web 2.0-Methode des Collaborative Tagging, das zu einer festen Größe des Wissensmanagements im Internet avanciert ist (Everitt und Mills, 2009). Auch die wissenschaftlichen Bibliotheken stehen vor der Aufgabe, die große Flut von Informationen in Gestalt digital publizierter Artikel und Werke möglichst inhaltlich erschlossen verfügbar zu machen. Entsprechend ist es nahe liegend, Tagging auch hier als Lösungsstrategie in Erwägung zu ziehen. Die Frage ist allerdings, ob die kollaborative Intelligenz der Nutzer für die Sacherschließung überhaupt eingesetzt werden kann. Während diese von der intuitiven und individuellen Wissensorganisation beim Tagging profitieren können, ist die große Freiheit bei der Vergabe von Deskriptoren für den Informationsexperten nämlich wie bereits angedeutet ein ambivalentes Phänomen: Kundennähe und kreatives Potenzial stehen der großen Menge völlig unkontrollierter Metainformationen gegenüber, deren inhaltliche Qualität und Aussagekraft nicht gesichert ist. Aus der Sicht stark strukturierter Wissensorganisation, wie sie etwa durch Fachreferenten und Bibliothekare realisiert wird, handelt es sich bei den zunächst unvermittelt nebeneinander stehenden Tags nämlich kurz gesagt um Chaos. Das in diesem Chaos aber auch sinnvolle Ideen und wertvolles Wissen als Gemeinschaftsprodukt erzeugt wird, kann mindestens durch den Erfolg der Sys-

teme, teilweise aber auch durch die bisherigen Forschungsergebnisse als bestätigt gelten.

## Tagging-Systeme als Forschungsgegenstand: Prioritäten und Fehlbestände

Während viel Aufwand darauf verwendet wurde und wird, die Strukturen und Beziehungen zwischen den Zeichenketten, die in Form von Tags in Tagging-Systemen vielfältige Funktionen erfüllen, aufzudecken, ist die Beschäftigung mit der tatsächlichen inhaltlichen Substanz dieser Zeichenketten deutlich unterrepräsentiert. Die Ursachen hierfür liegen zum einen in der „Sozialstruktur“ der interessierten Community, zum anderen aber auch in der unmöglich erscheinenden Aufgabe, jedes einzelne der hunderttausende von Tags intellektuell zu evaluieren. Dennoch ist eine valide inhaltliche Evaluierung von Tags ausschließlich intellektuell zu haben. Das einzelne Tag würde so nämlich nicht nur als leere Zeichenfolge betrachtet, die mit weiteren Zeichenfolgen mit einer mehr oder weniger großen Wahrscheinlichkeit in einem inhaltlich nicht näher definierten Zusammenhang steht. Sie würde vielmehr als Träger von de facto semantischem Gehalt betrachtet.

Eine Evaluierung von 73 wissenschaftlichen Studien und Beiträgen, die in den letzten zwei Jahren in namhaften Publikationen (z.B. *Lecture Notes in Computer Science*, *Journal of Information Science* oder *Library Review*) zum Thema erschienen sind, zeigt als immer wiederkehrende Themen: *a)* allgemeine Betrachtungen jeglicher Art, etwa zum Wissensmanagement oder zum generellen Potenzial von Tagging-Systemen, *b)* Gruppen, typische Strukturen und Verteilungen von Tags, Nutzern und/oder Ressourcen, *c)* Interaktionsbeziehungen zwischen Nutzern, Tagging-Systemen und/oder Umweltbedingungen, *d)* Analysen von Netzwerkstrukturen und/oder systemimmanenten Dynamiken sowie *e)* (eher selten ausschließlich, aber in fast allen Beiträgen als methodische Zielsetzung) die Verbesserung und Bereitstellung von Algorithmen und automatischen Verfahren. Die Ausdifferenzierung der Forschung zeigt sich außerdem in „Orchideenstudien“ wie etwa jene zur Charakterisierung von Angriffen auf Tagging-Systeme oder zu Reaktionen der Nutzer auf Veränderungen einer bestimmten Systemarchitektur. Erwartungsgemäß finden sich auch linguistisch orientierte Annäherungen, die sich etwa mit dem benutzten Vokabular, dem automatischen Auffinden und der Verknüpfung ähnlicher Begriffe oder der Anreicherung derselben mit Metainformationen beschäftigen.

## Defizite der Tagging-Forschung

Trotz aller Komplexität und Vielschichtigkeit der Beiträge werden fast immer sprachstatistische Methoden und/oder maschinelles Lernen (Clustering-Verfahren,

Ähnlichkeitsalgorithmen usw.) zur Analyse eingesetzt. Nur sehr vereinzelt und nachgeordnet finden sich intellektuelle Evaluierungen, noch seltener Befragungen zur Validierung der automatisch generierten Ergebnisse. Dies überrascht zunächst, da Tagging-Plattformen ja ganz wesentlich auf sprachlichen Ausdrücken basieren, die ausschließlich durch den Menschen valide interpretiert und hinsichtlich ihrer semantischen Qualität bestimmt werden können. Auf dem zweiten Blick zeigen die Schwerpunkte der untersuchten Studien jedoch schlicht die methodischen Präferenzen der interessierten Fachöffentlichkeit, deren Vertreter im Wesentlichen aus informatiknahen Disziplinen bzw. solchen mit entsprechenden Schnittstellen (etwa digitale Bibliotheksdienste, eLearning in der Pädagogik usw.) stammen. Auch ein erheblicher Teil der Nutzer selbst und fast alle „Poweruser“ sind Personen dieser Tätigkeitsfelder zuzuordnen, wenngleich der Trend zu mehr Interdisziplinarität im Zuge der allgemeinen Popularisierung der Systeme anhaltend ist. Die klassischen Geisteswissenschaften sind aber in jedem Fall deutlich unterrepräsentiert.

#### Folgerungen

Die Verengung der wissenschaftlichen Beschäftigung mit Tagging-Systemen auf die (unbestreitbar wichtigen) „technischen“ Aspekte bewirkt, dass die Diskussion der ebenso zentralen sozialen Dimension und auch typisch geisteswissenschaftliche Sichtweisen auf das Thema ausbleiben. Dies ist auch unter dem Eindruck vielfältiger Beiträge dieser Disziplinen zu Netzwerkthemen *allgemein* nicht nachvollziehbar (vgl. z.B. den Sammelband von Stegbauer, 2008). Darüber hinaus legen die vielfältigen (aufwändigen) Verfahren und Bestrebungen zur Nutzbarmachung von Tagging-Daten aber noch einen weiteren Schluss nahe: intelligent sind offenbar vor allem die Forscher, die im Nachhinein versuchen, das Chaos in den Tagging-Systemen zu ordnen und in den Griff zu bekommen. Das vermeintlich „intelligente“ Kollektiv hingegen - so scheint es - produziert zunächst einmal nur ziemlich wahllos Inhalte. Eine nachgeordnete Strukturierung dieser Inhalte fördert nur kurzfristige Übereinkünfte der Nutzer darüber zu Tage, welche Deskriptoren gerade „gut“ funktionieren, weil die Produktion von Ad-hoc-Innovationen ja stetig weitergeht. Dem Wunsch nach eindeutiger, idealer Weise hierarchischer Strukturierung steht also die Tatsache entgegen, dass Sprache als adaptives System abseits formaler Grammatiken ständigen Reorganisationen unterworfen und mithin niemals „fertig strukturiert“ oder zu strukturieren ist. Diese „semiotische Dynamik“, innerhalb derer eine Gruppe von Sprachagenten ihre Symbole ständig gegeneinander abgleicht, befördert den ständigen Wandel der Sprache (Steels, 2006). Insofern ist auch die Reorganisation der aus Sprache abgeleiteten Strukturen immer nur in Abhängigkeit von individuellen Zielsetzungen möglich. In diesem Beitrag wird allerdings nicht versucht, der Dynamik der Sprache über ihre Strukturierung durch feingliedrige automatische Justierungsmechanismen möglichst „dicht auf den Fersen“ zu bleiben. Vielmehr sollen die im Prozess der Dynamik erzeugten „Sprachprodukte“ auf ihre Eigenschaften hin geprüft und bewertet werden.

## Intelligenz und Kreativität im Web 2.0

Intelligenz bedeutet, „anschauliche sowie abstrakte Beziehungen zu erfassen, herzustellen und zu deuten und sich dadurch an neuartige Situationen anzupassen und sie gegebenenfalls durch problemlösendes Verhalten zu bewältigen“ (Hoffmann und Wolf, 2010).<sup>1</sup> Im Zusammenhang mit Web 2.0-Systemen wie Wikipedia, Flickr oder BibSonomy kommt nun immer wieder die Idee einer bestimmten Form von Intelligenz, nämlich der sogenannten „Schwarmintelligenz“, ins Gespräch. Es handelt sich dabei um ein Phänomen, welches zunächst aus der Naturbeobachtung von Staaten bildenden Insekten abgeleitet wurde. Relativ unintelligente Einzelwesen produzieren durch ihr kooperatives Verhalten eine emergente, übergreifende Ordnung, mittels derer sie zur Lösung komplexer Allokations- und Optimierungsprobleme befähigt werden. Dabei ist grundsätzlich keine klare hierarchische Struktur auszumachen. Für den Versuch, dieses Phänomen auch in den Informationsströmen des Webs wiederzuerkennen, finden sich ebenso Befürworter wie auch Gegner (vgl. etwa Lévy, 1997; Wehner, 2001; Buhse et al., 2008; Kreye, 2008 oder - besonders prägnant - der Interneterfinder und Visionär Jaron Lanier, Lanier, 2006). Einigkeit herrscht aber bezüglich der Ansicht, dass durch die „Nutzerschwärme“ unfassbare Mengen an Informationen bereitgestellt werden, die geordnet und priorisiert werden müssen.

### Intelligente Strukturen in Tagging-Systemen

Unserer Meinung nach können Web 2.0-Systeme in der Tat in bestimmter Hinsicht als Phänomen der Schwarmintelligenz betrachtet werden, weil in ihnen keine institutionellen oder wirtschaftlichen Interessen (abseits der Schaltung von Werbebannern, dem Anbieten spezieller Services o. ä.), mithin keine oder nur schwach ausgebildete hierarchische Strukturen die Produktion der Inhalte steuern. Wohl aber spielt der vom System vorgegebene *formale* Rahmen (z.B. allgemeine Zielsetzungen des Systems und entsprechende Funktionalitäten sowie der Grad an Strukturiertheit, den die Eingaben der Teilnehmer erfüllen müssen) und die absolute Anzahl der Eingaben und Teilnehmer insgesamt eine wichtige Rolle. Es wird also keine gänzlich eigene Ordnung aus sich selbst heraus produziert, sie muss vielmehr vorgegeben und stetig aktualisiert werden, um irgendetwas Sinnvolles entstehen zu lassen. Innerhalb der Ordnung kommt dann als zweite wichtige Dimension des Web 2.0 die Kreativität der Nutzer ins Spiel, ohne die keines der entsprechenden Systeme irgendeinen (Mehr-) Wert produzieren könnte. Ausschließlich die Nutzer liefern nämlich die „unfassbaren Mengen“ an Informationen und ein schöpferisches Vermögen, das sich in ihrem „Handeln oder Denken

---

1 Die umfassende und interdisziplinäre Diskussion darüber, was Intelligenz bzw. Kreativität ist oder sein könnte, kann an dieser Stelle nicht wiedergegeben werden. Die verwendete Definition beinhaltet jedoch alle zentralen Aspekte des Phänomens und kann insofern als allgemein anerkannt gelten.

realisiert und einerseits durch Neuartigkeit oder Originalität gekennzeichnet ist, andererseits aber auch einen sinnvollen und erkennbaren Bezug zur Lösung technischer, menschlicher oder sozialpolitischer Probleme aufweist“ (Hoffmann und Wolf, 2010). Die „Intelligenz“ eines Tagging-Systems hängt also insgesamt davon ab, wie intelligent es aufgebaut ist bzw. wie gut es das Verhalten seiner Nutzer kennt und voraussagen kann. Es muss für bestimmte Probleme einfach passende Mechanismen bereitstellen, die das System aus sich selbst heraus (d.h. der „Nutzerschwarm“ an sich) nicht hervorbringen kann.<sup>2</sup> Nur sehr begrenzt kann der Nutzerschwarm Konventionen, z.B. Präfixe oder Zeichenkombinationen für bestimmte Zwecke, entwickeln, die schließlich als allgemeine Sprachnorm eine Unzulänglichkeit des Tagging-Systems umgehen. Deshalb sind Letztere ohne konkretes Einwirken von Programmierern weitestgehend statisch und „dumm“, sie können sich selbst nicht weiterentwickeln und nur sehr begrenzt Probleme lösen. Hier liegt ein gewichtiger Grund, warum der Begriff Intelligenz zumindest im Zusammenhang mit Tagging-Systemen zu optimistisch ist. Die Plattformen stellen lediglich einen Rahmen zur Verfügung, der in Bezug auf seine Zielsetzung und aus der Perspektive seiner Nutzer klug oder weniger klug angelegt sein kann.

Wenn nun ein Tagging-System an sich nicht intelligent ist, so kann es dennoch - durchaus im Sinne der Schwarmintelligenz - Strukturen und Ordnungen innerhalb seiner festgelegten Systemumwelt emergieren lassen. Die Grundannahme dabei ist, dass häufig zusammen auftretende Konstellationen zwischen den Teilnehmern des Netzwerkes (Nutzer, Quelle, Tag, vgl. Abschnitt 2) eine inhaltliche Verbundenheit zwischen diesen implizieren (Wu und Zhou, 2009). Nutzer, die z.B. immer wieder germanistische Quellen taggen, können diesem Fachbereich zugeordnet werden. Quellen, die mit ähnlichen Tags versehen werden, sollten sich auch inhaltlich ähneln. Und die verschiedenen Tags eines Nutzers spannen ein inhaltliches Spektrum auf, das diesen bezüglich seiner Interessen qualifiziert. Diese Gruppen- und Clusterbildung emergiert in der Tat „organisch“ aus den Interaktionen des Netzwerkes selbst.<sup>3</sup> Die semiotische Dynamik (vgl. Abschnitt 1) sollte dabei in einer Menge von Tags die Dominanz von passenden und insofern „intelligenten“ Deskriptoren bewirken. Daraus ergibt sich aber zuallererst das Problem, was ein solches „kluges“ Tag überhaupt auszeichnet. Diesem Problem wird im Abschnitt 3 im Detail nachgegangen.

---

2 Steels nennt etwa ein Beispiel, bei dem die Suche nach dem Tag „nyc“ anstelle eines Bildes von New York ein Bild vom Hund eines Freundes, der in New York lebt, zeigt (Steels, 2006). Für solche Probleme müssen z.B. Möglichkeiten für nutzergenerierte Kategorien o.ä. im Tagging-System geschaffen werden, um eine inhaltliche Administration der gespeicherten Informationen (sprich: einer Zensur) zu vermeiden. Diese würde fundamental der Philosophie des Social Taggings widersprechen.

3 Eine ausführliche Beschäftigung mit sprachlichen (Mehrebenen-) Netzwerken findet sich z.B. bei Mehler et al., 2008.

## Zur Kreativität des Individuums

Nutzer von Tagging-Systemen bedienen sich verschiedener Strategien zur Bildung eines Tags, die je nach ihren Zielsetzungen und Eigenarten stark variieren können. Eine explorative Studie, in der 10 Probanden an der Hochschule der Medien in Stuttgart nach einem Tagging-Experiment ausführlich befragt wurden, ergab, dass 40% der Teilnehmer Tags ausschließlich aus den Wortformen der betreffenden Artikel wählten.<sup>4</sup> Die Hälfte der Teilnehmer nutzte Mischstrategien aus textbasierten, eigenen sowie Komposita-Tags. Lediglich ein einziger Teilnehmer personalisierte die verwendeten Tags und entnahm sie nicht aus den zu taggenden Artikeln. Die Tendenz zur simplen Übernahme bereits gegebener Informationsfragmente bei der Lösung der Tagging-Aufgabe zeigt sich auch in weiteren Ergebnissen der Studie. So zeigte der Vergleich von zwei Gruppen, bei denen die erste keine, die zweite Gruppe hingegen alle Tags von anderen Nutzern als Vorlage hatte, dass es eine starke Tendenz zur Übernahme vorhandener Tags gibt (auch auf die Gefahr hin, dass diese eventuell nicht ganz so treffend sind). Entsprechend hat sich in anderen Untersuchungen gezeigt, dass z.B. für Bookmarks als erste verwendete Tags regelmäßig auch später die Häufigkeitsliste der Tags für dieses Bookmark anführen (Wu und Zhou, 2009). In Anlehnung an das „Priming“-Konzept in der Psychologie, dass ersten Eindrücken während der Frühentwicklung von Lebewesen eine unwiderrufliche, prägende Bedeutung beimisst, kann dieses Phänomen als „Tag-Priming“ bezeichnet werden.

## Abgrenzung zur Sacherschließung

In den bisherigen Abschnitten dieses Beitrags wurde bewusst nicht von „Sacherschließung“ gesprochen. Dieser Begriff ist mit der professionellen Inhaltserschließung durch Informationsexperten (etwa Bibliothekaren) assoziiert, die sich bei der Erfassung einer Quelle auf verschiedene Formen kontrollierter Vokabulare stützen. In gewisser Weise komprimieren auch sie den Inhalt einer Quelle und ordnen ihn in hierarchische oder paradigmatische Gefüge ein. Naturgemäß sind ihre kreativen Möglichkeiten dabei allerdings erheblich eingeschränkt, da die Sicherheit vorgegebener Schlagworte oder relativ starrer Hierarchiepositionen kurzfristige dynamische Anpassungen und Verbesserungen konterkarieren. Die Prozesse der Erzeugung und der Anwendung eines Vokabulars sind bei dieser Form der Inhaltserschließung relativ streng getrennt. In Tagging-Systemen vollziehen sie sich dagegen immerfort, gleichzeitig und gemeinschaftlich. Diese Grundarchitektur bedingt gleichzeitig das Potenzial wie auch die „Gefahr“ des Chaos in den Systemen. Es hat sich aber gezeigt, dass gewisse Strukturvorgaben in Form von Kategorien u.ä. den Tagger bei der Vergabe positiv unterstützen können, sodass der Grad an Homogenität und Präzision im System steigt (Bar-Ilan et al., 2008). Die

---

4 Die Autoren bedanken sich bei Annika Hager für die Durchführung der Befragung.

Vergabe mit diesen Kategorien sollte allerdings freiwillig sein, zudem sollte der Nutzer eigene Kategorien zur Struktur hinzufügen können. Systeme zur Unterstützung des Nutzers können auf Seiten des Tagging-Systems selbst aber auch durch Mittlersysteme, die zugleich das Problem des Cross Taggings zu lösen versuchen, bereitgestellt werden (ein solches System findet sich etwa bei Peters und Weller, 2008).

## Gründe für Wortneubildungen

Die hohe Gebundenheit bei der Auswahl eines Tags an den Titel bzw. Text des betreffenden Artikels ist nicht überraschend, da sich der Inhalt desselben mit seinen eigenen Begriffen sehr einfach und schnell beschreiben lässt. Dies gilt umso mehr, wenn hierfür mehrere Ausdrücke benutzt werden. Das „Chaos“ in den Tagging-Systemen lässt sich aber am besten aus den vielfältigen Motivationen erklären, die die Tagger bei der Vergabe verfolgen. Tags können z.B. als Themenfinder, zur Quellenstrukturierung, als Literatursammlung, für private Bewertungen, zur Organisation von Teams oder zur *Quick&Dirty*-Speicherung eingesetzt werden. Ähnlich wie in alltagssprachlichen Kontexten führt der persönliche Wunsch, inhaltlichen Bedeutungsverschiebungen, Perspektivwechseln oder Differenzierungen Ausdruck zu verleihen, zur Manipulation bzw. zur kompletten Neugestaltung der sprachlichen Symbole. Je nach Motivation werden dann auch bestimmte Tag-Kategorien wahrscheinlicher.<sup>5</sup> Weitere Gründe für Sprachmanipulationen können z.B. sein (Lemnitzer und Jonas 2007):

- **Nachträgliche Differenzierung:** Beschreibung neuer Sachverhalte (z.B. „Hochladen“, „Dateitransfer“ oder Unterscheidungen wie „Analog-“ und „Digitalkamera“).
- **Persönliche/soziale Differenzierung:** Bildung und Abgrenzung sozialer Gruppen (z.B. in der zielgruppenspezifischen Werbung mit „inhaltsleeren“ Floskeln wie „Feel the difference“ oder „Happyologie“).
- **Sprachökonomie:** leicht verständliche und einfach zu bildende Wortverkürzungen wie z.B. „HSV-Transfer“ oder „Unterwegsradio“ (meist Gelegenheitsbildungen).
- **Kreativität:** z.B. „Mondscheintarif“ oder „Bierdeckelsteuer“.

Sollen also „Personalisierungen“ und/oder das spätere Wiederauffinden einer Quelle bei der Suche eines passenden Tags berücksichtigt werden, so müssen durch diese Zeichenfolge zusätzliche Eigenschaften ausgedrückt werden, die eine betreffende Ressource möglichst eindeutig von den anderen abgrenzt (Steels 2006). Dass solche speziellen Anforderungen durch den einfachen Rückgriff auf lexikalisch bereits verfügbare Ausdrücke erfüllt werden können, darf als Ausnah-

5 Unterschieden werden können etwa die Kategorien „Thema“, „Zeit“, „Ort“, „Art/Format“, „Autor“, „Meinung/Qualität“, „Benutzungskontext“ und „Selbstreferenz“ (Bischoff et al., 2009).



me gelten, zumal die übrigen Nutzer ja auf den gleichen allgemein verfügbaren Wortschatz zurückgreifen. Deshalb wird in einem solchen Fall die *kreative* Suche nach geeigneten Begriffen seitens des Nutzers wahrscheinlich. Die durch die Zielsetzung des Taggers und die Vorgaben des Tagging-Systems entstandene Lücke im mentalen Lexikon wird dann geschlossen, indem z.B. Ausdrücke zusammengezogen werden, die vorher nicht kompositorisch vereint waren. Eine andere Möglichkeit besteht darin, ganz neue Wörter zu bilden, um einen bestimmten Sachverhalt bzw. mehrere Dimensionen eines Sachverhaltes zu erfassen.

## Die kognitiven Investitionen der Tagger

Aus den Strategien, die bei der Vergabe von Tags üblicherweise angewendet werden, lässt sich ein erstes Kriterium für die Kreativität von Tagging-Systemen gewinnen: der Grad der semantischen Kompression. Diese ist nämlich keinesfalls ein triviales Problem, denn der Nutzer nimmt dabei innerhalb kurzer Zeit eine Vielzahl von (mehr oder weniger bewussten) Abwägungen vor, ehe er sich für eine bestimmte (komponierte, neue) Wortform entscheidet. Das Ziel dabei ist, einen mehr oder weniger komplexen und vielschichtigen Inhalt mit einer einzigen (oder sehr wenigen) Wortform(en) zu erfassen. Ähnlich einer SMS oder eines *Tweets* ist auch beim Tagging das Medium an der Formung des Inhaltes beteiligt: Tagging-Systeme motivieren und provozieren begriffliche Kompression. Gleichzeitig stellt sie die einfachste Form eines kreativen Aktes dar: der ursprüngliche Gegenstand (Textressource) wird unter Beibehaltung seiner individuell als wichtig beurteilten Kerninhalte verformt, reduziert und/oder um passende Aspekte ergänzt (Annotation). Die verdichtende Manipulation von Zeichenketten kann folglich als Motivator und als eine wichtige Größe zur Beurteilung der Kreativität eines Tagging-Systems betrachtet werden. Für das grundsätzliche Funktionieren des Tagging-Systems ist diese kognitive Leistung jedoch nicht unbedingt notwendig, es handelt sich ja nicht um einen „Kreativitätswettbewerb“. Die Verortung einer Quelle in einem System von weiteren Quellen - sei es in hierarchischer oder in paradigmatischer Hinsicht - kann entsprechend auch durch den einfachen Rückgriff auf bereits bestehende Ausdrücke gelingen. In diesem Fall zeigt sich sprachliche Kreativität in semantischen Verschiebungen bzw. Neu- und Redefinitionen von Wörtern. Welchen „Wert“ eine solche semantische Verschiebung hat, kann jedoch nicht an der Zeichenkette selbst abgelesen werden. Vielmehr handelt es sich um eine soziale Übereinkunft, ob ein bestimmtes sprachliches Symbol „sinnvoll“ und „aussagekräftig“ ist. Ein starker Beleg hierfür ist die Intensität der „Nachnutzung“ eines Tags, die operationalisiert und in die Beurteilung eines Tags miteinbezogen werden (vgl. Abschnitt 4) kann.

## Resümee: Intelligenz und Kreativität in Tagging Systemen

Die eigentliche Intelligenz und Kreativität eines Tagging-Systems liegt in seinen Nutzern und Administratoren. Diese Erkenntnis ist eigentlich wenig überraschend, sie ist aber angesichts der regelmäßig wiederkehrenden Anläufe, dem Internet oder seinen Subsystemen esoterische und/oder metaphysische Weihen zu Teil werden zu lassen, offenbar notwendig zu betonen. Es sei der globalen Vernetzung jedoch zugestanden, dass sie gewissermaßen ein neues Bewusstsein für den Status quo der menschlichen Evolution und eine dynamische und jederzeit zugängliche Enzyklopädie des menschlichen Wissens schafft, die eine gewaltige Beschleunigung der Erkenntnisproduktion nach sich zieht. Diese wird jedoch immer noch durch den Menschen geleistet. Ein Computer löst ebenso wie ein Netzwerk an sich kein einziges Problem. Ohne das Themenfeld an dieser Stelle zu überdehnen, kann bei Tagging-Systemen demnach in gleicher Weise für einen winzigen (aber relevanten) Informationsausschnitt aus dieser Enzyklopädie von einem klugen und nützlichen Organisationsinstrument gesprochen werden. Da sich die Intelligenz und Kreativität der Nutzer dieses Instrumentes sprachlich äußert, kann nicht zuletzt an seinen dominanten Zeichenfolgen eine Qualifizierung seines Potenzials festgemacht werden.

## Vermessungen zur Kreativität der Sprache

In dieser Untersuchung dienen sprachliche Ausdrücke als potenzielle Manifestationen von Kreativität zur Qualifizierung von Tagging-Systemen. Die Art und Weise der Tag-Formung, d.h. die Komplexität und das Ausmaß kognitiver Investitionen zur Rekombination sprachlicher Symbole, wird zu diesem Zweck messbar gemacht. Daher ist zunächst eine allgemeine Begutachtung aller Möglichkeiten notwendig, die ein Sprachsystem zur Generierung neuer Symbolkombinationen bereithält. Anschließend werden für Tagging-Systeme relevante Kombinationsformen bezüglich ihres kreativen Gehaltes eingeschätzt und in eine entsprechende Rangliste eingeordnet. Im empirischen Teil (vgl. Abschnitt 4) werden dann nach einem mehrstufigen Auswahlverfahren Kandidaten für besonders kreative Tags ausgesucht und qualifiziert.

## Neotags als Manifestationen sprachlicher Kreativität

Zugrunde gelegt wird das im Zusammenhang mit sprachlichen Ausdrücken allgemein anerkannte triadische Zeichenmodell von Ogden und Richards, das unter anderem jedem Wort einen Ausdrucks- (Zeichen, Symbol, Schriftzeichen usw.) und einen Bedeutungsbestandteil (Sinn, Inhalt, „Begriff“ usw.) zuordnet (Ogden und Richards, 1974). In dem Modell beantwortet die Ausdrucksseite die Frage, *wie* ein Wort gebildet wird, die Bedeutungsseite, *was* es bedeutet (welchen „Beg-

riff“ man sich von etwas macht), während das Ding (ein Sachverhalt, Gegenstand usw.) dasjenige ist, *warum* bzw. *wofür* ein Wort benutzt wird. Anhand dieser drei Aspekte eines sprachlichen Ausdrucks ergeben sich im Zusammenhang mit Neologismen verschiedene Definitionsprobleme.

### 1. Das Problem des „Neuen“

Durch die Unterscheidung von Form und Bedeutung ergeben sich bezüglich des „Neuen“ am Neologismus unterschiedliche Konstellationen. Demnach kann ein Neologismus verschiedenes sein:

- ein neu gebildetes Wort mit neuer Bedeutung (und damit ein „Neulexem“, Herberg, 2002b)
- ein neu gebildetes Wort mit alter Bedeutung (und damit ein „Neuformatismus“, Schippan, 2005) (z.B. „Hausmeister“ wird zu „Facility-Manager“)
- ein altes Wort, bei dem sich die Bedeutung verändert hat (Neusemantismus, Schippan, 2005; z.B. Maus: früher nur als Bezeichnung für das Lebewesen, heute auch als „PC-Steuerungsgerät“)

### 2. Das Problem der „Nachhaltigkeit“

Bewertet man situative Einmal-, Augenblicks- oder Gelegenheitsbildungen (zusammen auch „Okkasionalismen“ genannt) als Neologismen? Inwieweit müssen diese von der Allgemeinheit verwendet werden, um als „gültig“ betrachtet zu werden? Wann und wie lange war/ist ein bestimmtes Wort ein Neologismus bzw. ab wann ist es keiner mehr?

### 3. Das Problem des Bezugssystems

Welche Sprachnorm ist für die Bewertung von Neologismen anzusetzen? Sind Wörter aus Dialekten, Soziolekten und z.B. Fachsprachen Neologismen, wenn sie plötzlich in der Hochsprache allgegenwärtig sind? Sind Wörter aus der Hochsprache, die plötzlich in einem bestimmten Soziolekt, Dialekt oder in einer Fachsprache mit anderer Denotation verwendet werden Neologismen? Ist außerdem ein Wort, welches aus einer anderen Sprache stammt, erst ein Neologismus im Deutschen, wenn es nach deutschen Grammatikregeln flektiert bzw. dekliniert wird und bis dahin ein Fremdwort? Oder sind Fremdwörter noch vor ihrer Anpassung an das deutsche Grammatiksystem als Neologismen zu bewerten, wenn die anderen Faktoren für die Bewertung als Neologismus sprechen?

Der hier aufgezeigte Problemraum könnte nochmals erweitert werden: Sind auch Phrasen, Akronyme und Abkürzungen Neologismen, oder lässt sich dieses Prädikat nur einzelnen, vollständigen Wörtern verleihen? Wie sieht es mit der „Wohlgeformtheit“ aus, sind etwa simple neue Komposita (wie sie in Tagging-Systemen üblich sind) bereits Neologismen?

## Neotags

Im Zusammenhang mit Wortneubildungen zeigt sich wie dargestellt ein sehr komplexes Feld an möglichen Einflussgrößen, Kriterien und Problemen, die eine eindeutige Definition des Neologismus-Begriffs für jeden Untersuchungskontext immer wieder neu notwendig machen (vgl. z.B. Lemnitzer und Jonas, 2007; Herberg, 2002b oder Elsen, 2004). Tags können entgegen diesen Definitionen allerdings nur begrenzt als Neologismen bezeichnet werden, da ihr Entstehungshintergrund und ihre Benutzung nicht mit „normalen“ Sprachnutzungssituationen vergleichbar sind. Tags entstehen nicht dynamisch im Gespräch oder als neue Wörter in Fließtexten. Sie sind zunächst einmal funktional, tendenziell individuell und verfolgen nicht die Ziele einer allgemeinen Verständlichkeit oder der Wohlgeformtheit. Zur Vermeidung von Missverständnissen, aber auch, um der besonderen Herkunft Ausdruck zu verleihen, soll die für Tagging-Systeme angemessene Definition neuer sprachlicher Ausdrücke mit der Einführung einer intuitiv verständlichen Bezeichnung (diesmal in der Tat eines Neologismus) verbunden werden. Im Folgenden wird deshalb von „Neotags“ gesprochen werden. Zur Definition: Ein Neotag ist eine Wortneubildung, die unter den besonderen Gegebenheiten und Zweckmäßigkeiten eines Tagging-Systems durch einen Nutzer gebildet und einer Ressource zugeordnet wurde, um diese als mehr oder weniger komplexes Gebilde in einer einzelnen Wortform insoweit zu erfassen, dass eine spätere Identifikation, Organisation und/oder inhaltliche Bestimmung der Ressource problemlos möglich ist. Anhand der Neotags soll schließlich das kreative Potenzial eines Tagging-Systems sichtbar gemacht werden.

### *Definitiorische Eingrenzungen*

Neotags können ausschließlich Neuformatismen oder Neulexeme sein. Neusemantismen (Bedeutungsveränderungen identischer oder neuer Zeichenketten) können nur durch die intensive Betrachtung des Wortkontextes erschlossen werden, wobei auch hier immer ein Restrisiko der Fehlinterpretation erhalten bleibt. Es wäre allerdings eine Analyse der durch ein Tag bezeichneten Quellen mittels automatischer Verfahren (etwa Textmining-Verfahren) zu erwägen, was allerdings nicht Gegenstand dieses Beitrages ist.

Neotags können hingegen keine situativen Einmal-, Augenblicks- oder Gelegenheitsbildungen (Okkasionalismen) sein. Ein neues Wort muss erst eine gewisse Streuung erfahren, um als Neologismus anerkannt zu werden. Als Gradmesser für diese Etablierung werden oft die Aufnahme und die Konstitution in einem Standardwörterbuch genannt. Entsprechend müssen Neotag-Kandidaten mit verschiedenen Korpora abgeglichen werden, um ihre eventuell bereits bestehende lexikalische Existenz auszuschließen. Dabei kann eine Differenzierung in Korpora verschiedener Jahrgänge Aufschluss darüber geben, ab wann ein Ausdruck oder eine Variation eines Ausdrucks bereits im Wortschatz enthalten war. In Anlehnung an diese Kriterien wird ein Tag nur dann als Neotag bezeichnet, wenn es neben den anderen Voraussetzungen auch zum Standardvokabular des Tagging-

Systems gehört. Hierfür muss es von überdurchschnittlich vielen Nutzern überdurchschnittlich häufig benutzt werden. Außerdem werden die Steigungsgrade im Zeitverlauf berücksichtigt und auf dieser Basis ein Ranking vorgenommen. Regelmäßige Steigerungen von Nutzeranzahl und absoluter Häufigkeit implizieren eine hohe Gültigkeit und Verständlichkeit des Neotags. Wie bei der Bewertung von Neologismen wird für Neotags die allgemeine deutsche Hochsprache als Bezugsgröße gewählt (vgl. z.B. Herberg, 2002b: S. 196).<sup>6</sup> Ebenso wird mit neuen Phrasen und eindeutig neuen, durch den Nutzer entwickelten Akronymen und Abkürzungen verfahren, die sonst nicht als Neologismen betrachtet werden (Elsen, 2004). Da es sich bei Tagging-Systemen um per se mehrsprachig nutzbare Plattformen handelt bzw. dies explizit gewünscht und gefördert wird, ist der „Fremdwort-Aspekt“ als Kriterium zur Beurteilung der Durchdringung der Sprachen untereinander kaum sinnvoll auf solche Systeme anzuwenden.

## Ranking der Kreativität

Zur Bewertung der kreativen Leistung der Nutzer eines Tagging-Systems soll eine begründete Auswahl von Tag-Kandidaten durch ein Punktesystem messbar gemacht werden. Jedes Tag, das nicht aus dem Titel, dem Abstract, dem Volltext oder als lexikalisch existente Zeichenfolge übernommen wird, kann als mehr oder weniger „kreativ“ eingestuft werden. Grundlegend für das Ranking sind die vom Nutzer angewendeten Wortbildungsprozesse<sup>7</sup>, die je nach Komplexität eine Stufe zwischen 1 und 8 erreichen können, wobei 1 die „kreativste“ Form repräsentiert, die die höchsten kognitiven Investitionen erfordert. Nach eigenen Recherchen konnte kein bekanntes Kreativitätsranking aufgrund von Wortbildungsprozessen gefunden werden, sodass ein eigenes Ranking (Tabelle 1) erstellt wurde.

Zur weiteren Erklärung einiger der linguistischen Kategorien:

- Konversion (Wechsel der Wortklasse) (Stufe 6)

Zum Beispiel Wechsel von Verb oder Adjektiv zum Substantiv und umgekehrt.

- Kurzwortbildung (Stufe 5)

Etwa *Abkürzungen* (wie ein Wort ausgesprochen, z.B. Nato, Sars; Zusammenfassungen aus den Anfangsteilen von Wörtern, z.B. Kripo, Soko), *Kurzwörter* (Folge ausgesprochener Buchstaben, z.B. SPD, CDU), *Kopfwörter* (bestehen nur aus dem Anfang eines Wortes, z.B. Limo: Limonade) usw.<sup>8</sup>

- Kontamination (Verschmelzung) (Stufe 4)

6 Grundsätzlich können aber auch beliebige andere Wortsammlungen als Maßstab eingesetzt werden, etwa eine Sammlung neuer jugendsprachlicher Begriffe (vgl. z.B. Niegel et al., 2010).

7 Lemnitzer und Jonas, 2007

8 Weitere Formen sind Schwanzwörter (Bus: Omnibus, Rad: Fahrrad), Klammerwörter (Apfelblüte: Apfelbaumblüte), Rumpfwörter (Lisa: Elisabeth) und Initialwörter (E-Mail, S-Bahn).

Ähnlich der Komposition, allerdings verlieren die Wörter bei dieser Zusammenführung Teile ihrer eigenen Lautkette (z.B. „Breakfast“ und „Lunch“ werden zu „Brunch“).

- Multiple Mischformen (Stufe 3)

Als multiple Mischformen werden Neotags bezeichnet, die mehrere Wortbildungsprozesse in sich vereint tragen.

- Analogiebildung (Stufe 2)

Bei der Analogiebildung dienen nicht grammatikalische Regeln, sondern Wortbildungsmuster als Grundlage des Wortbildungsprozesses. Ein Beispiel hierfür ist das Wort „-gate“, welches seit dem Watergateskandal allgemein als Muster für politische Skandale benutzt wird. So werden andere Skandale mit dem Element „-gate“ gebildet: Teletubbiegate, Spycgate, Isargate, Bedgate (Lemnitzer und Jonas 2007).

- Neulexem (Stufe 1)

Völlig neue Wortkreationen, die mit alter oder neuer Bedeutung versehen sein können.

Tabelle 1: Ranking der Kreativität bei der Tag-Vergabe

Zeichenbildung	Stufe
Tag aus Titel	10
Tag aus Abstract/Text	9
Existent, aber nicht in Quelle	8
Komposition und Derivation (Aneinanderreihung/Affixe/Suffixe)	7
Konversion (unerwartete Wortklassenwechsel)	6
Kurzwortbildungen (Verkürzung)	5
Kontamination (Verschmelzung)	4
(Multiple) Mischformen	3
Analogiebildung	2
Neulexem	1

## Empirische Erhebung und Analyse

Im Folgenden wird ein Analyse-Modell vorgestellt, das durch die Kombination eines geisteswissenschaftlichen Zugangs mit (computer-)linguistischen und sprachstatistischen Verfahren intellektuelle Aussagen zur Qualität eines Tagging-Systems ermöglicht. Automatische Verfahren dienen zur ersten Strukturierung der Datenmengen, um sodann über verschiedene Filterungsmethoden zur Identifikation geeigneter Analyse Kandidaten zu gelangen. Von diesem „Kern“ der wichtigsten Tag-Kandidaten ausgehend, kann dann je nach Forschungsabsicht und -möglichkeit die Größe der Datenbasis reguliert werden. Exemplarisch wird das geschilderte Vorgehen anhand der Analyse des „kreativen Potenzials“ des Tagging-Systems „BibSonomy“ durchgeführt. Aufgegriffen wird dabei die immer wieder geäußerte These, dass in großen Kommunikationsverbänden „mehr“ Intelligenz als in autonomen Einzelwesen stecke. Als Datenbasis der Untersuchung dienen Datenabzüge aus BibSonomy, die etwa in sechsmonatigem Rhythmus bereitgestellt werden. Die 500 häufigsten, lexikalisch nirgends verzeichneten Tags aus dieser Datenmenge wurden mittels eines spezialisierten Kategoriensystems formal und inhaltlich evaluiert, damit schließlich ein „Kreativitätsranking“ zur Anwendung gelangen kann, das die sprachliche Schöpfungsfreude der Tagger bzw. des Tagging-Systems einzuschätzen ermöglicht.

### Datenstruktur

Da es zum Prinzip des Taggings gehört, dass die Nutzer bei der Vergabe von Deskriptoren in keiner Weise eingeschränkt werden, sind die zugrunde gelegten Daten höchst heterogen. Deshalb erfolgte als erstes eine Analyse der linguistischen Struktur des gesamten Tagging-Systems (Niemann, 2009). Die Separierung grammatikalischer Einheiten wurde mittels Part-of-Speech-Tagging („POS-Tagging“) realisiert. Bei dieser computerlinguistischen Methode wird ein zu analysierender Begriff mit einem Wörterbuch verglichen, um seine Wortart und seine Grundform zu bestimmen. Zusätzlich werden über verschiedene Heuristiken, etwa über die Großschreibung deutscher Substantive u.ä., Wahrscheinlichkeiten für eine Zuordnung ermittelt, sollte ein direkter lexikalischer Vergleich erfolglos bleiben. Im untersuchten Datensatz konnten durch POS-Tagging insgesamt 41.591 (42%) der Tags mindestens einer Sprache zugeordnet und mit einer Information über die jeweilige Wortart versehen werden. Knapp 43% dieser Tags sind in deutscher Sprache und in ihrer Mehrheit (80%) Substantive oder Eigennamen. Die englischen Tags sind hier wesentlich heterogener, sie enthalten anteilig doppelt so viele Adjektive und dreimal so viele Verben. Insgesamt stellen aber auch unter ihnen Substantive und Eigennamen die Mehrheit (72%).

### Identifikation des Kreativen Potenzials

Die Anwendung von POS-Tagging und Matching-Verfahren hatte primär den Zweck, eine Vorstrukturierung der Datenmenge zu erreichen, welche eine Auswahl von Kandidaten besonders kreativer Tags ermöglicht. In den BibSonomy-Daten ist eine internationale Nutzergemeinde offenbar geworden, die klar von englisch- und deutschsprachigen Taggern dominiert wird. Lexikalisch konnte eine annähernde Drittelung der Daten festgestellt werden, die in die Gruppen „existent“, „Wortketten“ und „nicht existent“ (Abkürzungen, Neologismen bzw. Neukombinationen von Sprache jeglicher Art) zerfiel. Als besonders interessant im Sinne der These von „Intelligenz“ und „wertvollem Wissen“ in den Daten kann diese letzte Gruppe gelten. Die ausführliche Beschreibung dieser sprachlichen Phänomene wurde bereits im Abschnitt 3 geleistet.

### Intellektuelle Erschließung I: Inhaltliche Struktur

Im ersten Teil wurden zunächst die „Hardfacts“ der Tags (Ausdruckslänge, Nutzeranzahl und Wortart) untersucht. Die intellektuell erschlossenen Wortarten (Nomen, Verb, Adjektiv, Pronomen und Präposition) wurden entsprechend ihrer üblichen Definition vergeben, während für die besonderen Gegebenheiten in Tagging-Systemen weitere Wortklassen (Akronyme, Abkürzungen, Neologismen, Eigennamen und Nichtwörter) nach eigenen Identifikationsmerkmalen entwickelt wurden. Da Tags sehr häufig Kompositionen mehrerer Wortarten bzw. -klassen sind, wurde jedem Teilausdruck eines Tags zusätzlich seine individuelle Wortart zugeordnet (z.B. „bioinformatics“: „bio“ = Abkürzung, „informatics“ = Nomen).

### Ergebnisse des ersten Analyseabschnitts

Die lexikalischen Strukturen der Tags erlauben zwar noch keine Rückschlüsse auf konkrete semantische Gehalte, dafür liefern sie aber erste wichtige Hinweise für die grundsätzliche inhaltliche Beschaffenheit der Tags:

#### Ausdruckslänge

Die Anzahl der Zeichen eines jeden Tags wurden automatisch gezählt. Zusätzlich liegen Vergleichswerte aus der „Korpusbasierten Grund-/Wortformenliste“ (DeReWo) des Instituts für Deutsche Sprache als natürlichsprachlicher Korpus vor.<sup>9</sup> Dieser weist bezüglich der Häufigkeitsverteilung bestimmter Wortlängen Ähnlichkeiten mit jener aller Tag-Daten auf. Beide Korpora bestehen zum überwiegenden Teil aus Ausdrücken mit fünf bis vierzehn Zeichen. Bei den Tags ist die Verteilung allerdings insgesamt flacher, wobei ein deutlicher Ausreißer nach oben bei Längen von drei bis vier Zeichen besteht. Dies ist auf die große Anzahl von Akronymen und Abkürzungen zurückzuführen. Zudem fällt die Kurve in den Bereichen über vierzehn Zeichen wesentlich langsamer ab als im Vergleichskor-

9 <http://www.ids-mannheim.de/kl/projekte/methoden/derewo.html>



pus, was auf häufige Zusammensetzungen lexikalisch ursprünglich getrennter Wörter schließen lässt. Betrachtet man nur die 500 häufigsten Tags, so fällt ein deutlicher Ausreißer bei drei Zeichen auf, dem ein abrupter Einbruch bei fünf Zeichen und ein erneut steiler Anstieg folgen. Diese überproportionale Häufigkeit ist erneut durch die große Menge von Abkürzungen zu erklären, da sich anschließend ab einer Länge von fünf Zeichen beide Verteilungen bezüglich ihrer relativen Häufigkeiten wieder ähneln.

### Nutzeranzahl

Die Anzahl unterschiedlicher Nutzer, die gemeinsam ein Tag benutzen, gibt Aufschluss über den Anteil der „collaborativeness“, also der „Vergemeinschaftung“, im Tagging-System. Zudem ist dieser Wert zusammen mit der absoluten und der relativen Häufigkeit des Tags ein Hinweis für die allgemeine Gültigkeit eines Neotags. Für das BibSonomy-System gilt, dass gut 60% der Tags von nur einem oder zwei Nutzern verwendet wurden. Dieser klassische „Long Tail“<sup>10</sup> besteht also aus etwa 23.000 sehr speziellen Wortformen, die offenbar nur auf die individuellen Bedürfnisse einzelner Nutzer passen. 16,32% der Tags werden von drei bis fünf Nutzern verwendet. Diese 6.133 Wortformen beschreiben vermutlich sehr gut das inhaltliche Spektrum des Tagging-Systems bzw. bilden vereinzelt das Organisationsbedürfnis kleinerer Gruppen ab, die sich z.B. für den Aufbau gemeinsamer Ressourcenlisten auf ein bestimmtes Tag geeinigt haben. Die restlichen gut 22% der Tags werden von mindestens sechs Nutzern verwendet, wobei mit zunehmender Nutzeranzahl eine unabhängige Nutzung der betreffenden Wortform immer wahrscheinlicher wird. Die in diesen 8.433 Tags aufgefundenen Neotags können als besonders aussichtsreiche Kandidaten für allgemein gültige Neuschöpfungen gelten, die gleichzeitig kreative und allgemein verständliche Züge tragen.

Der Vergleich der Tag-Daten aller Tags mit den Nutzerverteilungen der 500 am häufigsten Vergebenen zeigt signifikante Unterschiede bis zu einer Anzahl von fünf Nutzern. Erwartungsgemäß werden die sehr häufigen Tags wesentlich seltener von nur einem oder zwei Nutzern verwendet. Der Anteil der im eigentlichen Sinne kollaborativen Tags ist dagegen mit über 40% doppelt so hoch als in der Vergleichsmenge, zudem spielen hier Fünfergruppen keine besondere Rolle mehr. Diejenigen Nutzer, die es alleine mit einem Tag in die „Top 500“ der häufigsten Tags geschafft haben, können als „Poweruser“ bezeichnet werden. Von besonderem Interesse wird es sein, aus welchen Fachbereichen diese Tagging-Pioniere mehrheitlich stammen.

### Wortarten

Die Ergebnisse der Wortartverteilung liefern je nach Zählart, d.h. der Berücksichtigung der einzelnen Glieder eines Tags (z.B. bei „visualcommunication“ das Adjektiv „visual“ und das Nomen „Communication“), nur leicht variierende Werte. Wortarten, die mit unter 2% vertreten sind, wurden hier aufgrund ihrer fehlen-

10 Anderson, 2004 sowie bezogen auf Tags Barnett, 2006.

den Relevanz nicht berücksichtigt. Für die 500 häufigsten Tags zeigt sich eine klare Dominanz nominativer Wortformen (zwischen 45 und 50%), die als Ergebnisse der Motivation einer inhaltlichen Erschließung betrachtet werden können. Dies gilt grundsätzlich auch für die zweitgrößte Gruppe, die aus 129 Akronymen besteht (20%). Eigennamen dienen als Autoren bzw. Nutzernamen zumeist der Grobstrukturierung der eigenen Literatursammlung und sind im Sinne des kreativen Potenzials nicht weiter relevant. Ähnlich verhält es sich mit den ebenfalls mit 10% noch signifikant vertretenen Adjektiven, die insbesondere als allein stehende Deskriptoren nur in Bezug auf eine konkrete Quelle Sinn machen bzw. als sehr allgemeine Attribute (z.B. „social“, „own“ oder „visual“) kaum Diskriminationskraft besitzen.

## Intellektuelle Erschließung II: Kreatives Potenzial

Um das kreative Potential der „nicht existenten“ Zeichenketten zu untersuchen, wurden die 500 häufigsten Tags dieser Gruppe (die 44,6 % der Nutzung des Gesamtsystems repräsentieren) intellektuell (grammatikalisch und inhaltlich) erschlossen.

### Neotag-Identifikation

Für die Identifikation der Neotags innerhalb der Tagmenge wurden zunächst alle bekannten Akronyme und Abkürzungen (25 % der ersten 500 Tags) aussortiert, da diese keine innovative, vom Nutzer selbst entwickelte Zeichenkette darstellen. Anschließend wurde überprüft, ob ein Tag bereits im *British National Corpus*<sup>11</sup> (BNC) vorhanden war, wodurch es sich als Neotag disqualifizieren würde. Darüber hinaus wurde getestet, ob es im *Corpus of Contemporary American English*<sup>12</sup> (COCA) einen Treffer vor dem Jahr 2006 gab. Da das Vorkommen eines Tags in BibSonomy seit 2006 nachgewiesen werden kann, kann ab diesem Zeitpunkt nicht mehr eindeutig entschieden werden, an welcher Stelle die Wortform zum ersten Mal aufgetaucht ist. Sollte es allerdings bereits zuvor aufgetreten sein, so kann es sich nach unserer Definition nicht um ein Neotag handeln. Alle verbliebenen Tags wurden als Neotags bewertet, da sie in irgendeiner Form neue Manipulationen von Sprache repräsentierten.

---

11 <http://www.natcorp.ox.ac.uk>; Das Korpus enthält Texte bis zum Entstehungsjahr 1976.

12 <http://www.americancorpus.org>

## Spreading

Ein zentraler Aspekt der These über das kreative Potenzial in Tagging-Systemen ist, dass sich „kluge“ neue Deskriptoren schnell und stetig verbreiten. Um diesen positiven Wiederhall durch das Netzwerk zu untersuchen, wurden die nach den vorgeschalteten Filterverfahren verbliebenen 150 Neotags auf ihren Steigungsgrad im Zeitraum vom 30.03.2006 bis zum 01.01.2010 überprüft. Die absoluten Nutzerzahlen sowie die absoluten Häufigkeiten eines Tags wurden zu neun Stichtagen normalisiert und verglichen. Anschließend wurden die Differenzen zwischen allen Stichtagen, sowohl für die Nutzeranzahl als auch für die Häufigkeit der Tags, berechnet und aufsummiert. Je höher diese Gesamtsteigerungsrate ist, desto schneller hat sich ein bestimmtes Tag über die letzten 4 Jahre in BibSonomy verteilt und je eher handelt es sich um eine gelungene kreative Leistung eines Nutzers. Zu beachten ist, dass das Spreading zwar ein wichtiger Indikator für die allgemeine Verständlichkeit und die Akzeptanz eines Tags ist, über die Komplexität der Wortform und ihren Bildungsprozess selbst jedoch wenig aussagt. Für diese zweite, wichtige Dimension der Kreativität kommt deshalb das Bewertungssystem wie in Abschnitt 3.2 erläutert zum Einsatz.

## Ergebnisse des zweiten Analyseabschnitts

Zu beobachten ist, dass Neotags häufig aus der Komposition zweier Wörter entstehen (Stufe 7, vgl. auch Tabelle 2). Auch die Übernahme von neuen Wörtern direkt aus dem Titel, dem Abstract oder dem Text der getaggtten Ressource tritt häufig auf (Stufen 9 und 10). Neben diesen beiden Phänomenen ist die Verkürzung (Stufe 5) eines Wortes, welches dann als Tag dient, ein beliebtes Verfahren, das bereits unter den ersten 500 Tags mit einem Anteil von 25 % aufgefallen war. Neotags, die eine höhere kognitive Leistung erfordern, sind hingegen eher dünn gesät. Unter den 15 Neotags mit dem höchsten Steigungsgrad finden sich aber auch einige sehr interessante Kandidaten. Das Tag „socialnetworking“ z.B., welches durch die Komposition von „social“ und „network“ zum Neotag „socialnetwork“ wurde, hatte sich nach seiner Bildung zusätzlich durch eine Konversion vom Nomen zum Verb weiter von seinen Grundformen emanzipiert. Auch die Neotags „semweb“, „socialnets“ und „wwwkap2“ nehmen an mehreren synchronen Wortbildungsprozessen teil. So ist „semweb“ eine Komposition aus „sem“ und „web“, die nur aufgrund der vorausgegangenen

Verkürzung von „semantic“ auf „sem“ stattfinden konnte. Hervorzuheben sind nicht zuletzt Analogiebildungen (Stufe 2), die durchaus in BibSonomy zu finden sind. Die Ausdrücke „open“ in „openaccess“ oder „opensource“ sowie „social“ in „socialnets“ und „socialnetworking“ können als Analogiebildungselemente aufgefasst werden. Außerdem wurde jedem Tag ein Kompressionsgrad zugewiesen, der die Anzahl der miteinander verknüpften semantischen Ebenen anzeigt. So hat das Tag „OpenSource“ beispielsweise die zwei semantischen Komponenten „open“ und „source“. Entsprechend der Komplexität der Wortneu-

bildungen zeigte sich, dass in der Mehrheit der Fälle lediglich zwei Komponenten zusammengeführt wurden, der Kompressionsgrad einzelner Tags also überwiegend gering ist.

Tabelle 2: Top 15 Neotags sowie einige ausgewählte kreative Kandidaten

Rang (Nachnutzung)	Tag	Steigung	Bewertung
1	Statecharts	3,66010622	7
2	Mythesis	2,83569057	7
3	Folkrank	2,80379841	9
4	Socialnetworking	2,77894016	2
5	to-read	2,77378644	7
6	OpenSource	2,73755705	2
7	Tagora	2,73239376	9
8	Bibsonomy	2,73239376	9
9	Openaccess	2,53415483	2
10	Visualcommunication	2,48995848	7
11	Eval	2,45024911	5
12	Toread	2,40823997	7
13	Graphicdesign	2,39988648	7
14	semantic-web	2,39794001	7
15	Myown	2,29264014	7
30	Semweb	1,83078209	3
57	ByZz	1,07918125	5
58	Ipe	1,00372787	5
69	Socialnets	0,82736927	2
71	wwwkap2	0,81131961	3

## Resümee und Ausblick

In diesem Artikel wurde ein Modell zur inhaltlich-intellektuellen Analyse eines Tagging-Systems vorgestellt und am Beispiel der Kreativität bei der Tagvergabe erprobt. Anlass hierfür war neben den Defiziten der Tagging-Forschung die These wahrscheinlich wertvoller Informationen im „Wortformen-Chaos“ eines Tagging-Systems, dessen theoretische und empirische Fundierung kritisch hinterfragt wurde. Laut des vorgestellten Annotationsdreiecks könnte das Tagging insbesondere die Innovationsfunktion im Rahmen der durch den *Information Overload* notwendigen distributiven Inhaltserschließung erfüllen. „Neotags“ wurden deshalb als Kandidaten innovativer Erschließungsversuche identifiziert und ein linguistisch begründetes System zur Beurteilung derselben entwickelt. Anhand des konkreten Beispiels BibSonomy wurde zunächst die Funktionsweise der Filterverfahren (POS-Tagging, Korpus-Vergleich, Spreading) erläutert und auf diesem Wege eine Menge aussichtsreicher Tag-Kandidaten für kreative Wortneubildungen ermittelt. Die Analyse und Einordnung dieser ersten Kandidaten in das Ranking zeigte, dass nur eine geringe Anzahl von Tags bei ihrer Bildung eine hohe kognitive Investition erfordert hat. Auch wenn sich keine Neuwörter im strengen Sinne innerhalb der Neotags fanden, so ist dennoch in bestimmten Fällen eine nennenswerte kreative Leistung zu beobachten, insbesondere, wenn mehrere Wortbildungsprozesse kombiniert auftraten. Um in diesem Zusammenhang zu repräsentativen Aussagen zu gelangen, ist allerdings die Ausweitung der zunächst explorativen Betrachtung von 500 Tags notwendig. Das vorgestellte Analysemodell hat sich hierfür als fruchtbar und praktikabel gezeigt, zumal bisher kein vergleichbarer Ansatz zur systematischen Bestimmung der Kreativität von Sprachprodukten existiert. Weitere Forschungsbemühungen sollten deshalb die Optimierung und Erprobung des Modells, nicht zuletzt auch in anderen Sprachkontexten wie etwa der Werbesprache oder in Diskurstexten, ins Auge fassen.

Für den bibliothekarischen Kontext sind Tagging-Systeme ein ernst zu nehmendes Phänomen, denn in gewisser Hinsicht stellen sie die „wilde Konkurrenz“ des professionellen Wissensmanagements dar. Die Besorgnis im Hinblick auf die Präzision und die Vollständigkeit einer Recherche mittels Tags ist dabei gerechtfertigt. Gleichzeitig erfreut sich das voraussetzungsfreie Tagging ebenso wie die intuitiv zu bedienenden Suchmaschinen einer besonders hohen Beliebtheit bei den Nutzern. Letztere befassen sich in den seltensten Fällen mit der Frage, ob ein Rechercheergebnis auch im bibliothekarischen Sinne „gut“ und vollständig ist, oder eben nicht. Informationsexperten müssen sich also mit der parallelen Existenz aller drei Erschließungsmethoden befassen und sich als legitime „Sachwalter“ begrifflicher Ordnungen empfehlen, die ganz offensichtlich Sinn und Struktur in das „Chaos“ bringen können. Mit dieser anspruchsvollen Aufgabe können sie sich intensiv beschäftigen, weil sie sich von anderen Pflichten befreien: der sprachlichen Dynamik können sie durch die Beobachtung der Tagging-Plattformen „in Echtzeit“ folgen, effiziente Recherchen werden durch automatische Verfahren auf der Basis ihrer Strukturierungsleistung erreicht. Die zunehmende Verknüpfung der

„OPACs“ mit sozialen Netzwerken, aber auch der Einzug von Suchmaschinentechnologie in die Online-Kataloge weisen in diese Richtung. Erneut handelt es sich hier um eine technisch basierte „Pionierarbeit“, die von Informationsexperten kritisch und konstruktiv begleitet werden sollte.

## Literaturverzeichnis

- (Anderson, 2004) Anderson, C.: The Long Tail: The future of entertainment is in the millions of niche markets at the shallow end of the bitstream. *Wired Magazine*, 12(10), 2004, S. 170–177.
- (Bar-Ilan et. al., 2008) Bar-Ilan, Judit; Shoham, Snunith; Idan, Asher; Miller, Yitzchak und Shachak, Aviv: Structured versus unstructured tagging: a case study. *Online Information Review*, Volume 32, Number 5, 2008, S. 635–647.
- (Barnett, 2006) Barnett, A.: The Long Tail of Tags. 2006. URL: <http://alexbarnett.net/blog/archive/2006/09/16/The-Long-Tail-of-Tags.aspx>
- (Bischoff et. al., 2009) Bischoff, K., Firan, C. S., Kadar, C., Nejd, W., und Paiu, R.: Automatically Identifying Tag Types. In: Ronghuai Huang, Qiang Yang, Jian Pei, João Gama, Xiaofeng Meng, und Xue Li, (Hrsg.), ADMA, Jahrgang 5678 von *Lecture notes in computer science*. Springer, 2009, S. 31–42.
- (Buhse et. al., 2008) Buhse, W., Mühlner, J., und Rausch, A.: Gespräch mit Peter Kruse. *DNAdigital*, 2008, S. 81–99.
- (Elsen, 2004) Elsen, H.: Neologismen: Formen und Funktionen neuer Wörter in verschiedenen Varietäten des Deutschen, Jahrgang 477 von *Tübinger Beiträge zur Linguistik*. Narr, Tübingen, 2004.
- (Everitt; Mills, 2009) Everitt, D. und Mills, S.: Cultural anxiety 2.0. *Media, Culture & Society*, 31(5), 2009, S. 749–768.
- (Herberg, 2002a) Herberg, D.: Der lange Weg zur Stichwortliste: Aspekte der Stichwortselektion für ein allgemeinsprachliches Neologismenwörterbuch. In: Haß-Zumkehr, U., Kallmeyer, W., Zifonun, G., und Stickel, G., (Hrsg.), *Ansichten der deutschen Sprache: Festschrift für Gerhard Stickel zum 65. Geburtstag*, Jahrgang 25 von *Studien zur deutschen Sprache*. Narr, Tübingen, 2002, S. 237–250.
- (Herberg, 2002b) Herberg, D.: Neologismen in der deutschen Gegenwartssprache. *Probleme ihrer Erfassung und Beschreibung*. *Deutsch als Fremdsprache* 39, 2002, S. 195–200.
- (Hoffmann; Wolf, 2008) Hoffmann, B. und Wolf, K. H.: *Brockhaus Enzyklopädie Online*, 2010.
- (Kreye, 2008) Kreye, A.: *Das Ende der Weisheit der Vielen: Serie: Kapitalismus in der Krise*. *Süddeutsche Zeitung*, 2008.
- (Lanier, 2006) Lanier, J.: *Digital Maoism: The Hazards of the New Online Collectivism*. Edge, 2006.
- (Lemnitzer; Jonas, 2007) Lemnitzer, L. und Jonas, J.: *Von Aldianer bis Zauselquote: Neue deutsche Wörter; wo sie herkommen und wofür wir sie brauchen*. Narr, Tübingen, 2007.
- (Lévy, 1997) Lévy, P.: *Die kollektive Intelligenz: Für eine Anthropologie des Cyberspace*. *Kommunikation & neue Medien*. Bollmann, Mannheim, 1997.

- (Mehler et. al., 2008) Mehler, A., Frank-Job, B., Blanchard, P., und Eikmeyer, H.-J.: Sprachliche Netzwerke. In: Stegbauer, C. (Hrsg.), *Netzwerkanalyse und Netzwerktheorie: Ein neues Paradigma in den Sozialwissenschaften*, Jahrgang 1 von Springer-11776 /Dig. Serial], 2008, Seite 413–427. VS Verlag für Sozialwissenschaften / GWV Fachverlage GmbH Wiesbaden, Wiesbaden.
- (Niegel et. al., 2010) Niegel, C., Straeter, G., und Betz, E.: *Hä? ? Jugendsprache unplugged 2010: Deutsch - Englisch - Spanisch - Französisch - Italienisch*. Langenscheidt, Berlin, 2010.
- (Niemann, 2009) Niemann, C.: *Intelligenz im Chaos: Collaborative Tagging als neue Form der Sacherschließung*. In: Barth, R., (Hrsg.), *Die Lernende Bibliothek 2009: Wissensklausur, Unvermögen oder Paradigmenwechsel?*, Jahrgang 33 von Churer Schriften zur Informationswissenschaft, Chur, 2009, S. 59–66.
- (Ogden; Richards, 1974) Ogden, C. K. und Richards, I. A.: *Die Bedeutung der Bedeutung: Eine Untersuchung über den Einfluß der Sprache auf das Denken und über die Wissenschaft des Symbolismus = The meaning of meaning*. Theorie. Suhrkamp, Frankfurt am Main, 1. Aufl. edition, 1974.
- (Peters; Weller, 2008) Peters, Isabella und Weller, Katrin: *Tag Gardening for Folksonomy Enrichment and Maintenance*. *Webology*, Volume 5, Nummer 3, 2008.
- (Schippan, 2005) Schippan, T.: *Neologismen und Archaismen. Fallstudien*. In: Cruse, D. A., Ungeheuer, G., Wiegand, H. E., Steger, H., und Brinker, K., (Hrsg.), *Lexikologie /Lexicology, Handbücher zur Sprach- und Kommunikationswissenschaft*, S. 1373–1380. de Gruyter, Berlin, 2005.
- (Steels, 2006) Steels, L.: *Collaborative tagging as distributed cognition*. *Pragmatics Cognition*, 2006, 14: S. 287–292.
- (Stegbauer, 2008) Stegbauer, C., (Hrsg.): *Netzwerkanalyse und Netzwerktheorie: Ein neues Paradigma in den Sozialwissenschaften*, Jahrgang 1 von Springer-11776 /Dig. Serial]. VS Verlag für Sozialwissenschaften / GWV Fachverlage GmbH Wiesbaden, Wiesbaden, 1. Aufl. edition, 2008.
- (Toffler, 1990) Toffler, A.: *Future shock*. Bantam Books, New York, 1990
- (Wehner, 2001) Wehner, R.: *Miniaturgehirne und kollektive Intelligenz: Zur Evolution biologischer Komplexität; Rede; Dies academicus 2001 anlässlich der 168. Stiftungsfeier der Universität Zürich*. *Zürcher Universitätsschriften*. Univ., Zürich, 2001.
- (Wu; Zhou, 2009) Wu, C. und Zhou, B.: *Semantic Relatedness in Folksonomy*. In: NISS '09: *Proceedings of the 2009 International Conference on New Trends in Information and Service Science*, 2009, S. 760–765. IEEE Computer Society.

