

Folksonomies in Wissensrepräsentation und Information Retrieval

Isabella Peters und Wolfgang G. Stock, Düsseldorf

Die populären Web 2.0-Dienste werden von Prosumern – Produzenten und gleichzeitig Konsumenten – nicht nur dazu genutzt, Inhalte zu produzieren, sondern auch, um sie inhaltlich zu erschließen. Folksonomies erlauben es dem Nutzer, Dokumente mit eigenen Schlagworten, sog. Tags, zu beschreiben, ohne dabei auf gewisse Regeln oder Vorgaben achten zu müssen. Neben einigen Vorteilen zeigen Folksonomies aber auch zahlreiche Schwächen (u. a. einen Mangel an Präzision). Um diesen Nachteilen größtenteils entgegenzuwirken, schlagen wir eine Interpretation der Tags als natürlichsprachige Wörter vor. Dadurch ist es uns möglich, Methoden des Natural Language Processing (NLP) auf die Tags anzuwenden und so linguistische Probleme der Tags zu beseitigen. Darüber hinaus diskutieren wir Ansätze und weitere Vorschläge (Tagverteilungen, Kollaboration und aktorspezifische Aspekte) hinsichtlich eines Relevance Rankings von getaggeten Dokumenten. Neben Vorschlägen auf ähnliche Dokumente („more like this!“) erlauben Folksonomies auch Hinweise auf verwandte Nutzer und damit auf Communities („more like me!“).

Folksonomies in Knowledge Representation and Information Retrieval

In Web 2.0 services „prosumers“ – producers and consumers – collaborate not only for the purpose of creating content, but to index these pieces of information as well. Folksonomies permit actors to describe documents with subject headings, „tags“, without regarding any rules. Apart from a lot of benefits folksonomies have many shortcomings (e.g., lack of precision). In order to solve some of the problems we propose interpreting tags as natural language terms. Accordingly, we can introduce methods of NLP to solve the tags' linguistic problems. Additionally, we present criteria for tagged documents to create a ranking by relevance (tag distribution, collaboration and actor-based aspects). Besides recommending similar documents („more like this!“) folksonomies can be used for the recommendation of similar users and communities („more like me!“).

Inhaltserschließung durch die kollektive Intelligenz in Web 2.0-Diensten

In den frühen Jahren des World Wide Web sind einige wenige Experten in der Lage, Wissen über dieses Medium zu verteilen. Die Mehrzahl all derjenigen, die mit dem WWW zu tun haben, beschränkt

Der Artikel ist eine übersetzte und überarbeitete Fassung eines Vortrags der Verfasser bei der 70. Jahrestagung der American Society for Information Science and Technology in Milwaukee, Wisconsin, am 24. Oktober 2007. Isabella Peters trug eine Version des Vortrags am 29. Oktober 2007 beim Center for Intelligent Information Retrieval (University of Massachusetts, Amherst) vor. Der Studienaufenthalt von Isabella Peters wurde durch eine anteilige Finanzierung der Deutschen Forschungsgemeinschaft (DFG) ermöglicht. Wir danken der DFG für das gewährte Reise-stipendium.

ihren Umgang mit dem Web ausschließlich auf die passive Nutzung des Mediums. Mit dem Beginn des 21. Jahrhunderts kommen Dienste auf, die sehr leicht zu bedienen sind und die damit Nutzern gestatten, Content problemlos zu publizieren. Aus dem (passiven) Nutzer wird nunmehr zusätzlich ein (aktiver) Webautor. Der Konsument von Wissen ist gleichzeitig auch sein Produzent geworden, ein „Prosumer“ im Sinne Tofflers (1980). Da die Autoren (zumindest manchmal) ihre Dokumente wechselseitig korrigieren und fort-schreiben, kann man

in diesem Zusammenhang durchaus von „kollektiver Intelligenz“ sprechen: „With content derived primarily by community contribution, popular and influential services like Flickr and Wikipedia represent the emergence of ‚collective intelligence‘ as the new driving force behind the evolution of the Internet“ (Weiss, 2005, 16). „Kollektive Intelligenz“ entsteht dank der Zusammenarbeit von Autoren und Nutzern in „kollaborativen Diensten“, die zusammengefasst als „Web 2.0“ (O'Reilly, 2005) etikettiert werden können. Solche Dienste widmen sich beispielsweise dem Führen von „Tagebüchern“ (Weblogs) und deren Suche (z.B. Technorati), dem Aufbau einer Enzyklopädie (z.B. Wikipedia), der Ordnung von Lesezeichen zu Webseiten (Del.icio.us), von Bildern (Flickr) oder von Videos (YouTube). Die Inhalte von Diensten, insofern diese sich komplementär ergänzen, werden gelegentlich als „mash-ups“ (Weiss, 2005, 23) zusammengeführt (beispielsweise Housingmaps.com als „mash-up“ aus Immobilieninformationen aus Craigslist mit Landkarten und Satellitenbildern aus Google Maps). Die Kooperation endet nicht beim Bereitstellen von Content, sondern schließt bei einigen Web-2.0-Services auch die inhaltliche Erschließung des bereitgestellten Wissens mit ein. In Abbildung 1 sehen wir eine „tag cloud“ (Sinclair & Cardew-Hall, 2008) aus den Dingen, die die Nutzer in ihrem Leben noch tun oder erreichen wollen. Die Häu-

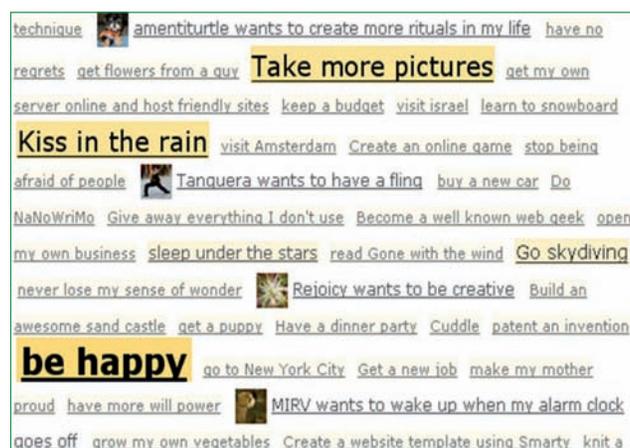


Abbildung 1: Tag Cloud der Homepage „43 Things“. Quelle: <http://www.43things.com>.

figkeit der Worte, der Tags, wird durch die Schriftgröße angedeutet. Am häufigsten wurde *be happy*, am zweithäufigsten *take more pictures* und weniger oft *get a new job* als Lebensziel genannt.

Diese Art der freien Schlagwortvergabe durch jedermann bezeichnen wir als „Folksonomy“, wobei die freien Schlagworte hier „Tags“ genannt werden. Die Indexierung mithilfe von Folksonomies ist demnach „Tagging“. Peter Merholz (2004) bezeichnet das Verfahren als „metadata for the masses“, der Schriftsteller James Surowiecki (2004) nennt es „the wisdom of crowds“ (McFedries, 2006, 80). Folksonomies sind eine Form der Inhaltserschließung im Sinne einer freien Verschlagwortung von Dokumenten im Internet (Furnas et al., 2006). Die Produzenten oder auch die Nutzer der Dokumente werden selbst zu Indexern. Es gibt keine zentrale Kontrolle, die die Tags und die Arbeit der Prosumer prüft. Der Begriff „Folksonomy“ als Kunstwort aus „Folk“ und „Taxonomy“ geht auf einen Beitrag in einem Blog zur Informationsarchitektur zurück, in dem Gene Smith (2004) Thomas Vander Wal zitiert:

„Last week I asked the AlfIA (i.e. the „Asilomar Institute for Information Architecture“, d. A.) member's list what they thought about the social classification happening at Furl, Flickr and Del.icio.us each of these systems people classify their pictures/bookmarks/web pages with tags ..., and then the most popular tags float on the top ... Thomas Vander Wal, in his reply, coined the great name for these informal social categories: a folksonomy. Still, the idea of socially constructed classification schemes (with no input from an information architect) is interesting. Maybe one of these services will manage to build a social thesaurus“.

Smith gebraucht das Wort „classification“ zur Umschreibung von Folksonomies. Dies weist – genau wie Taxonomie – in eine falsche Richtung. Folksonomies sind gerade *keine* Klassifikationen, denn sie arbeiten weder mit Notationen noch mit Relationen. Wichtig erscheint uns der Hinweis Smith', dass man, aufbauend auf Folksonomies, Thesauri kooperativ erstellen kann.

Weite Verbreitung finden Folksonomies bei den Web 2.0-Diensten, sie können aber genauso in firmeneigenen Intranets angewendet werden (Fichter, 2006), z.B. um Corporate Blogs, Podcasts oder Vodcasts (Peters, 2006a; Peters, 2006b; Peters & Stock, 2006), Corporate Bookmarking-Dienste (Millen, Feinberg, & Kerr, 2006) oder Schwarze Bretter (Murison, 2005) zu erschließen. In Spiteri (2006 und 2007) wird der Gebrauch von Folksonomies bei Katalogen öffentlicher

Bibliotheken diskutiert. Smith (2006) und Trant (2006a, 2006b) berichten über den Einsatz dieser Indexierungsmethode in Museen. Heller (2007) stellt die Nutzung von Folksonomies beim Aufbau gemeinschaftlich erstellter Bibliographien dar, ähnlich Hänger und Krätzsch (2007) bei der Erschließung ansonsten unindexierter Dokumente. Stock (2007b) ist ein Plädoyer für ein Mash-up von bewährten „traditionellen“ Indexierungsmethoden und Folksonomies bei kommerziellen Informationsdiensten (siehe auch Peters, 2007)

Broad und Narrow Folksonomies

Mit Thomas Vander Wal (2005) lassen sich zwei Formen von Folksonomies unterscheiden: broad, weite, und narrow, enge, Folksonomies (siehe auch Dye, 2006). In der Broad Folksonomy (Abbildung 2) können viele verschiedene Nutzer (A bis F in der Abbildung) ein Dokument mit Tags versehen. So wird der Dokumentinhalt aus zahlreichen Ansichten mit gleichen, ähnlichen oder völlig verschiedenen Schlagworten beschrieben (1 bis 5). Das bekannteste Web 2.0-Angebot, das diese Art der multiplen Verschlagwortung nutzt, ist der Social Bookmarking-Dienst Del.icio.us. Auch für den Wissenschaftsbereich existieren mit CiteULike, Connotea (Hammond, Hannay, Lund, & Scott, 2005; Lund, Hammond, Flack, & Hannay, 2005) und BibSonomy (Hotho, Jäschke, Schmitz, & Stumme, 2006c) entsprechende Systeme.

In Narrow Folksonomies werden Tags nur einmal pro Dokument vergeben und auch nur einmal erfasst. Deswegen können nur neue Tags zugeordnet werden. Daher besteht keine Möglichkeit, Taghäufigkeiten zu zählen und Verteilungen zu beobachten. Häufig vergibt der Autor (oder der Content Creator) die ersten Tags; manchmal ist es auch anderen Usern erlaubt, weitere Tags hinzuzufügen. Dieses Vorgehen erinnert an die Inhaltserschließung mit den kontrollierten Vokabularien aus Nomenklaturen, Thesauri oder Klassifikationssystemen (Stock & Stock, 2008, Kap. 11 bis 13) durch professionelle Indexer; abgesehen davon, dass in Folksonomies nur unkontrollierte Terme zum Einsatz kommen. Web 2.0-Dienste, die mit Narrow Folksonomies arbeiten, sind u. a. Flickr (Fotos; Beaudoin, 2007), Technorati (Blogeinträge) und YouTube (Videos).

Narrow Folksonomies können keine bestimmten Häufigkeitsverteilungen der Indexierungs-Tags zeigen, da bei ihnen alle Tags gleichwertig sind (alle kommen genau einmal vor). Allerdings ist es möglich, Tagverteilungen aus den Such-Tags zu gewinnen. Die Frage ist hier: Welches Dokument wird mit welchen Tags gefunden? Jedem (angezeigten und angeklick-

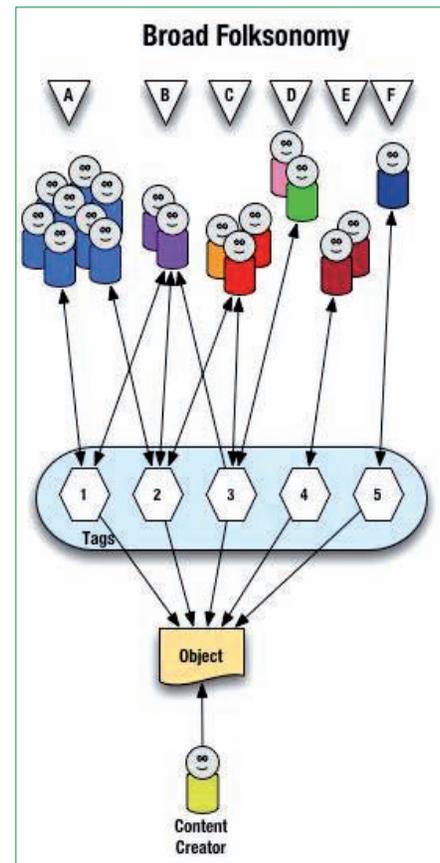


Abbildung 2: Folksonomy mit Mehrfachvergabe gleicher Tags („Broad Folksonomy“).
Quelle: Vander Wal (2005).

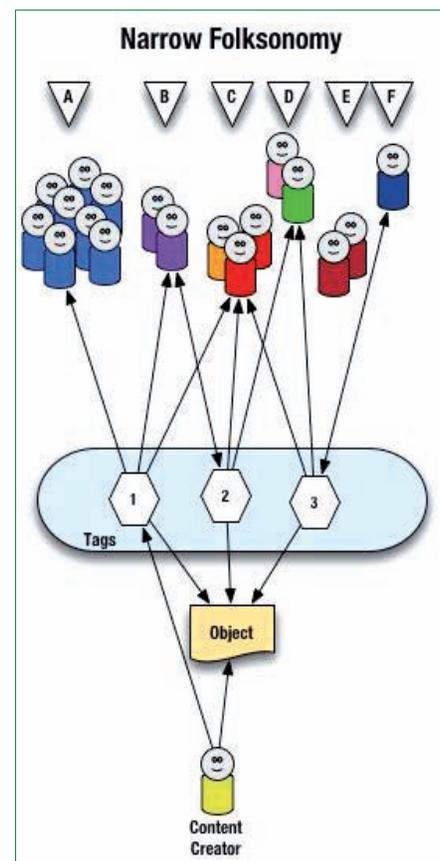


Abbildung 3: Folksonomy mit singulärer Vergabe von Tags („Narrow Folksonomy“).
Quelle: Vander Wal (2005).

ten) Dokument werden diejenigen Terme als Such-Tags zugeordnet, mit denen der jeweilige User erfolgreich das Dokument recherchiert hat. Solche Untersuchungsoptionen bestanden eigentlich schon immer; u. W. fanden sie jedoch noch nie Einsatz.

Im Vergleich zu den Broad Folksonomies werden Narrow Folksonomies i. d. R. mit weniger Tags indexiert. Das ist ein Grund, warum hier kein besonders langer Long Tail wird entstehen können. Die Häufigkeitsverteilungen der Such-Tags können aber durchaus gewisse Regelmäßigkeiten aufweisen.

Häufigkeitsverteilungen der Tags

Vander Wal (2005), Shirky (2005), Munk & Mork (2007) u. a. stellen bei Broad Folksonomies fest, dass die Verteilung der indexierten Tags einer Power Law-Kurve nach Lotka's Gesetz (Egghe & Rousseau, 1990; Egghe, 2005) gleicht. Diese Verteilung folgt der Gleichung

$$f(x) = C / x^a,$$

bei der C eine Konstante, x der Rang des gegebenen Tags und a ein konstanter Wert (normalerweise zwischen 1 und 2) ist. Sollte diese Annahme richtig sein, wären am linken Anfang der Kurve nur wenige Tags mit hohen Werten platziert, das rechte Ende der Kurve würde allerdings aus zahlreichen nahezu gleichhäufigen Tags bestehen und den sog. „Langen Schwanz“, den „Long Tail“, bilden. Wie sieht dies konkret aus?

In Abbildung 4 wird deutlich, dass sich bestimmte Tags durch ihre Häufigkeit von anderen Tags deutlich abgrenzen. Die Tags *politik* und *bundestag* auf der linken Seite der Kurve vereinen fast 80 Prozent der gesamten Taghäufigkeit auf sich, während die rechte Seite der Kurve mit Tags wie *culture*, *institutionen* und *news* zwar eine größere Vielfalt der Content-Beschreibung aufweist, die aber jeweils nicht so häufig genannt werden. Diese Menge an niedrigfrequenten nahezu gleichhäufigen Tags bildet den Long Tail. Nach dem Aufrufen der Webseite am Beginn der Kurve (*politik*, *bundestag*) den Website-Content in einem hohen Maße adäquat und allgemein, die Long Tail-Tags ihn jedoch sehr speziell beschreiben. Die Kurve zur Webseite *www.bundestag.de* und somit die Tagverteilung folgt hier dem Power Law mit einem Exponenten von ungefähr $a = 1$. Eine idealtypische Power Law-Verteilung zeigt Abbildung 5.

In einer Untersuchung gemeinsamen Auftretens von Tags („co-tags“ in Analogie zu den „co-citations“ der Zitationsanalyse) können Lux, Granitzer und Kern (2007) zeigen, dass rund 80% der

Tags derart mit anderen Tags vorkommen, dass diese anderen Terme einem Power Law folgen. Wie lässt sich dies erklären? Einen sehr interessanten theoretischen Ansatz legen Cattuto (2006) sowie Cattuto, Loreta und Pietronero (2007) vor. Auch hier geht es – wie bei Lux et al. – um das Auftreten der Tags auf einer Makroebene (der gesamten Datenbank oder experimentell abgetrennter Teile davon) und nicht – wie in unserem Beispiel (Abbildung 4) – um die Mikroebene, der Zuordnung gewisser Tags zu konkreten Dokumenten. Cattuto et al. wenden den Ansatz der „Semiotics Dynamics“ auf die Kreation neuer bzw. dem Verwenden alter Tags in einer Folksonomy an. „Folksonomies ... do exhibit dynamical aspects also observed in human language, such as the emergence of naming conventions, competition between terms, takeovers by neologisms, and more“ (Cattuto, 2006, 33). Das Erklärungsmodell hat seinen Ur-

sprung im Yule-Simon-Ansatz, der Wahrscheinlichkeiten für das Auftreten von Worten in Texten thematisiert. An jeder Stelle eines Textes gilt: Ein bestimmtes Wort hat die Wahrscheinlichkeit p , ein neues Wort zu sein (d.h., noch nicht im bisherigen Text vorzukommen), oder es hat die Wahrscheinlichkeit $1 - p$, eine Kopie eines bereits vorhandenen Wortes zu sein. Die Größe von $1 - p$ hängt davon ab, wie häufig das gegebene Wort schon in Text vorgekommen ist, wobei ein einfacher Zusammenhang hergestellt wird: Je häufiger das Wort bereits vorkommt, desto wahrscheinlicher ist es, dass es erneut Anwendung findet. Der Yule-Simon-Ansatz ist damit eine auf Texte angewandte Variante des altbekannten „Erfolg gebiert Erfolg“ oder des „Wer hat, dem wird gegeben“. Um dieses Modell auf Folksonomies anzuwenden, muss man nur statt von „Worten in Texten“ von „Tags in Datenbanken“ sprechen. Je häufiger ein Tag bereits in einem Informationsdienst auftritt, desto höher wird

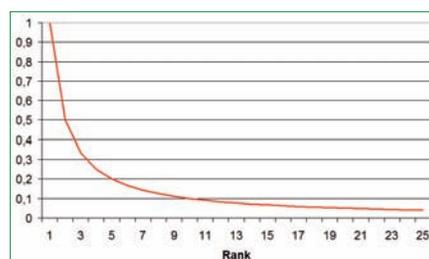


Abbildung 5: Power Law.

seine Wahrscheinlichkeit, dass er auch erneut zur Indexierung herangezogen wird. Cattuto et al. verfeinern den Yule-Simon-Ansatz um eine Zeitkomponente derart, dass neuere Tags eine größere Indexierungswahrscheinlichkeit haben als alte. „It seems more realistic to assume that users tend to apply recently added tags more frequently than old ones, according to a skewed memory kernel“ (Cattuto, 2006, 35). Abbildung 6 zeigt das Erklärungsmodell im Überblick. Die Wahrscheinlichkeit p drückt aus, dass ein völlig neuer Tag zur Indexierung herangezogen wird, $1 - p$ entsprechend, dass ein alter gewählt wird. Letztere Wahrscheinlichkeit hängt von der Zeit ab, zu der der (alte) Tag genutzt worden ist. Die Zeitangaben sind im Gedächtnis Q gespeichert und steuern eine Gewichtung nach der Regel „je neuer, desto wahrscheinlicher“.

seine Wahrscheinlichkeit, dass er auch erneut zur Indexierung herangezogen wird. Cattuto et al. verfeinern den Yule-Simon-Ansatz um eine Zeitkomponente derart, dass neuere Tags eine größere Indexierungswahrscheinlichkeit haben als alte. „It seems more realistic to assume that users tend to apply recently added tags more frequently than old ones, according to a skewed memory kernel“ (Cattuto, 2006, 35). Abbildung 6 zeigt das Erklärungsmodell im Überblick. Die Wahrscheinlichkeit p drückt aus, dass ein völlig neuer Tag zur Indexierung herangezogen wird, $1 - p$ entsprechend, dass ein alter gewählt wird. Letztere Wahrscheinlichkeit hängt von der Zeit ab, zu der der (alte) Tag genutzt worden ist. Die Zeitangaben sind im Gedächtnis Q gespeichert und steuern eine Gewichtung nach der Regel „je neuer, desto wahrscheinlicher“.

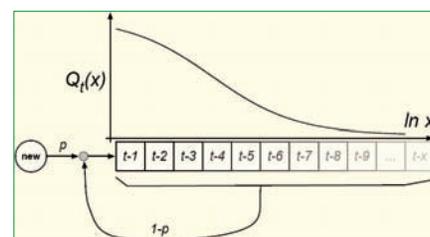


Abbildung 6: Ein Yule-Simon-Prozess der Auswahl von Indexierungs-Tags mit Berücksichtigung des Alters der Tags. Quelle: Cattuto, Loreta, & Pietronero (2007, Fig. 5).

Für die Makroebene erscheint der Ansatz von Cattuto und des Yule-Simon-Modells vielversprechend. Offen ist, ob sich diese Theorie auch auf die Mikroebene übertragen lässt. Wenn ja, müssten sich die Tags auf der Dokumentenebene auch nach dem Power Law verteilen. Dies scheint aber nicht – oder zumindest nicht immer – zuzutreffen. Lux et al. (2007) konnten (auf der Makroebene) nur 80% der Co-Tag-Verteilungen als Power Law auszeichnen. 20% folgen offenbar anderen Regeln. Betrachten wir ein zweites Beispiel einer dokumentspezifischen Tag-Verteilung (Abbildung 7)!

Wie man sieht, folgt die Tagverteilung hier nicht dem Power Law. Es scheint, als würden sich hier zwei „Lange

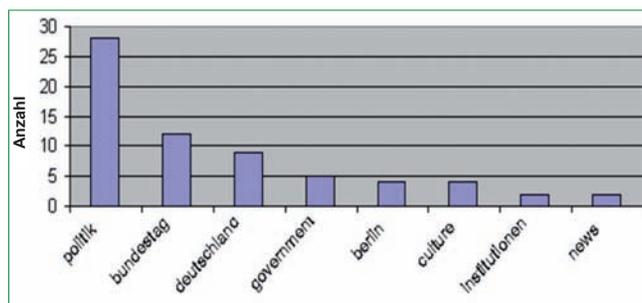


Abbildung 4: Tagverteilung zur Webseite *www.bundestag.de*. Quelle: <http://del.icio.us> (Stand: November 2007; Anzahl indexierender User: 60).

Schwänze“ bilden: rechts der bekannte „Long Tail“ und links ein „Long Trunk“, der „Lange Rüssel“. Die Tags grenzen sich in ihren Häufigkeitsverteilungen nicht stark genug voneinander ab; es werden also viele Tags ähnlich häufig genannt. Der lange Rüssel nimmt bei www.asis.org die Tags *associations*, *library*, *information*, *ia*, *technology* und *professional* auf. Hiernach hat die Verteilung einen Wendepunkt, an den sich der lange Schwanz anschließt. Die Tags im langen Rüssel beschreiben den Inhalt der Webseite – anderes als im Fall der ersten zwei, drei Tags bei einer Power Law-Verteilung – nur wenig treffend. Die Tags *associations* und *technology* sind äußerst allgemein, und auch die restlichen Terme geben nur vage Hinweise auf den Content. Hier hat die kollektive Intelligenz offenbar Schwierigkeiten, ein „wahres“ Bild des Dokuments zu zeichnen. Es scheint jedoch auch invers-logistische Verteilungen zu geben, bei denen die Tags links vom Wendepunkt eine Webseite adäquat beschreiben. Margaret Kipp (zitiert in Peters, 2008) kann sogar demonstrieren, dass bei ihrem Beispiel die Tags im langen Rüssel im Zeitverlauf konstant bleiben. Für Kipp haben diese Power Tags den Charakter von kontrolliertem Vokabular – lediglich durch die Community der Prosumer „geprüft“. Stock (2006) konnte in einer theoretischen Studie zeigen, dass es mindestens zwei verschiedene Verteilungen von Relevanz gibt: die bekannte Power Law-Verteilung und die invers-logistische Verteilung (siehe Abbildung 8). Die invers-logistische Verteilung hat viele mehr oder minder relevante Tags im „Long Trunk“ und zusätzlich ebenfalls einen Long Tail. Dabei folgt sie der Formel

$$f(x) = e^{-C'(x-1)^b}$$

bei der *e* die Euler'sche Zahl und *x* der Rang des Tags ist. *C'* (im Beispiel von Abbildung 8 ist es 0,1) ist eine Konstante und der Exponent *b* ist stets ungefähr 3. In vielen Fällen ist der „Long Trunk“ kürzer als der „Long Tail“.

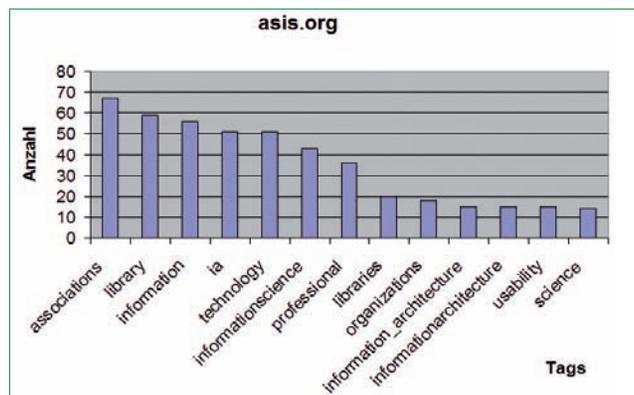


Abbildung 7: Tagverteilung zur Webseite www.asis.org. Quelle: <http://del.icio.us> (Stand: November 2007; Anzahl indexierender User: 293).

Einige Untersuchungen scheinen von der stillschweigenden Voraussetzung auszugehen, es gäbe ausschließlich die Power Law-Verteilung. Da bei beiden Verteilungskurven ein langer Schwanz auftritt, ist in der Tat eine Verwechslungsmöglichkeit gegeben. Man muss demnach besonderes Augenmerk auf den Anfang der Verteilung legen. Das entscheidende Kriterium, das ein Power Law ausschließt, ist das Vorliegen des langen Rüssels. Approximiert man eine Kurve unter Auslassen der ersten Rangplätze (besonders eklatant beispielsweise bei Capocci und Caldarelli (2007, Fig. 3) oder bei Cattuto (2006, Fig. 5)), so sind Fehlzuschreibungen nicht auszuschließen. Bei der Beobachtung von Tags sollten beide Verteilungen im Auge behalten werden, da diese bei der Entwicklung und Erstellung von Information Retrieval-Werkzeugen für getaggte Dokumente wieder zur Sprache kommen müssen. Es haben nämlich die Tags im langen Rüssel bei der invers-logistischen sowie die ersten *n* Tags beim Power Law eine Sonderstellung, die wir sowohl für eine spezifische Rechercheoption als auch für das Relevance Ranking ausnutzen wollen.

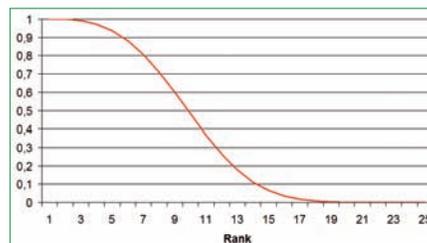


Abbildung 8: Invers-logistische Verteilung. Quelle: Stock (2006), S. 1127 (modifiziert).

Nicht unerwähnt bleiben sollen Dienste, die informatrische Analysen von Tags anbieten. So zeigt beispielsweise Technorati Zeitreihen von Blogbeiträgen in Bezug auf Themen (für ein Beispiel siehe Stock & Weber, 2006), während Cloudalicious Zeitreihen von Tags zu bestimmten URLs in Delicio.us generiert (Russell, 2006). Hiermit öffnet sich ein Forschungsfeld für die Analyse von Regelmäßigkeiten zeitlicher Tagverteilungen.



Abbildung 9: Getaggte Foto in Flickr. Quelle: Flickr.com.

Ofness – Aboutness – Ikonologie

Wir kommen zu einem nicht unerheblichen Problem bestimmter Anwendungen von Folksonomies, nämlich zur kollaborativen Erschließung nicht-textueller Dokumente. Manche Web 2.0-Dienste verzeichnen und sammeln Fotos (z.B. Flickr) oder Videos (z.B. YouTube). Nicht-textuelle Dokumente (und auch einige fiktionale Texte) haben nicht nur eine Ebene des jeweils Thematisierten, im Gegensatz etwa zu wissenschaftlichen Zeitschriftenartikeln. Fachliche textuelle Dokumente tragen auf einer einzigen semantischen Ebene Content, gewöhnlich mit „Aboutness“ beschrieben (Stock & Stock, 2008, 31-36). Nach Panofsky (2006) gibt es für Kunstwerke drei verschiedene semantische Stufen der Interpretationsmöglichkeiten (Smith, 2006). Wir wollen Panofskys Theorie an einem Beispiel aus Flickr erklären (siehe Abbildung 9). Das Foto zeigt eine Gruppe von Skulpturen („Mönch – Arzt – Händler“), die von der deutschen Künstlerin Katharina Fritsch erstellt wurden und im Düsseldorfer Museum K21 ausgestellt werden. Auch die zu dem Foto annotierten Tags sind aufgeführt. Panofskys grundlegendes semantisches Level wird prä-ikonographisch genannt und beschreibt die Welt der primären bzw. natürlichen Objekte. Auf dem Foto sind drei Figuren zu sehen: eine rote Figur mit einem Pferdefuss, eine weiße mit einem Totenkopf und eine schwarze bekleidet mit einer Kutte. Für dieses Level der Interpretation benötigt der Interpret keine weiteren Kenntnisse als die aus alltäglicher Erfahrung. In der Informationswissenschaft nennt man das Ergebnis der prä-ikonographischen Interpretation die Ofness eines Dokuments (Lancaster, 2003; Layne, 2002; Markey, 1986; Shatfort, 1986; Turner, 1995). Die Ofness-Tags in diesem Beispiel sind (interessanterweise nur in englischen Termen ausgedrückt) *black*, *white* und *red*. Das ikonographische Level umfasst die sekundären Objekte; dies ist die Welt

der Bilder, Anekdoten und Allegorien (Panofsky, 1975, 50). Der Interpret muss hier einiges an Erfahrung und Wissen im soziokulturellen sowie thematischen Umfeld des Kunstwerkes für die Interpretation mitbringen. Dieses Level bezeichnet man als die Aboutness des Dokuments. Aboutness-Tags sind *monk*, *doctor* und *dealer* sowie die deutschen Übersetzungen davon. (Nunmehr arbeitet unser Tagger mit englischen und mit deutschsprachigen Worten.)

Das dritte Level bezeichnet Panofsky als ikonologisch; es besteht aus dem „wahren“ Inhalt, aus der Welt der symbolischen Werte (Panofsky, 1975, 50). Umfassende Kenntnisse des thematischen Gebiets sind hier für eine adäquate Interpretation notwendig. Informationswissenschaftler berücksichtigen normalerweise dieses Level nicht in der Inhaltserschließung; User der Web 2.0-Dienste können aber den semantischen Gehalt ihrer Tags selbst bestimmen. In dem Beispiel ist nur ein, allerdings sehr schwacher, ikonologischer Tag zu finden: *art* (bzw. *Kunst*). Eine stärkere ikonologische Beschreibung könnte z.B. *unheilige Dreifaltigkeit* sein.

Darüber hinaus werden auch nicht-inhaltliche Informationen vom Nutzer zum Dokument hinzugefügt, wie z.B. der Name des Museums, *K21*, oder der Name der Künstlerin, *Katharina Fritsch*. Diese Arten von Tags beschreiben in der Diktion von Peter Ingwersen (2002) die Isness des Dokuments. Katharin Fritsch *ist* (is) die Künstlerin des Objekts; *K21 ist* (is) das ausstellende Museum. Folksonomies machen keinen Unterschied zwischen den verschiedenen Beschreibungen- und Interpretationsebenen; sie fließen zu einem einzigen semantischen Level zusammen.

Nach Ingwersen (2002) werden vier Aspekte von Aboutness unterschieden: a) die Autor-Aboutness beschreibt den Content, wie er ist, b) die Indexer-Aboutness die zweckgebundene Interpretation des Contents, c) die Anfrage-Aboutness den Content als Ausdruck der Suchanfrage und letztlich d) die User-Aboutness, die Interpretation des Contents durch den Nutzer. Das Taggen von Fotos, Videos usw. ist als nutzerzentriert zu klassifizieren, d.h. beim Level 1 handelt es sich um die User-Ofness, Level 2 umfasst die User-Aboutness und Level 3 die User-Ikonologien; hinzu tritt als vierte Ebene die Isness aus Nutzersicht. All diese verschiedenen Stufen von Interpretationen werden in einer Tag Cloud zusammengefasst und nicht unterschieden. Trotz dieser Problematik stellt Smith in Bezug auf das Taggen von Museumsobjekten durch die Nutzer fest: „Tagging has the potential of increasing access to artwork images and records dramatically for searchers of all levels of expertise“ (Smith, 2006, 13). Studien über den Ein-

satz neuer nutzerzentrierter Dokumentationsmethoden in Museen zeigen, dass die meisten der vom Nutzer kreierten Schlagworte nicht in der bisher verwendeten Wissensordnung vorkommen, sondern neu sind (Trant, 2006b, 21) und dass der Einsatz von Folksonomies neue Möglichkeiten schafft, ein Kunstwerk zu finden (Trant, 2006a).

In zahlreichen informationswissenschaftlichen Veröffentlichungen wird das Thema Inhaltserschließung von Fotos und Bildern behandelt (z.B. Rasmussen, 1997), allerdings nicht mit Bezug auf Folksonomies und Web-2.0-Dienste, sondern beispielsweise auf Pressebilder.

Ornager (1995, 214) hat dazu herausgefunden, dass sich die Inhaltsanalyse von Fotos und Bildern auf hauptsächlich fünf Kategorien stützt: „named person (who), background information about the photo (when, where), specific events (what), moods and emotions (shown or expressed), size of photo.“ Markkula und Sormunen (2000, 270 f.) stellen fest, dass „(t)he most often used index terms referred to specifics, i.e. to individual objects, places, events and linear time, and to the theme of the photo“. Zur Inhaltserschließung und zur Beschreibung der abgebildeten Objekte verwenden die Indexer häufig Nomen, jedoch werden Handlungen und Ereignisse auch mit Verben beschrieben („Czechs celebrate the victory...“; Markkula & Sormunen, 2000, 271). Eine wichtige Beobachtung ist: „(T)he output of the indexing process seemed to be quite inconsistent“ (Markkula & Sormunen, 2000, 273). Diese Indexierungsinkonsistenz dürfte in der doch stark subjektiven Färbung der Bildbetrachtung und -auswertung begründet liegen (Neal, 2007).

Die Vorteile von Folksonomies

Folksonomies sind eine neue Methode, um den Inhalt von Dokumenten zu erfassen. Die Prosumer – oder die „Akteure“ im Sinne der Analyse kognitiver Arbeit (Fidel, 2006) – indexieren selbst das Dokument. Dabei spiegelt die Folksonomy die Sprache und das Wissen der Prosumer authentisch wider (Quintarelli, 2005). Diese Art der Inhaltserschließung führt zu einer Vielzahl an „multiple interpretations“, unterschiedlichen (und manchmal unvereinbaren) Meinungen und „multicultural views“ von ein und

derselben Information (Peterson, 2006). Diese „shared intersubjectivities“ ermöglichen es dem Nutzer „to benefit, not just from their own discoveries, but from those of others“ (Campbell, 2006, 10).

Folksonomies komplettieren die Werkzeuge und Methoden der Wissensrepräsentation, wie dies in Abbildung 10 skizziert ist. Wir unterscheiden drei Gruppen von Akteuren, die Dokumente indexieren: Autoren, professionelle Indexer und Nutzer (Kipp, 2006b). Hierbei werden unterschiedliche Wege der Indexierung beschriftet, die wahrscheinlich auch unterschiedliche Merkmale der Dokumente

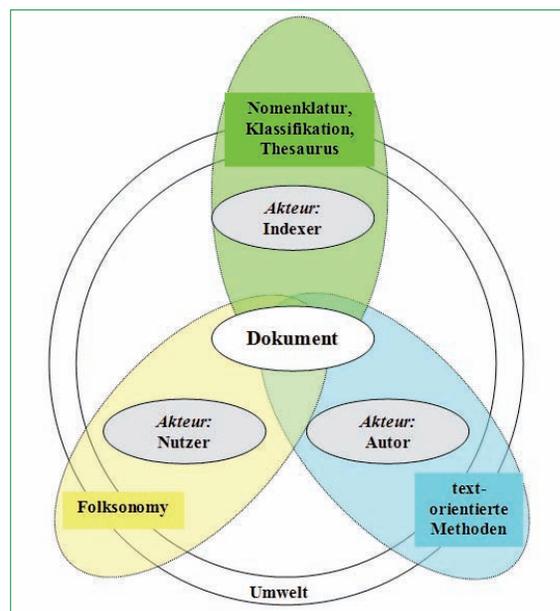


Abbildung 10: Methoden der Wissensrepräsentation und deren Akteure. Quelle: Stock & Stock (2008, S. 41).

jeweils in den Vordergrund rücken. Textorientierte Methoden machen in einem großen Maße Gebrauch von der Sprache des Autors, z.B. bei der Textwortmethode oder bei der Zitationsindexierung. Im Gegensatz zu diesen Methoden berücksichtigen Folksonomies nicht nur die Sprache des Autors, sondern auch die Sprache der Nutzer. Wissensordnungen wie Nomenklaturen, Thesauri oder Klassifikationssysteme benötigen für ihren Gebrauch Interpreten oder Übersetzer. Zum einen sind dies die Experten, die das Vokabular erstellen, und zum anderen die, die das Vokabular zur Inhaltserschließung benutzen. Die Entwickler von kontrollierten Vokabularien müssen zunächst die „literature, needs, actors, tasks, domains, activities, etc.“ (Mai, 2006, 17) analysieren, um daraus das kontrollierte Vokabular zu erstellen. Dies ist ohne Zweifel eine zeitaufwändige und teure Prozedur. Dagegen sind Folksonomies günstig im Unterhalt, denn die Pflege und Indexierung geschieht durch die freiwillige und kostenlose Zusammenarbeit vieler Nutzer. „Tagging has dramatically lower costs because there is

no complicated, hierarchically organized nomenclature to learn. User simply create and apply tags on the fly" (Wu, Zubair, & Maly, 2006, 111). Darüber hinaus gibt es eine hohe Bereitschaft seitens der User, selbst zu taggen und so den Taggingprozess in der Community anzuregen und zu verstärken (Fichter, 2006).

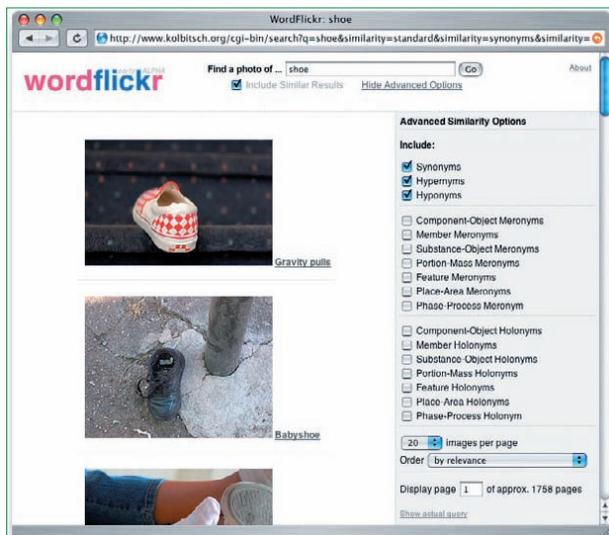


Abbildung 11: WordFlickr: Synthese aus Web-2.0-Dienst und Thesaurus. Quelle: Kolbitsch (2007a, S. 85).

Aus praktischen Gründen ist es wohl unmöglich, alle URLs, Blogbeiträge, Fotos und Videos im World Wide Web intellektuell mit kontrollierten Vokabularien zu erschließen. Algorithmische Suchmaschinen wie Google oder Yahoo! können textuelle Dokumente automatisch indizieren. Für Dokumente wie Fotos oder Videos gestaltet sich eine rein automatische, content-basierte Erschließung jedoch schwierig, da sie keine Metadaten enthalten.

Welchen Nutzen bewirken Folksonomies? Wir wollen an dieser Stelle einige kurze Überlegungen anstellen.

1) Die Entwicklung und die Pflege von bestehenden kontrollierten Vokabularien kann von Folksonomies profitieren (Aurnhammer, Hanappe, & Steels, 2006; Christiaens, 2006; Gendarmi & Lanubile, 2006; Macgregor & McCulloch, 2006; Mika, 2005; Zhang, Wu, & Yu, 2006). Die Tags, ihre Häufigkeit und ihre Verteilungen können als Quelle für neue kontrollierte Terme, für Term-Modifikationen und Term-Löschungen dienen. Diese „bottom-up categorization“ (Vander Wal, 2004) garantiert schnelle Reaktionszeiten bezüglich neuer Themen und Innovationen in der Wissensdomäne und ist damit ein wichtiges Instrument sowohl zum Neuaufbau als auch zur Pflege von Nomenklaturen, Klassifikationssystemen und Thesauri. Doch inwieweit können Folksonomies praktisch zur Entwicklung von kontrollierten Vokabularien, von Ontologien bis hin zum

Semantischen Web beitragen (Al-Khalifa, 2007; Specia & Motta, 2007)? Die bis jetzt entwickelten Modelle arbeiten hauptsächlich mit Co-Occurrences (Schmitz, 2006), einfachen Cluster-Algorithmen (Grahl, Hotho, & Stumme, 2007) oder dem Vektorraum-Modell (Dimensionen: Dokumente, Vektoren: Tags, Tagähnlichkeit:

Cosinus) (Heymann & Garcia-Molina, 2006). Die entstandenen Ähnlichkeitswerte können als Grundlage für einen Ähnlichkeitsgraphen dienen, bei dem die Position des Tags Auskunft über die hierarchische Lokalisierung in einer „latent hierarchical taxonomy“ (Heymann & Garcia-Molina, 2006, 4) geben kann.

2) Folksonomies ergänzen kontrollierte Vokabularien in der Recherchepraxis. Kann man bei Folksonomies und Ontologien überhaupt von Rivalen sprechen? Für Tom Gruber (2006, 994) ist dies „nonsense, and it is time to embrace a

unified view“. Die Grundidee hierbei ist es, Tags in ihrem semantischen Umfeld darzustellen, aus dem der Nutzer zusätzliche Suchargumente gewinnen kann. Da Folksonomies nie explizit paradigmatische Umfeldterme nennen, müssen diese aus anderen Wissensordnungen gewonnen werden. Infrage kommen alle fachlichen wie linguistischen Begriffsordnungen. WordNet (Miller, 1998) konnte bereits erfolgreich bei Daten aus Delicio.us (Laniado, Eynard, & Colombetti, 2007) sowie bei Flickr (Kolbitsch, 2007b) zum Einsatz kommen. Abbildung 11 zeigt die experimentelle Oberfläche von WordFlickr, einem Mash-up aus WordNet und Flickr.

3) Journalisten oder Wissenschaftler müssen an der Hürde der Gatekeeper und ihrer Qualitätskontrolle vorbei, wollen sie ihre Werke veröffentlichen. Es stellt sich die Frage, ob es solche Gatekeeper auch im Internet gibt. Clay Shirky (2005) sagt dazu: „The Web has an editor, it's everybody“. Eine Qualitätskontrolle des Contents findet statt – jedoch erst nach seiner Veröffentlichung. Je mehr Nutzer ein Dokument taggen, desto mehr Relevanz scheint dieses Dokument für sie zu haben. Ist dies aber eine ernstzunehmende Qualitätskontrolle? Wird etwas zu „geprüfter“ Qualität, nur weil viele Leute dies so sehen? (Wenn viele Studenten bei einer Mathematiklausur die gleiche – falsche – Lösung bringen, wird diese nicht dadurch qualitativ wertvoll,

sondern bleibt falsch.¹) Quantität bedeutet nicht Qualität. Andererseits weist es in eine bestimmte Richtung, wenn viele Nutzer ein Stück Information mit *stupid* und ein anderes mit *cool* taggen. Dieser Content könnte für das Relevance Ranking verwertet werden.

4) Das Retrieval mit Folksonomies kann auf zweifache Weise erfolgen: a) der Nutzer gibt einfach mögliche Tags in die Suche ein oder b) er nutzt die Tags, um sich zum gewünschten Dokument „durchzuhangeln“. Dieser letzte Punkt ist für Adam Mathes (2004) die zentrale Eigenschaft und Stärke von Recherchen mit Folksonomies – Serendipity. „The long tail paradigm is also about discovery of information, not just about finding it“ (Quintarelli, 2005). Das Suchen mit Tags ist für Laien viel einfacher und schneller als die Recherche mit elaborierten Information Retrieval-Werkzeugen, wie z.B. der Internationalen Patentklassifikation (IPC). Einige Nutzer verzichten sogar völlig auf die Eingabe von Suchargumenten und klicken sich durch die Tag Clouds zur gewünschten Information durch (Sinclair & Cardew-Hall, 2008).

5) Es gibt Tags, die in Folksonomies zum ersten Mal auftauchen, sog. Neologismen. In Mathes (2004) werden die Beispiele *sometait hurts* (für „so meta it hurts“) und *flicktion* diskutiert. „Although small, there is a quick formation of new terms to describe what is going on, and others adopting that term and the activity it describes“ (Mathes, 2004). Dieser unerwartete und unvorhersehbare Gebrauch von Sprache und Tags reflektiert deutlich die „communication and ad-hoc group formation facilitated through metadata“ (Mathes, 2004).

6) Betrachtet man alle Dokumente, Tags und Nutzer als Knoten in einem Netzwerk, so kann man leicht die Communities identifizieren, die ein Thema oder ein Interessensgebiet teilen: „Collaborative tagging systems have the potential of becoming a technological infrastructure for harvesting social knowledge“ (Wu, Zubair, & Maly, 2006, 114).

7) Auf der Basis gemeinsam verwendeter Tags oder gemeinsam indizierter Dokumente wird es möglich, Folksonomies für Zwecke von impliziten kollaborativen Recommendersystemen zu nutzen (Diederich & Iofciu, 2006; Niwa, Doi, & Honiden, 2006).

8) Ein weiterer Vorteil von Folksonomies ist (auch wenn es dafür noch keine wissenschaftliche Evidenz gibt), dass sie dem Nutzer die Schwierigkeiten, aber

¹ Dieses illustrative Beispiel verdanken wir unserem Studenten Peter Bucker.

auch Nützlichkeit von Inhaltserschließung und Indexierungsmethoden näher bringen und ihn dafür sensibilisieren.

Wir wollen die Vorteile der Folksonomies noch mal zusammenfassen. Folksonomies:

- spiegeln die Sprache der Nutzer authentisch wider,
- erlauben verschiedene Interpretationen,
- sind eine günstige Form der Inhaltserschließung,
- sind die einzige Möglichkeit, Massen-Information im Web zu erschließen,
- sind Termquellen für die Entwicklung und Pflege von Ontologien und kontrollierten Vokabularen,
- geben die Qualitätskontrolle an die Nutzer weiter,
- erlauben konkretes Suchen und Browsing,
- berücksichtigen Neologismen,
- tragen dazu bei, Communities zu identifizieren,
- geben eine Basis für Recommendersysteme,
- sensibilisieren Nutzer für die Inhaltserschließung.

Die Nachteile der Folksonomies

Es resultieren zahlreiche Probleme aus dem Fehlen eines kontrollierten Vokabulars. „Lack of precision is a function of user behaviour, not the tags themselves“, wiegelt Shirky (2004) ab. Nichtsdestotrotz findet man in Folksonomies oftmals Uneindeutigkeiten, wie verschiedene Wortformen, Nomen in Plural, Nomen in Singular oder Abkürzungen, die das Suchen und Finden von Dokumenten erschweren. Weitere Beispiele präsentiert der Web 2.0-Dienst Del.icio.us (siehe oben Abbildung 7). Da hier jeweils nur ein Wort als Tag eingegeben werden kann, werden Phrasen (*informationarchitecture*) aus zwei oder mehr Wörtern kreierte oder einzelne Wörter mit einem Unterstrich (*information_architecture*) verbunden. Abkürzungen (*ia*) finden Verwendung; Singular- (*library*) wie Pluralformen (*libraries*) kommen nebeneinander vor. Es fehlt die Zusammenfassung von Synonymen, auch über Sprachgrenzen hinweg, und die Trennung von Homonymen. Daneben finden sich viele Rechtschreib- oder Tippfehler und unterschiedliche Schreibweisen. Daher lassen sich Folksonomies auch als unstrukturierte Liste von Schlagwörtern bezeichnen (Reamy, 2006). „Currently most users don't give much thought to the way they tag resources, and bad or 'sloppy' tags are ten-a-penny in Folksonomies“ (Guy & Tonkin, 2006). Das Tagging-Verhalten der User ist also unberechenbar und nicht vorhersehbar, was das Information Retrieval drastisch erschwert. Ein paar

Zahlen sollen dies verdeutlichen: Nach Guy und Tonkin (2006) sind fast 40 Prozent der Flickr-Tags und 28 Prozent der Del.icio.us-Tags „either misspelt, from a language not available via the software used, encoded in a manner that was not understood by the dictionary software, or compound words consisting of more than two words or a mixture of languages“.

Die Prosumer, die ein Dokument erschließen, handeln in unterschiedlichen Kontexten und aus verschiedenen Motivationen: Ein Nutzer annotiert Tags zu einem Dokument aus beruflichen Hintergründen, ein anderer hat sein Freizeitvergnügen im Hinterkopf. Es fehlt daher ein einheitliches, grundlegendes Level der Inhaltserschließung. „There is [...] systematic variation across individuals in what constitutes a basic level“ (Golder & Huberman, 2006, 200). Erfahrungen, sprachliche Ausdrucksstärke, kognitive Fähigkeiten und Motivationen sind alles Faktoren, die den Grad der Inhaltserschließung bestimmen und beeinflussen. Darüber hinaus muss berücksichtigt werden, dass „a significant amount of tagging, if not all, is done for personal use rather than public benefit“ (Golder & Huberman, 2006, 207).

Web 2.0-Services finden weite Verbreitung und erfreuen sich wachsender Popularität – und das auf der ganzen Welt. Die Problematik, die hier auftaucht, betrifft die Vielsprachigkeit unter den Usern. Das „merging of languages“ (Gordon-Murnane, 2006) führt zu Problemen in der intersprachlichen Synonymie oder Homonymie (z.B. deutsch: *Gift* und englisch: *gift*) und schwächt die Präzision der Suchresultate bzw. zeigt u. U. falsche Ergebnisse. Darüber hinaus gibt jeder Nutzer seine Tags vornehmlich (auch) in seiner Muttersprache ein und kann wahrscheinlich nur einen begrenzten Teil anderssprachiger Tags verstehen – und somit auch nur nach einem begrenzten Teil in der gesamten Datenbank suchen und darauf zugreifen.

Folksonomies machen im Gegensatz zu Thesauri und Klassifikationssystemen keinen Gebrauch von Relationen zwischen den Begriffen, sog. paradigmatischen Relationen, wie z.B. Äquivalenzrelation, Hierarchierelation oder Assoziationsrelation. Jedoch bestehen in der aktuellen Co-Occurrence von mehreren Tags zu einem Dokument zwischen diesen syntagmatische Relationen (Stock, 2007a, 451). Wir wollen dies an einem Beispiel verdeutlichen (Abbildung 12).

Die Tags *UK* und *england* sind durch eine geographische Meronymie verbunden, während die Tags *seat* und *stadium* durch eine Komplex-Komponenten-Relation gekennzeichnet

net sind. Diese syntagmatischen Relationen sind eigentlich paradigmatisch und könnten zur Suchfrageerweiterung oder -einschränkung genutzt werden (Weller & Peters, 2007; Peters & Weller, 2008). Jedoch sind diese nützlichen Relationen in der Tagliste verborgen und werden nicht von Folksonomies eingesetzt.

Professionell erstellte Metadaten werden in verschiedene Felder segmentiert, z.B. in *Dokumenttyp*, in *Autor* oder in die *Notationen eines Klassifikationssystems*. Die Inhaltserschließung berücksichtigt hier die formalen Aspekte eines Dokuments und die Aspekte der Aboutness und unterscheidet sie voneinander. Folksonomies vernachlässigen diese strikte Trennung von unterschiedlichen Metadaten, machen aber sehr wohl Gebrauch von Tags, die nicht nur den Inhalt des Dokuments beschreiben, wie z.B. einen Namen für den Besitzer des Dokuments oder den Hinweis, dass es sich um ein Buch handelt. Daneben werden auch Wertungen vorgenommen (*stupid*). Diese user-spezifischen Tags beschreiben oder evaluieren ein Dokument nur aus der Sicht des einen Users, sodass diese Tags „are virtually meaningless to anybody except their creators“ (Pluzhenskaia, 2006, 23). Zahlreiche andere Tags können performativ genannt werden, d.h. dass sie als Ausdruck einer geplanten oder erledigten Aktivität annotiert werden, z.B. *toread* oder *todo* (Kipp, 2006). Auch synkategorische Tags kommen vor, d.h. sie können nur im Kontext verstanden werden. Ein typisches Beispiel für einen synkategorischen Tag ist *me* bzw. *ich*, wenn er zur Beschreibung der Eigenaufnahme des Fotografen annotiert wird. Es ist auch nicht auszuschließen, dass einige der Tags absichtlich als Spam-Tags von Nutzern verwendet werden. „(I)t involves an unethical user who propagates [...] tags in order to corrupt a system“ (Kroski, 2005). Beim Tagging von Fotos und Videos kommt die Schwierigkeit hinzu, dass sich unterschiedliche Levels der Indexierung bezüglich Ofness, Aboutness, Ikonologie und Isness miteinander vermischen, sich nicht unterscheiden

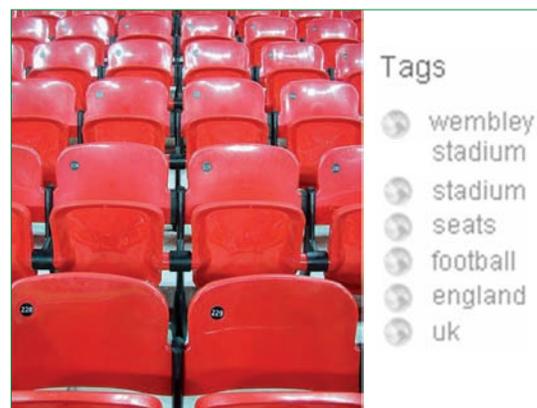


Abbildung 12: Syntagmatische und paradigmatische Relationen zwischen Tags. Quelle: Flickr.com.

den lassen und in ihrer Inter- und Intra-indexerkonsistenz erheblich schwanken. Wir wollen die Nachteile der Folksonomies noch einmal kurz festhalten:

- fehlendes kontrolliertes Vokabular,
- verschiedene Levels der Indexierung,
- Vermischung von Sprachen,
- versteckte paradigmatische Relationen bleiben ungenutzt,
- fehlende Trennung von formalen bzw. bibliographischen Tags und Aboutness-Tags,
- Spam-Tags, nutzer-spezifische Tags und andere uneindeutige Schlagworte,
- Verschmelzung von Ofness, Aboutness, Ikonologie und Isness.

NLP – eine Problemlösung für Folksonomies?

Um die oben genannten Probleme zu lösen, könnten sich zwei Vorgehensweisen etablieren, die sich gegenseitig ergänzen. Zum einen kann man sich auf die Indexer bzw. Prosumer konzentrieren und versuchen, diese zu einer besseren „tag literacy“ (Guy & Tonkin, 2006) zu erziehen. Dazu muss man zunächst das Nutzerverhalten bezüglich Tagging beobachten (siehe z.B. Bar Ilan et al.; 2006; Lin et al., 2006; Morrison, 2007; Veres, 2006a oder Winget, 2006). Die Nutzerforschung steht hinsichtlich der Folksonomies allerdings erst am Anfang ihrer Bemühungen.

Um die Nutzer bei der Auswahl „richtiger“ Tags zu unterstützen und sie in der Tagvergabe zu trainieren, bietet es sich an, Tags durch das System vorzuschlagen (MacLauri, 2005; Xu, Fu, Mao, & Su, 2006). Dieses Vorschlagssystem kann auf zwei Ebenen arbeiten. Wir unterscheiden zwischen syntaktischen Vorschlägen (z.B. den Tag *Bild* durch den Tag *Bilder* ergänzen) und relationalen Hinweisen (z.B. der Nutzer annotiert den Tag *Bild*, und das System schlägt *Grafik* vor, weil unser Nutzer bereits den Tag *Grafik* verwendet hat). Der Einsatz von weitergehender Tag-Recommendation bei der Indexierung ist nicht frei von Problemen – im Gegensatz zu Vorschlägen von Such-Tags im Retrieval (s. u.!). Schlägt ein System nämlich dem Indexer die jeweils bereits am häufigsten vergebenen Tags eines Dokuments vor und orientieren sich die indexierenden Nutzer tatsächlich daran, so entsteht – in einer Art self-fulfilling prophecy – stets eine Tag-Verteilung nach dem Power Law. Zum anderen kann man die Tags als Elemente der natürlichen Sprache ansehen und sie mit den etablierten Methoden des Natural Language Processing (NLP) bearbeiten (Stock, 2007a, Kap. 14 bis 18). Studien zeigen, dass circa 90 Prozent der Tags Nomen sind (Guy & Tonkin, 2006).

Ob auch andere Wortarten, vor allem Verben, einmal häufig genutzt werden, bleibt abzuwarten. Da Adjektive an Nomen gebunden sind, müssen wir diese stets mitbeachten. Aus Gründen der Einfachheit nehmen wir an, dass der größte Teil der Tags Nomen (einschließlich Adjektiven) sind. Deswegen können wir während der Tag-Bearbeitung mittels NLP-Verfahren alle anderen Wortformen vernachlässigen.

Für die NLP-Verarbeitung der Tags bevorzugen wir einen wort-basierten Ansatz, da dieser bessere Möglichkeiten zur Bearbeitung bietet (siehe Abbildung 13 die drei untersten Ebenen). Eine (hier nicht weiter verfolgte) alternative Möglichkeit bietet die Implementierung von *n*-Gramm-Verfahren (Stock, 2007a, Kap. 13).

Problematisch ist, dass wir nicht davon ausgehen können, dass alle Dokumente getaggt sind. Für textuelle Dokumente können wir jedoch einen Ersatz generieren: Bei Blogs haben wir den gesamten Text als Basis, bei Fotos und Videos zumindest die Titel, Beschreibungen und Kommentare. Die wichtigsten Terme bestimmen wir automatisch (Brooks & Montanez, 2006a; 2006b). Mit dem Ergebnis aus $WDF \cdot IDF$ lässt sich ein Ranking erstellen; die ersten *n* (etwa $n = 3$) Terme werden als Tags bestimmt. Dieses Verfahren darf nicht zu häufig Anwendung finden, da die automatische Indexierung dem Grundgedanken des intellektuellen Taggings zuwiderläuft. Eine Studie von Al-Khalifa und Davis (2007) zeigt im Trend, dass die von den Prosumern indexierten Tags nur wenig mit Termen korrelieren, die auf der Basis automatischer Verfahren gewonnen werden.

Nach der Spracherkennung und dem Parsing müssen wir die kontextspezifischen synkategorematischen Tags, wie *me* und *ich*, bearbeiten. Für diese Problemfälle schlagen wir vor, sie automatisch durch den Nutzernamen des angemeldeten Users zu ersetzen. Sucht dieser spezielle Nutzer nach sich selbst, indem er den Tag *me* nutzt, erhält er die passenden Resultate. Alle nicht angemeldeten User erhalten für eine Suchanfrage mit *me* keine Ergebnisse. Nur die Suche nach dem Benutzernamen zeigt ein Resultat. Der Bearbeitungsalgorithmus folgt den typischen NLP-Aufgaben (Fehlererkennung, Wortformzusammenfassung, Identifikation von Eigennamen, Phrasenerkennung und Dekomposition), wie in Abbildung 13 dargestellt. „Höhere“ Bearbeitungsschritte verlangen nach einem Mash-up aus Folksonomy und kontrolliertem Vokabular. Während der Erkennung von Homonymen und Synonymen muss man Wissensordnungen wie beispielsweise WordNet (Miller, 1998) zur Hilfe nehmen. Außerdem könnte es hilfreich sein, bei der Homonymunterscheidung auch Co-Occurrence-Statistiken der Tags einzubeziehen (Butterfield et al., 2006).

Man stelle sich ein Tag Cluster vor, das die Tags *Java*, *Perl* und *Programmierung* enthält, und ein weiteres Cluster, welches *Java*, *Sumatra* und *Indonesien* als Tags aufweist. Das System könnte beide Tag-Cluster anzeigen und den User, der nach *Java* gesucht hat, selbst auswählen und entscheiden lassen, welche Bedeutung des Tags er tatsächlich gesucht hat. Ein Merging von Thesauri, Klassifikationssystemen oder Ontologien mit Folksonomies kann die Nutzung von Relationen (wie Hierarchie oder Assoziationsrelation) ermöglichen (Gruber, 2007) und so die Suchanfrage erweitern bzw. eingrenzen. Im Sinne des multilingualen Retrievals erleichtert der Einsatz von mehrsprachigen maschinenlesbaren Wörterbüchern den Zugriff auf fremdsprachige Dokumente.

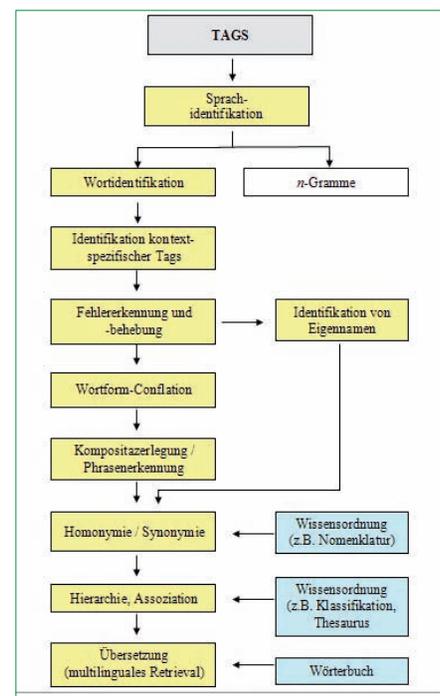


Abbildung 13: Der Aufgaben-Algorithmus der Tag-NLP.

Recherchen nach Power Tags

Wir wollen kurz einen Blick zurück auf die Abbildungen 5 und 8 werfen. Hier gibt es zwei Möglichkeiten, den Einfluss der Power Tags im Retrieval zu verstärken: a) Im Fall einer Power Law-Verteilung werden die ersten *n* Tags genutzt. *n* wird hier – in Abhängigkeit von dem Exponenten *a* – eher klein gehalten; bei $a = 1$ bietet sich beispielsweise $n = 4$, bei $a = 2$ etwa $n = 2$ an. b) Im Fall der invers-logistischen Verteilung wählen wir alle Tags des linken „Long Trunk“ bis zum Wendepunkt der Kurve aus. Es wäre wünschenswert, eine zusätzliche, optionale Retrievalfunktionalität zu installieren, die nur Ergebnisse aus Übereinstimmungen von Suchargument

und Power Tag erzielt („Power Tags only“). Demnach würde das Beispiel in Abbildung 4 nur mit den Tags *politik* und *bundestag* gefunden werden. Alle anderen Terme und Tags blieben unberücksichtigt. Wir erhoffen uns durch das Angebot von Power Tags als Suchoption eine Verbesserung der Präzision der Suchergebnisse. Begründen können wir diese Vermutung durch das gegenläufige Verhältnis von Recall und Precision. Steigt das eine, fällt das andere. Im Falle der Recherche ausschließlich nach Power Tags wird der Recall – da der gesamte dokumentspezifische „Lange Schwanz“ abgeschnitten wird – drastisch verkleinert.

Relevance Ranking der getaggtten Dokumente

Ein bislang noch weitgehend ungelöstes Problem ist das Ranking von getaggtten Dokumenten, die man als Antwort auf seine Suchanfrage erhält. Die Menge der Dokumente, die man nach der NLP-Bearbeitung bekommt, ist nämlich noch unsortiert.

Es gibt eine Patentanmeldung von Yahoo! für Flickr, die sich mit diesem Problem auseinandersetzt (Butterfield et al., 2006). Für ein Ranking nach „Interestingness“ in Narrow Folksonomies wurden fünf generelle Kriterien definiert: 1) die Zahl der Tags zu einem Dokument, 2) die Zahl der Nutzer, die ein Dokument taggen, 3) die Zahl der Nutzer, die das Dokument nach einer Suche erhalten, 4) die Zeit (je älter desto weniger relevant) und 5) die Relevanz der Tags. Zusätzlich gibt es zwei weitere Rankingkriterien für den „personalized interestingness rank“, 6) Nutzerpräferenzen (z.B. Favoritenlisten) und 7) der Wohnort des Nutzers. Beim Faktor der Zeit melden wir Bedenken an. Es ist nicht notwendig, dass ein Dokument im Laufe der Zeit stets an Relevanz verliert. Dann wäre die Relevanz von, sagen wir, der „Mona Lisa“ heutzutage nahe Null, was Kunsthistoriker sehr wahrscheinlich bestreiten würden. Butterfield et al. berücksichtigen u. E. zu Recht die kollaborativen Aspekte, die wir bei Web 2.0-Diensten vorfinden; sie vernachlässigen jedoch insbesondere altbekannte Retrievalmodelle wie beispielsweise das Vektorraummodell oder die Linktopologie. Auch sollte man konkrete Aktionen der Prosumer nicht außer Acht lassen.

Für das Relevance Ranking der getaggtten Dokumente gibt es nämlich drei Mengen von Faktoren:

- die (informationslinguistisch „gereinigten“) Tags selbst,
- die Kollaboration in Web 2.0-Diensten,
- nutzer-spezifische Rankingkriterien (siehe Abbildung 14).

Jede Faktorenmenge und auch jeder einzelne Faktor kann zudem gewichtet werden. Jeder einzelne der Faktoren führt entweder zu einem Ansteigen (bei positiven Faktoren) oder zu einem Abfallen (bei negativen Faktoren) des Retrievalstatuswertes der Dokumente.

Set eins berücksichtigt die konkreten Tags und bearbeitet sie für das Relevance Ranking. Das Vektorraummodell kann genutzt werden, um die Ähnlichkeit zwischen Dokumenten zu bestimmen (Stock, 2007a, Kap. 20). Die Dimensionen sind dabei die verschiedenen Tags in der Datenbank, der Wert der Dimension wird durch $WDF * IDF$ bestimmt (1a). Der WDF-Wert errechnet sich aus der dokumentspezifischen Häufigkeit ($freq$) eines bestimmten Tags t in Relation zur Gesamthäufigkeit L aller Tags, die dem jeweiligen Dokument d (auch u. U. mehrfach) zugeordnet worden sind:

$$WDF(t,d) = [ld (freq(d,t) + 1) / ld L.$$

Der IDF-Wert bezieht die Gewichtung eines Tags t auf sein Vorkommen in der gesamten Datenbank. Sei N die Mächtigkeit der Menge aller Datensätze in der Datenbank und n die Anzahl derjenigen Dokumente, die mittels t indexiert worden sind, so gilt:

$$IDF(t) = ld (N/n).$$

Die Dokumente werden durch Vektoren abgebildet, und die Ähnlichkeit zwischen Dokument und Suchfrage wird letztlich durch den Cosinus (1b) bestimmt. Der Wert des WDF hängt in Broad Folksonomies von der Anzahl der Index-Tags ab; bei Narrow Folksonomies wird der WDF-Wert dagegen durch die Zahl der Such-Tags determiniert.

Es gibt Ansätze, die für das Ranking einen modifizierten PageRank einsetzen. Das Motto heißt hier: „The basic notion is that a resource which is tagged with important tags by important users becomes important itself“ (Hotho, A., Jäschke, R., Schmitz, C., & Stumme, G., 2006a, 417). Jedoch ist diese Schlussfolgerung problematisch und kann zu unbefriedigenden Ergebnissen führen. Die gleiche Studie stellt den (bei BibSonomy eingesetzten) FolkRank-Algorithmus vor, mit dem Ziel „to focus the ranking around the topics defined in the preference vector“ (419). Der FolkRank verfolgt die Idee eines „Super-Posters“ oder „Super-Autors“, der eine große Menge an Con-

tent veröffentlicht und damit ein Experte (in einem bestimmten Bereich) zu sein scheint (1c). Daher sollten Inhalte dieser User beim Ranking von Suchergebnissen höher gewertet werden (Hotho, A., Jäschke, R., Schmitz, C., & Stumme, G., 2006a, 2006b). Abhängig von ihrer Häufigkeitsverteilung werden manche Tags als Power Tags markiert (1d). Der WDF-IDF-Gewichtungswert von Power Tags kann mittels eines festzulegenden Faktors f ($f > 1$) justiert werden.

Set zwei bezieht die aktive Zusammenarbeit der User als positive Faktoren mit in das Ranking ein. So können die Klick-Raten (2a) bezogen auf einzelne Dokumente zum Ranking herangezogen werden (Culliss, 1997). Für Jung, Herlocker und Webster (2007) sind die Klick-Raten ein implizites Relevance Feedback in der Internetsuche; sie sind im hohen Maße ein kollaborationsorientiertes Ranking-Kriterium im Sinne des Web 2.0. Möglicherweise ist auch die Anzahl der unterschiedlichen Nutzer, die ein Dokument indexieren, ein sinnvoller Ranking-Faktor (2b). Hoch frequentierte Diskussionen auf der Basis gewisser Dokumente verweisen auf Dokumente von hoher Bedeutung für die Community. Die Zahl der Kommentare zu einem Dokument (2c) kann die Stellung im Ranking bestimmen. Insbesondere Blogpostings, aber auch andere Dokumente wie Bilder und Videos, insofern auf diese Links verweisen, bieten Potentiale für einen quantitativen Ausdruck ihrer linktopologischen Stellung entweder mittels des Kleinberg-Algorithmus' (Kleinberg, 1999) oder des PageRank (Brin & Page, 1998). Autorities (Dokumente mit vielen In-Links) befinden sich im Zentrum des Interesses von bestimmten Themen und sollten daher höher gerankt werden.

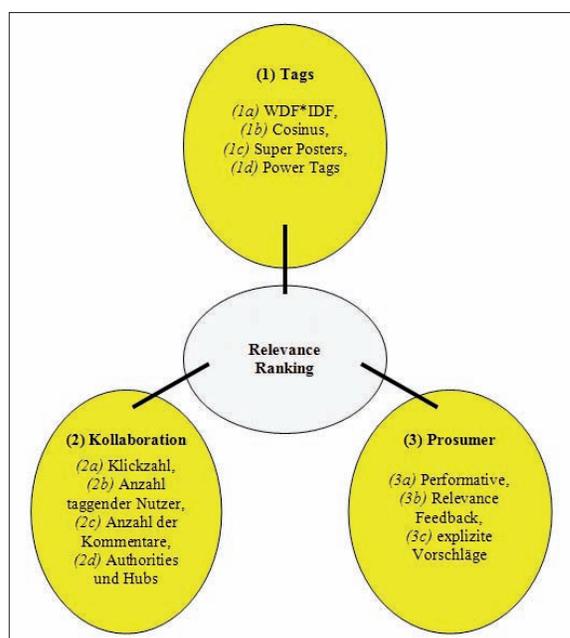


Abbildung 14: Kriterien für das Relevance Ranking von getaggtten Dokumenten.

Das dritte Set der Rankingkriterien bezieht sich auf den Nutzer selbst. Dokumente, die mit *todo* oder *toread* getaggt sind, scheinen sehr wichtig für den Nutzer zu sein und sollten daher ihre Stellung im Ranking erhöhen (3a). Negative Werturteile (wie z.B. *stupid*) führen – abhängig von der Häufigkeit ihres dokumentenspezifischen Vorkommens – zu einer Senkung des Retrievalstatuswertes. Man kann auch den Nutzer aktiv am Ranking beteiligen und ihm die Gelegenheit zum Feedback geben. Das kann auf zweifache Weise geschehen: (3b) das Relevance Feedback geschieht nach der Anzeige und Bewertung der Suchergebnisse (z.B. Rocchio, 1971) oder (3c) mittels eines formalen Recommendersystems, wie es z.B. in einer Sternchen-Bewertung oder mittels eines Frage-Dialogs nach dem Stil „Hat Ihnen dieses Dokument geholfen?“ gegeben ist. Der User bekommt die Gelegenheit, das Dokument zu bewerten und so aktiv an der Bestimmung des jeweiligen Retrievalstatuswertes mitzuwirken – ein demokratischer Weg des Relevance Ranking. Beide Faktoren (3b und 3c) kön-

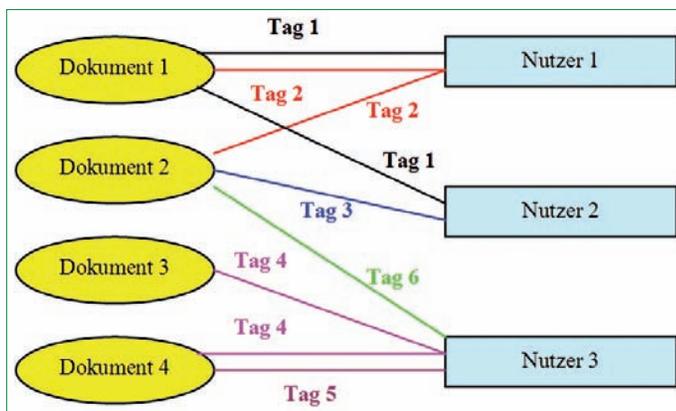


Abbildung 15: Dokumente, Tags und Nutzer in einer Folksonomy.

nen in das Ranking einfließen; einmal implizit als Gewichtungsfaktor für den Retrievalstatuswert und einmal explizit („Andere Nutzer fanden dieses Dokument sehr hilfreich“ oder „Das Dokument erhält 5 von 6 Relevanzpunkten“). Beide Möglichkeiten haben u. W. bis jetzt noch keinen Eingang in die Web 2.0-Dienste gefunden.

Aus Nutzersicht wichtig ist bei solch ausgefeilten Rankingalgorithmen, dass er sie ganz oder teilweise abschalten kann. Ein Drop-Down-Menü scheint eine optimale Vorgehensweise, um die Nutzer in die Einstellung der Rankingfaktoren der Suchresultate einzubeziehen. Er selbst kann entscheiden, nach welchen Kriterien er die Ergebnisse sortieren möchte. Vielleicht möchte ein User lieber nach formalen Kriterien (Veröffentlichungszeitpunkt, Autor usw.) sortieren oder er bevorzugt aus unserem Werkzeugkasten für Relevance Ranking nur ganz bestimmte Kriterien.

Vorschlagssysteme auf der Basis von Folksonomies: Recommendation von Dokumenten und Nutzern

Innerhalb einer Folksonomy sind wir mit drei unterschiedlichen Aspekten konfrontiert (Marlow, Naaman, Boyd, & Davis, 2006; Yeung, Gibbins, & Shadbolt, 2008):

- die zu beschreibenden Dokumente,
- die Tags (Worte), die zur Beschreibung gewählt werden,
- die Nutzer (Prosumer), die die Indexierung ausführen.

Nutzer untereinander sowie Dokumente untereinander sind in einem sozialen Netzwerk miteinander verknüpft, wobei als Pfade jeweils die Tags fungieren. Dokumente sind erstens *thematisch* miteinander verbunden, wenn sie durch dieselben Tags indexiert worden sind. Die Dokumente 1 und 2 sowie 3 und 4 aus Abbildung 15 sind jeweils thematisch verknüpft (Dokumente 1 und 2 über Tag 2; Dokumente 3 und 4 über Tag 4).

Zusätzlich sind Dokumente zweitens *über gemeinsame Nutzer* verknüpft. So sind die Dokumente 1 und 2, 3 und 4, aber auch 2 und 4 über ihre Nutzer verknüpft. Nutzer sind miteinander verknüpft, wenn sie entweder dieselben Tags verwenden oder dieselben Dokumente indexieren. Nutzer sind *thematisch* verbunden, wenn sie dieselben Tags verwenden (im Beispiel die Nutzer 1 und 2 über Tags 1 und 2); sie sind *über gemeinsame Dokumente* gekoppelt, wenn sie diese jeweils inhaltlich beschreiben (Nutzer 1, 2 und 3 über Dokument 2).

Wir wollen diese Zusammenhänge ausnutzen, um auf der Basis der Stellung eines Tags, eines Nutzers und eines Dokuments im Netzwerk automatisch Vorschläge zu generieren (Diederich & Iofciu, 2006; Höhfeld & Kwiatkowski, 2007, 272 f.). Unsere Ausführungen konzentrieren sich auf Empfehlungen ähnlicher Dokumente (ausgehend von einem Musterdokument: „More like this!“) sowie von ähnlichen Nutzern (ausgehend von einem bestimmten Nutzer, beispielweise sich selbst: „More like me!“), da beide Vorschläge über recht einfache Algorithmen auf der Basis von Co-Occurences arbeiten.

Wir wollen diese Zusammenhänge ausnutzen, um auf der Basis der Stellung eines Tags, eines Nutzers und eines Dokuments im Netzwerk automatisch Vorschläge zu generieren (Diederich & Iofciu, 2006; Höhfeld & Kwiatkowski, 2007, 272 f.). Unsere Ausführungen konzentrieren sich auf Empfehlungen ähnlicher Dokumente (ausgehend von einem Musterdokument: „More like this!“) sowie von ähnlichen Nutzern (ausgehend von einem bestimmten Nutzer, beispielweise sich selbst: „More like me!“), da beide Vorschläge über recht einfache Algorithmen auf der Basis von Co-Occurences arbeiten.

Möchte man durch die Empfehlungen weiterer Tags dem Nutzer Synonyme und Quasi-Synonyme an die Hand geben, so eignen sich Co-Occurences nicht, denn Synonyme kommen selten gemeinsam syntagmatisch verbunden als Tags vor (ein User, der sein Bild mit *Hochzeit* indexiert hat, wird dieses kaum zusätzlich mit *Eheschließung* und *Heirat* beschreiben). Hier sind andere Algorithmen gefordert. Der Fall wird genau dann leicht, wenn wir auf eine Begriffsordnung (wie beispielsweise WordNet) zurückgreifen können, da dann die Empfehlungen aus den dort eingetragenen Synonymen gebildet werden.

Der erste Schritt bei Empfehlungen von Dokumenten und Nutzern besteht darin, Ähnlichkeiten zwischen dem Ausgangsdokument bzw. -nutzer und allen anderen Dokumenten (Nutzern) der Datenbank zu berechnen. Dazu kommen einschlägige Algorithmen wie Jaccard-Sneath, Dice oder Cosinus-infrage (Stock & Stock, 2008, Kap. 20). Möchte man eine Community (also ähnliche Nutzer), ausgehend von Nutzer 1, identifizieren, so errechnet man nach Jaccard-Sneath

$$\text{Ähnlichkeit (Nutzer 1, Nutzer } i) = \frac{g}{a + b - g}$$

Die Formel findet zweifach Anwendung: 1) a ist die Anzahl der Tags, die Nutzer 1 vergeben hat, b die Anzahl der Tags, die Nutzer i zum Indexieren heranzog und g ist die Anzahl derjenigen Tags, die die beiden Nutzer gemeinsam verwenden; 2) a ist nunmehr die Anzahl der Dokumente, die Nutzer 1 ausgewertet hat, b die Anzahl der von Nutzer i indexierten Dokumente und g die Zahl der von beiden beschriebenen Dokumente.

Der zweite Schritt besteht entweder in einer einfachen oder in einer elaborierteren Vorgehensweise. Der einfache Weg fasst – in der Reihenfolge der beiden errechneten Ähnlichkeiten – die jeweils ersten k Dokumente bzw. Nutzer (k kann auf etwa 20 eingestellt werden) zu einer Quasiklasse zusammen (Methode der k-nearest neighbors). Ein aufwändigerer zweiter Weg arbeitet mit der Clusteranalyse (z.B. Single Linkage oder Complete Linkage) oder vergleichbaren Verfahren. Auch hier erhalten wir Quasiklassen – nun jedoch möglicherweise mehrere und diese ggf. in einer hierarchischen Ordnung. Dem Nutzer werden die Quasiklassen als ähnliche Dokumente bzw. als ähnliche Nutzer (oder als „seine“ Community) vorgeschlagen.

Es gibt bisher nur sehr wenige empirische Studien zum Einsatz von Recommendation bei Folksonomies. Erwähnenswert ist die Untersuchung von Jäschke, Marinho, Hotho, Schmidt-Thieme und Stumme (2007), die die Wirkung von vorgeschlagenen Suchtags auf die Retrievalleistung (anhand von BibSonomy und dem Mu-

sikdienst Last.fm) untersuchten. Der Recall-Precision-Graph zeigt durchgehend Steigerungen der Retrievalqualität beim Einsatz von Recommendation-Methoden.

Fazit und offene Forschungsprobleme

Folksonomies haben in informationswissenschaftlicher Sicht sowohl Vorteile (authentische Sprache der Nutzer, Vereinigung unterschiedlicher Sichtweisen auf ein Dokument, Quelle für die Entwicklung kontrollierter Vokabularien, Ermöglichung von Suchen und Browsen, Identifikation von Communities usw.) als auch Nachteile (Mangel an Präzision, unterschiedliche Indexierungsebenen, Vermischung von Sprachen, versteckte paradigmatische Relationen, Spam-Tags, nutzerspezifische Tags und andere unbrauchbare Terme, Verschmelzung der semantischen Ebenen bei Bildern und Videos). Da die Vorteile überwiegen, müssen wir dafür Sorge tragen, dass die Nachteile möglichst minimiert werden. Ein erster Schritt der Problemlösung ist die Behandlung der Tags als natürlichsprachige Eingaben, auf die (bekannte wie neue) Methoden der Informationslinguistik (Natural Language Processing) angewandt werden. Damit sollten sich zumindest Tags unterschiedlicher Sprachen trennen, die Tags auf Grund- oder Stammformen bringen sowie Eingabefehler korrigieren lassen. Über die Kopplung der Folksonomies mit kontrollierten

Vokabularien sollten – in einem gewissen Maße – Homonyme getrennt und Synonyme zusammengeführt werden; ebenso können die semantischen Relationen der eingesetzten Wissensordnungen (z.B. die Hierarchien) bei der Recherche nutzbar werden. Liegen maschinenlesbare Wörterbücher vor, wird auch multilinguales Retrieval möglich.

Eine große Herausforderung ist die Kreation passender Algorithmen des Relevance Ranking. Hier sehen wir drei Kriterienbündel, die Berücksichtigung finden müssen: die (via NLP „gereinigten“) Tags selbst (WDF*IDF, Vektorraummodell, Berücksichtigung von Super-Postern sowie von Power Tags), Merkmale der Kollaboration (Klickraten, Anzahl der indexierenden Nutzer, Anzahl der Kommentare und linktopologische Ansätze bei verlinkten Dokumenten) sowie Aspekte der Prosumer (Performative, Relevance Feedback, formale Bewertungen). Ein neues Suchfeld nach *Nur Power Tags* verringert den Ballast und steigert somit die Präzision der Suchresultate.

Über Ähnlichkeitsberechnungen und den Einsatz von Clusteranalyse oder verwandter Verfahren erhalten wir Algorithmen, um sowohl ähnliche Dokumente („More like this!“) zu finden als auch um ähnliche Nutzer („More like me!“) und damit Communities zu identifizieren.

Das Forschungsgebiet der Folksonomies ist noch sehr jung. Teilweise haben wir außer Ansätzen noch nicht viel Konkretes an brauchbaren (d.h. in Informa-

tionsdiensten praktisch anwendbaren) Resultaten vorzuweisen. Wir sehen folgende offenen Forschungsfragen:

- Bisher nicht untersucht sind die Such-Tags. Insbesondere bei Narrow Folksonomies brauchen wir diese jedoch, um Tagverteilungen herzustellen und – soweit dort auch vorhanden – Power Tags zu identifizieren.
- Wir wissen, dass viele Tagverteilungen bei Broad Folksonomies dem Power Law folgen; wir wissen aber auch, dass dies nicht auf alle dokumentenspezifischen Tagverteilungen zutrifft. Wie groß sind die Anteile der Tagverteilungen auf der Mikroebene, die dem Power Law und die der invers-logistischen Kurve folgen? Gibt es darüber hinaus weitere Verteilungsformen?
- Das Entstehen einer Power-Law-Verteilung ist durch den Yule-Simon-Ansatz theoretisch zu fassen. Wie erklären wir aber die Entwicklung einer invers-logistischen Verteilung?
- In Web 2.0-Diensten werden häufig Bilder und Videos indexiert. Wir haben als Problem die Verschmelzung der semantischen Ebenen bei der Indexierung durch Tags (Ofness, Aboutness, Ikonologie, Isness) markiert. Des Weiteren kennen wir die Inter-Indexer-Inkonsistenz bei der Indexierung solcher Dokumente. Eine Lösung können wir nicht bieten.



CONTENTdm—Digital Collection Management Software

Werden Sie Teil einer aktiven Anwendergemeinschaft mit einem gemeinsamen Ziel: wichtige Bestände, Sammlungen und Kulturschätze zu digitalisieren und einem breiten Publikum zugänglich zu machen.

CONTENTdm bietet die vollständige Lösung zur Speicherung, Verwaltung, Selektion und Internet-Präsentation Ihrer digitalen Sammlungen.

Mit CONTENTdm generieren Sie aus Ihren Spezialsammlungen vollständig suchbare Online-Ressourcen, auf die von jedem Browser aus zugegriffen werden kann.

CONTENTdm—digital collection management software



Grünwalder Weg 28g
82041 Oberhaching, Deutschland
T +49-(0)89-613 08 300
F +49-(0)89-613 08 399
E deutschland@oclc.org

www.oclc.org

- Unsere Lösungsoptionen für die Mängel bei Tags beruhen auf informativ-linguistischen Ansätzen. Können wir uns in der Tat auf die Behandlung von Substantiven und Adjektiven beschränken und Verben außer acht lassen? Wie häufig kommen Verben überhaupt als Tags vor?
- In den dokumentspezifischen Indextaxen sind paradigmatische Relationen „versteckt“. Wie lassen sich die versteckten semantischen Relationen aufdecken?
- Bislang setzt kein System durchgehend NLP zur Bearbeitung von Tags ein. Entsprechend fehlen Erfahrungen mit informativ-linguistisch bearbeiteten Tags.
- Wir behaupteten, dass die Retrievaloption „Power Tags only“ Verbesserungen der Retrievalqualität in Form einer Steigerung der Precision mit sich bringt. Hier fehlt noch jede empirische Bestätigung.
- Der vorgestellte Werkzeugkasten für Kriterien des Relevance Ranking für getaggte Dokumente ist ein erster Vorschlag, Algorithmen klassischer Retrievalmodelle mit den Gegebenheiten der Web 2.0-Dienste (Kollaboration, Prosumer) zu verbinden. Hier fehlt nicht nur die technische Umsetzung, sondern auch – im Anschluss daran – deren Evaluation.
- Nach welchen Algorithmen kann man Nutzern ähnliche Tags (vor allem Synonyme und Quasi-Synonyme ihrer ursprünglichen Suchterme) vorschlagen, wenn keine Begriffsordnung einsetzbar ist?
- Eignen sich Verfahren nach k-nearest neighbors oder nach der Clusteralanalyse für die Identifikation von ähnlichen Dokumenten und von Communities? Werden sie von den Nutzern angenommen?

Literatur

Al-Khalifa, H.S. (2007). Automatic Document-Level Semantic Metadata Annotation Using Folksonomies and Domain Ontologies. PhD-Thesis. University of Southampton / School of Electronics and Computer Science.

Al-Khalifa, H.S., & Davis, H.C. (2007). Exploring the value of folksonomies for creating semantic metadata. *International Journal on Semantic Web and Information Systems* 3(2007)1, 12-38.

Aurnhammer, M., Hanappe, P., & Steels, L. (2006). Augmenting navigation for collaborative tagging with emergent semantics. *Lecture Notes in Computer Science* 4273, 58-71.

Bar-Ilan, J., Shoham, S., Idan, A., Miller, Y., & Shachak, A. (2006). Structured vs. unstructured tagging – A case study. In: *Proceedings of the 15th International WWW Conference. Collaborative Web Tagging Workshop*.

Beaudoin, J. (2007). Flickr image tagging: Patterns made visible. *Bulletin of the American Society for Information Science and Technology* 34(2007)1, 26-29.

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30, 107-117.

Brooks, C.H., & Montanez, N. (2006a). An analysis of the effectiveness of tagging in blogs. In: *Nicolov, N., Salvetti, F., Liberman, M., & Martin, J.H. (eds.): Computation Approaches to Analyzing Weblogs. Papers from the 2006 AAAI Spring Symposium (S. 9-15)*. Menlo Park, Cal: AAAI Press. (Technical Report SS-06-03. American Association for Artificial Intelligence).

Brooks, C.H., & Montanez, N. (2006b). Improved annotation of the blogosphere via autotagging and hierarchical clustering. In: *Proceedings of the 15th International World Wide Web Conference (S. 625-632)*. New York: ACM.

Butterfield, D.S., Costello, E., Fake, C., Henderson-Begg, C.J., & Mourachow, S. (2006). Interestingness ranking of media objects. Patentanmeldung Nr. US 2006/0242139 A1. Eingereicht: 8.2.2006. Publiziert: 26.10.2006.

Butterfield, D.S., Costello, E., Fake, C., Henderson-Begg, C.J., Mourachow, S., & Schachter, J.E. (2006). Media object metadata association and ranking. Patentanmeldung Nr. US 2006/0242178 A1. Eingereicht: 8.2.2006. Publiziert: 26.10.2006.

Campbell, D.G. (2006). A phenomenological framework for the relationship between the semantic Web and user-centered tagging systems. In: *17th ASIS&T SIG/CR Classification Research Workshop*.

Capocci, A., & Caldarelli (2007). Folksonomies and clustering in the collaborative system CiteULike. (arXiv:0710.2835v2).

Carlin, S.A. (2007). Social Tagging. Schlagwortvergabe durch User als Hilfsmittel zur Suche im Web. Ansatz, Modelle, Realisierungen. – Boizenburg: vwh.

Cattuto, C. (2006). Semiotic dynamics in online social communities. *European Physical Journal C* 46(2006)Suppl. 2, 33-37.

Cattuto, C., Loreto, V., & Pietronero, L. (2007). Semiotic dynamics and collaborative tagging. *Proceedings of the National Academy of Sciences of the United States of America* 104(2007)5, 1461-1464.

Christiaens, S. (2006). Metadata mechanisms: From ontology to folksonomy ... and back. *Lecture Notes in Computer Science* 4277, 199-207.

Culliss, G.A. (1997). Method for organizing information. Patent-No. US 6.006.222. Granted: Dec. 21, 1999. Applied: Jan. 8, 1997.

Diederich, J., & Iofciu, T. (2006). Finding communities of practice from user profiles based on folksonomies. In *Proceedings of the 1st International Workshop on Building Technology Enhanced Learning Solutions for Communities of Practice (TEL-CoPs'06)*.

Dye, J. (2006). Folksonomy: A game of high-tech (and high-stakes) tag. *E-Content* 29(2006)3, 38-43.

Egghe, L. (2005). *Power Laws in the Information Production Process: Lotkaian Informetrics*. Amsterdam: Elsevier Academic Press.

Egghe, L., & Rousseau, R. (1990). *Introduction to Informetrics*. Amsterdam: Elsevier.

Fichter, D. (2006). Intranet applications for tagging and folksonomies. *Online* 30(2006)3, 43-45.

Fidel, R. (2006). An ecological approach to the design of information systems. *Bulletin of the American Society for Information Science and Technology* 33(2006)1, 6-8.

Furnas, G.W., Fake, C., von Ahn, L., Schachter, J., Golder, S., Fox, K., Davis, M., Marlow, C., & Naaman, M. (2006). Why do tagging systems work? In: *CHI '06 Extended Abstracts on Human Factors in Computing Systems (S. 36-39)*. New York: ACM.

Gendarmi, D., & Lanubile, F. (2006). Community-driven ontology evolution based on folksonomies.

Lecture Notes in Computer Science, 4277, 181-188.

Golder, S.A., & Huberman, B.A. (2006). Usage patterns of collaborative tagging systems. *Journal of Information Science* 32(2006)2, 198-208.

Gordon-Murnane, L. (2006). Social bookmarking, folksonomies, and Web 2.0 tools. *Searcher - The Magazine for Database Professionals* 14(2006)6, 26-38.

Grahl, M., Hotho, A., & Stumme, G. (2007). Conceptual clustering of social bookmarking sites. In: *Proceedings of I-KNOW '07, Graz, Austria, September 5-7, 2007 (S. 356-264)*.

Gruber, T. (2006). Where the social Web meets the semantic Web. *Lecture Notes in Computer Science* 4273, 994.

Gruber, T. (2007). Ontology of folksonomy: A mash-up of apples and oranges. *International Journal on Web Semantics and Information Systems* 3(2007)1, 1-11.

Guy, M., & Tonkin, E. (2006). Folksonomies: Tidying up tags? *D-Lib Magazine* 12(2006)1.

Hänger, C., & Krätzsch, C. (2007). Collaborative Tagging als Service von Hochschulbibliotheken. In: *Ball, R. (Hrsg.): Wissenschaftskommunikation der Zukunft. 4. Konferenz der Zentralbibliothek Forschungszentrum Jülich (S. 123-134)*. Jülich: Forschungszentrum Jülich GmbH / Zentralbibliothek.

Hammond, T., Hannay, T., Lund, B., & Scott, J. (2005). Social bookmarking tools. A general review. Part 1. *D-Lib Magazine*, 12(1).

Heller, L. (2007). Bibliographie und Sacherschließung in der Hand vernetzter Informationsbenutzer. *Bibliothek*, 31(2), 162-172.

Herget, J., & Hierl, S. (2007). Top down versus bottom up: Knowledge representation in transition. From controlled vocabulary to folksonomies. In: *Proceedings of the 2nd International CODATA Symposium on Generalization of Information (S. 150-157)*.

Heymann, P., & Garcia-Molina, H. (2006). Collaborative creation of communal hierarchical taxonomies in social tagging system. *InfoLab Technical Report; 2006-10*. Stanford: Stanford University / Computer Science Dept.

Höfheld, S., & Kwiatkowski, M. (2007). Empfehlungssysteme aus informationswissenschaftlicher Sicht – State of the Art. *Information – Wissenschaft und Praxis* 58(2007)5, 265-276.

Hotho, A., Jäschke, R., Schmitz, C., & Stumme, G. (2006a). Information retrieval in folksonomies: Search and ranking. *Lecture Notes in Computer Science* 4011, 411-426.

Hotho, A., Jäschke, R., Schmitz, C., & Stumme, G. (2006b). Trend detection in folksonomies. *Lecture Notes in Computer Science*, 4306, 56-70.

Hotho, A., Jäschke, R., Schmitz, C., & Stumme, G. (2006c). BibSonomy: A social bookmark and publication system. In: *Proceedings of the Conceptual Structure Tool Interoperability Workshop at the 14th International Conference on Conceptual Structures*.

Huang, H. (2006). Tag distribution analysis using the power law to evaluate social tagging systems: A case study in the Flickr database. In *17th ASIS&T SIG/CR Classification Research Workshop. Abstracts of Posters (S. 14-15)*.

Ingwersen, P. (2002). Cognitive perspectives of document representation. In: *CoLIS 4: 4th International Conference on Conceptions of Library and Information Science (S. 285-300)*. Greenwood Village: Libraries Unlimited.

Jäschke, R., Marinho, L., Hotho, A., Schmidt-Thieme, L., & Stumme, G. (2007). Tag recommendations in folksonomies. *Lecture Notes in Artificial Intelligence* 4702, 506-514.

Jung, S., Herlocker, J.L., & Webster, J. (2007). Click data as implicit relevance feedback in web

- search. *Information Processing & Management*, 43, 791-807.
- Kim, H.G., Hwang, S.H., Kang, Y.K., Kim, H.L., & Yang, H.S. (2007). An agent environment for contextualizing folksonomies in a triadic context. *Lecture Notes in Computer Science* 4496, 728-737.
- Kim, H.L., Hwang, S.H., & Kim, H.G. (2007). FCA-based approach for mining contextualized folksonomy. In: *Proceedings of the 2007 ACM Symposium on Applied Computing* (S. 1340-1345). New York: ACM.
- Kipp, M.E.I. (2006a). @toread and cool: Tagging for time, task and emotion. In: *17th ASIS&T SIG/CR Classification Research Workshop. Abstracts of Posters* (S. 16-17).
- Kipp, M.E.I. (2006b). Complementary or discrete contexts in online indexing: A comparison of user, creator and intermediary keywords. *Canadian Journal of Information and Library Science* 30(2006)3, in prep.
- Kipp, M.E.I. (2007). Tagging for health information organisation and retrieval. In: *Proceedings of the North American Symposium on Knowledge Organization*, vol. 1 (S. 63-74).
- Kipp, M.E.I., & Campbell, D.G. (2006). Patterns and inconsistencies in collaborative tagging systems: An examination of tagging practices. In: *Proceedings of the 69th Annual Meeting of the American Society for Information Science & Technology* (Vol. 43), Austin, TX (CD-ROM).
- Kleinberg, J. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM* 46(1999)5, 604-632.
- Kolbitsch, J. (2007a). *Aspects of Digital Libraries*. Diss. TU Graz.
- Kolbitsch, J. (2007b). WordFlickr: A solution to the vocabulary problem in social tagging systems. In: *Proceedings of I-MEDIA ,07 and I-SEMANTICS ,07*. Graz, Austria, September 5-7, 2007 (S. 77-84).
- Kroski, E. (2005). The Hive Mind: Folksonomies and user-based tagging. [Blog post; 2005-12-07]. Online: <http://infotangle.blogspot.com/2005/12/07/the-hive-mind-folksonomies-and-user-based-tagging>.
- Lancaster, F.W. (2003). *Indexing and Abstracting in Theory and Practice*. Champaign: University of Illinois. 3rd ed.
- Laniado, D., Eynard, D., & Colombetti, M. (2007). A semantic tool to support navigation in a folksonomy. In: *Proceedings of the 18th Conference on Hypertext and Hypermedia* (S. 153-154). New York: ACM.
- Layne, S. (2002). Subject access to art images. In: *Baca, M. (ed.): Introduction to Art Image Access* (S. 1-19). Los Angeles: Getty Research Institute.
- Lin, X., Beaudoin, J.E., Bui, Y., & Desai, K. (2006). Exploring characteristics of social classification. In: *17th ASIS&T SIG/CR Classification Research Workshop*.
- Lüth, J. (2007). Inhaltserschließung durch Nutzerinnen und Nutzer. Ergebnisse eines Tests mit Internetquellen der virtuellen Fachbibliothek EconBiz. In: *Ockenfeld, M. (Hrsg.): Information in Wissenschaft, Bildung und Wirtschaft*. 29. Online-Tagung der DGI. 59. Jahrestagung der DGI (S. 113-120). Frankfurt am Main: DGI.
- Lund, B., Hammond, T., Flack, M., & Hannay, T. (2005). Social bookmarking tools (II). A case study - Connotea. *D-Lib Magazine*, 11(2005)4.
- Lux, M., Granitzer, M. & Kern, R. (2007). Aspects of broad folksonomies. In: *Proceedings of the 18th International Conference on Database and Expert Systems Applications (DEXA 2007)* (S. 283-287). Washington, DC: IEEE Computer Society.
- Macgregor, G., & McCulloch, E. (2006). Collaborative tagging as a knowledge organisation and resource discovery tool. *Library Review* 55(2006)5, 291-300.
- MacLaurin, M.B. (2005). Selection-based item tagging. Patentanmeldung Nr. US 2007/0028171 A1. Eingereicht: 29.7.2005. Publiziert: 1.2.2007.
- Mai, J.M. (2006). Contextual analysis for the design of controlled vocabularies. *Bulletin of the American Society for Information Science and Technology* 33(2006)1, 17-19.
- Markey, K. (1986). *Subject Access to Visual Resource Collections*. Westport: Greenwood.
- Markkula, M., & Sormunen, E. (2000). End-user searching challenges indexing practices in the digital newspaper photo archive. *Information Retrieval* 1(2000), 259-285.
- Marlow, C., Naaman, M., Boyd, D., & Davis, M. (2006). HT06, tagging paper, taxonomy, Flickr, academic article, to read. In: *Proceedings of the 17th Conference on Hypertext and Hypermedia* (S. 31-40). New York: ACM.
- Mathes, A. (2004). *Folksonomies - Cooperative Classification and Communication Through Shared Metadata*. Urbana, Ill.: University of Illinois Urbana-Campaign / Graduate School of Library and Information Science.
- McFedries, P. (2006). Folk wisdom. *IEEE Spectrum*, 43(2006)Febr., 80.
- Merholz, P. (2004). Metadata for the masses. [Blog post, 2004-10-19]. Online: www.adaptivepath.com/publications/essays/archives/000361.php.
- Mika, P. (2005). Ontologies are us: A unified model of social networks and semantics. *Lecture Notes in Computer Science* 3729, 522-536.
- Millen, D.R., Feinberg, J., & Kerr, B. (2006). DO-GEAR. Social bookmarking in the enterprise. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (S. 111-120). New York: ACM.
- Miller, G.A. (1998). Nouns in WordNet. In *Fellbaum, C. (ed.), WordNet. An Electronic Lexical Database* (S. 23-46). Cambridge, Mass., London: MIT Press.
- Müller-Prove, M. (2007). Taxonomien und Folksonomien - Tagging als neues HCI-Element. In *i-com*, 6(1), 14-18.
- Morrison, P.J. (2007). Why are they tagging, and why do we want them to? *Bulletin of the American Society for Information Science and Technology* 34(2007)1, 12-15.
- Munk, T.B., & Mork, K. (2007). Folksonomy, the power law and the significance of the least effort. *Knowledge Organization* 34(2007)1, 16-33.
- Murison, J. (2005). Messageboard topic tagging: User tagging of collectively owned community content. In: *Proceedings of the 2005 Conference on Designing for User eXperience* (art. no 5). New York: American Institute of Graphic Art.
- Neal, D. (2007). Folksonomies and image tagging: Seeing the future? *Bulletin of the American Society for Information Science and Technology* 34(2007)1, 7-11.
- Niwa, S., Doi, T., & Honiden, S. (2006). Web page recommender systems based on folksonomy mining for ITNG'06 submissions. In: *Third Conference on Information Technology: New Generation (ITNG'06)* (S. 388-393). Washington, DC: IEEE Computer Society.
- Noruzi, A. (2006). Folksonomies: (un)controlled vocabulary? *Knowledge Organization* 33(2006)4, 199-203.
- Noruzi, A. (2007). Folksonomies: Why do we need controlled vocabulary? *Webology* 4(2), editorial 12.
- O'Reilly, T. (2005). What is Web 2.0. Design patterns and business models for the next generation of software. Online: www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html.
- Ornager, S. (1995). The newspaper image database. Empirical supported analysis of users' typology and word association clusters. In: *Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (S. 212-218). New York: ACM.
- Ott, E.S. (2006). Picture tagging. Patentanmeldung Nr. US 2007/0079321 A1. Eingereicht: 17.2.2006. Publiziert: 5.4.2007.
- Panofsky, E. (1975) *Sinn und Deutung in der bildenden Kunst*. Köln: DuMont.
- Panofsky, E. (2006). *Ikonographie und Ikonologie*. Köln: DuMont.
- Paolillo, J.C., & Penumarty, S. (2007). The social structure of tagging internet videos on del.icio.us. In: *Proceedings of the 40th Hawaii International Conference on System Sciences*.
- Peters, I. (2006a). Inhaltserschließung von Blogs und Podcasts im betrieblichen Wissensmanagement. In *Ockenfeld, M. (Hrsg.): Content. 28. Online-Tagung der DGI, 58. Jahrestagung der DGI. Proceedings* (S. 143-151). Frankfurt am Main: DGI.
- Peters, I. (2006b). Against folksonomies: Indexing blogs and podcasts for corporate knowledge management. In: *Jezzard, H. (ed.): Preparing for Information 2.0. Online Information 2006. Proceedings* (S. 93-97). London: Learned Information Europe.
- Peters, I. (2007). Web 2.0 und Fachinformation. *Folksonomies im Web 2.0. Password* 22(2007)5, 10-11.
- Peters, I. (2008). ASIS&T Annual Meeting 2007: Milwaukee, Wisconsin. *Joining Research and Practice: Social Computing and Information Science. Password* 23(2008)1, 16-19.
- Peters, I., & Stock, W.G. (2006). Corporate Blogs im Wissensmanagement. *Wissensmanagement* 8(2006)6, 40-41.
- Peters, I., & Stock, W.G. (2007a): *Web 2.0 im Unternehmen. Wissensmanagement* 9(2007)4 22-25.
- Peters, I., & Stock, W.G. (2007b): Folksonomy and information retrieval. In: *Proceedings of the 70th Annual Meeting of the American Society for Information Science and Technology Vol. 45* (S. 1510-1542). (CD-ROM).
- Peters, I., & Weller, K. (2008). Paradigmatic and syntagmatic relations in knowledge organization systems. *Information - Wissenschaft und Praxis*, 59(2008)2, 100-107.
- Peterson, E. (2006). Beneath the metadata. Some philosophical problems with folksonomies. *D-Lib Magazine* 12(2006)11.
- Pluzhenskaia, M. (2006). Folksonomies or fauxsonomies: How social is social bookmarking? In: *17th ASIS&T SIG/CR Classification Research Workshop. Abstracts of Posters* (S. 23-24).
- Quintarelli, E. (2005). Folksonomies: Power to the people. Paper presented at the ISKO Italy Unimib meeting, Milan, June 24, 2005.
- Rasmussen, E.M. (1997): Indexing images. *Annual Review of Information Science and Technology* 32, 169-196.
- Reamy, T. (2006). Folksonomies and complexity theory: Evolved information structures. In: *Jezzard, H. (ed.): Preparing for Information 2.0. Online Information 2006. Proceedings* (S. 111-113). London: Learned Information Europe.
- Rocchio, J.J. (1971). Relevance feedback in information retrieval. In: *Salton, G. (ed.): The SMART Retrieval System - Experiments in Automatic Document Processing* (S. 313-323). Englewood Cliffs, N.J.: Prentice Hall.
- Russell, T. (2006). Clouidalicious: Folksonomy over time. In: *Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries* (S. 364). New York: ACM.
- Schmitz, C., Hotho, A., Jäschke, R., & Stumme, G. (2006). Mining association rules in folksonomies. In: *Batagelj, V., Bock, H.H., Ferligoj, & Žiberna, A. (eds.): Data Science and Classification* (S. 261-270). Berlin, Heidelberg: Springer.

Schmitz, P. (2006). Inducing ontology from Flickr tags. In: Proceedings of the 15th International WWW Conference. Collaborative Web Tagging Workshop.

Shatfort, S. (1986). Analyzing the subject of a picture. A theoretical approach. Cataloguing and Classification Quarterly 6(1986)3, 39-62.

Shirky, C. (2004). Folksonomy. [Blog post; 2004-08-24]. Online: <http://many.corante.com/archives/2004/08/25/folksonomy.php>

Shirky, C. (2005). Ontology is overrated: Categories, Links, and Tags. www.shirky.com/writings/ontology_overrated.html.

Sinclair, J., & Cardew-Hall, M. (2008). The folksonomy tag cloud: When is it useful? Journal of Information Science 34(2008)1, 15-29.

Smith, G. (2004). Folksonomy: Social classification. [Blog post; 2004-08-03]. http://atomic.org/archives/2004/08/folksonomy_social_classification.html.

Smith, M.K. (2006). Viewer tagging in art museum: Comparisons to concepts and vocabularies of art museum visitors. In: 17th ASIS&T SIG/CR Classification Research Workshop.

Specia, L., & Motta, E. (2007). Integrating folksonomies with the semantic Web. Lecture Notes in Computer Science, 4519, 624-639.

Spiteri, L.F. (2006). The use of folksonomies in public library catalogues. The Serials Librarian 51(2006)2, 75-89.

Spiteri, L.F. (2007). Structure and form of folksonomy tags: The road to the public library catalogue. Webology 4(2007)2, art. 41.

Spyns, P., de Moor, A., Vandebussche, J., & Meersman, R. (2006). From folksonomies to ontologies: How the twain meet. Lecture Notes in Computer Science 4275, 738-755.

Stock, W.G. (2006). On relevance distributions. Journal of the American Society for Information Science and Technology 57(2006)8, 1126-1129.

Stock, W.G. (2007a). Information Retrieval. Informationen suchen und finden. München, Wien: Oldenbourg.

Stock, W.G. (2007b). Folksonomies and science communication. A mash-up of professional science databases and web 2.0 services. Information Services & Use, 27(2007)3, 97-103.

Stock, W.G., & Stock, M. (2008). Wissensrepräsentation. Informationen auswerten und bereitstellen. München, Wien: Oldenbourg.

Stock, W.G., & Weber, S. (2006). Facets of informetrics. Information – Wissenschaft und Praxis 57(2006)8, 385-389.

Surowiecki, J. (2004). The Wisdom of Crowds. New York: Doubleday (dt. Die Weisheit der Vielen. München: Bertelsmann, 2005).

Toffler, A. (1980). The Third Wave. New York: Morrow.

Tonkin, E. (2006a). Folksonomies: The fall and rise of plain-text tagging. Ariadne, 47.

Tonkin, E. (2006b). Searching the long tail: Hidden structure in social tagging. In: 17th

ASIS&T SIG/CR Classification Research Workshop.

Trant, J. (2006a). Exploring the potential for social tagging and folksonomy in art museums: Proof of concept. New Review of Hypermedia and Multimedia 12(2006)1, 83-105.

Trant, J. (2006b). Social classification and folksonomy in art museums: Early data from the steve.museum.tagger prototype. In: 17th ASIS&T SIG/CR Classification Research Workshop.

Vander Wal, T. (2004). Feed on this. [Blog post; 2004-10-03]. www.vanderwal.net/random/category.php?cat=153.

Vander Wal, T. (2005). Explaining and showing broad and narrow folksonomies. [Blog post; 2005-02-21]. www.vanderwal.net/random/category.php?cat=153.

Veres, C. (2006a). The language of folksonomies: What tags reveal about user classification. Lecture Notes in Computer Science 3999, 58-69.

Veres, C. (2006b). Concept modeling by the masses: Folksonomy structure and interoperability. Lecture Notes in Computer Science 4215, 325-338.

Voß, J. (2007). Tagging, folksonomy & co. – Renaissance of manual indexing? In: Obwald, A., Stempfhuber, M., & Wolff, C. (Hrsg.): Open Innovation. Neue Perspektiven im Kontext von Information und Wissen. ISI 2007 (S. 243-254). Konstanz: Universitätsverlag.

Wang, X., Bai, R., & Liao, J. (2007). Chinese weblog pages classification based on folksonomy and support vector machines. Lecture Notes in Computer Science 4476, 309-321.

Weiss, A. (2005). The power of collective intelligence. netWorker, 9(2005)3, 16-23.

Winget, M. (2006). User-defined classification on the online photo sharing site Flickr ... Or, how I learned to stop worrying and love the million typing monkeys. In: 17th ASIS&T SIG/CR Classification Research Workshop.

Weller, K., & Peters, I. (2007). Reconsidering relationships for knowledge representation. In: Proceedings of I-KNOW '07. Graz, Austria, September 5-7, 2007 (S. 493-496).

Wu, H., Zubair, M., & Maly, K. (2006). Harvesting social knowledge from folksonomies. In: Proceedings of the 17th Conference on Hypertext and Hypermedia (S. 111-114). New York: ACM.

Xu, Z., Fu, Y., Mao, J., & Su, D. (2006). Towards the semantic Web: Collaborative tag suggestions. In: Proceedings of the 15th International WWW Conference. Collaborative Web Tagging Workshop.

Yeung, C.M.A., Gibbins, N., & Shadbolt, N. (2008). Mutual contextualization in tripartite graphs of folksonomies. Lecture Notes in Computer Science 4825, 966-970.

Zhang, L., Wu, X., & Yu, Y. (2006). Emergent semantics from folksonomies: A quantitative study. Lecture Notes in Computer Science 4090, 168-186.

Inhaltliche Erschließung, Zusammenarbeit, Folksonomy, Wissensrepräsentation, Information Retrieval, Prosumer, Broad Folksonomy, Narrow Folksonomy, Index-Tag, Such-Tag, Tag-Verteilung, Long Tail, Long Trunk, Power Law, invers-logistische Verteilung, Yule-Simon-Prozess, Ofness, Aboutness, Isness, Flickr, Delicious, YouTube, Natural Language Processing, Relevance Ranking, Interestingness Ranking, ähnliche Dokumente, ähnliche Nutzer, Community

DIE AUTOREN

Isabella Peters, M.A.



ist wissenschaftliche Mitarbeiterin der Abteilung für Informatikswissenschaft der Heinrich-Heine-Universität Düsseldorf. Ihre Forschungsschwerpunkte liegen bei Web-2.0-Diensten, Folksonomies, der Blog- und Podcastosphäre sowie im betrieblichen Wissensmanagement. isabella.peters@uni-duesseldorf.de

Univ.-Prof. Dr. Wolfgang G. Stock



ist Leiter der Abteilung für Informatikswissenschaft der Heinrich-Heine-Universität Düsseldorf. Seine Forschungen liegen vor allem in Gebieten des Information Retrieval, der Wissensrepräsentation und der Informatik. stock@phil-fak.uni-duesseldorf.de



Effektive Wissensverteilung mit DOMESTIC

Entdeckungs- und Erinnerungs-Recherche!

Alles dazu unter ...

www.domestic.de