

Das Entstehen von Semantik in BibSonomy

Andreas Hotho¹, Robert Jäschke^{1,2}, Christoph Schmitz¹, Gerd Stumme^{1,2}

¹ Fachgebiet Wissensverarbeitung, Fachbereich Mathematik/Informatik

Universität Kassel, Wilhelmshöher Allee 73, 34121 Kassel

<http://www.kde.cs.uni-kassel.de>

² Forschungszentrum L3S, Expo Plaza 1, 30539 Hannover, <http://www.l3s.de>

Zusammenfassung

Immer mehr Soziale-Lesezeichen-Systeme entstehen im heutigen Web. In solchen Systemen erstellen die Nutzer leichtgewichtige begriffliche Strukturen, so genannte Folksonomies. Ihren Erfolg verdanken sie der Tatsache, dass man keine speziellen Fähigkeiten benötigt, um an der Gestaltung mitzuwirken. In diesem Artikel beschreiben wir unser System BibSonomy. Es erlaubt das Speichern, Verwalten und Austauschen sowohl von Lesezeichen (Bookmarks) als auch von Literaturreferenzen in Form von BibTeX-Einträgen. Die Entwicklung des verwendeten Vokabulars und der damit einhergehenden Entstehung einer gemeinsamen Semantik wird detailliert diskutiert.

1 Einleitung

Eine neue Art von Anwendungen, die auf die Beteiligung der Nutzer und einfache Benutzerschnittstellen baut, bildet den Kern des entstehenden „Web 2.0“. Diese Anwendungen ergänzen die aktuellen Bemühungen im Bereich des Semantic Web. Dabei spielen benutzerbezogene Plattformen wie z.B. Wikis, Blogs und soziale Systeme zum Teilen von Ressourcen eine immer zentralere Rolle. Solche Systeme dienen zunächst zum Veröffentlichen von Informationen, sie können aber auch als einfache Wissensmanagementsysteme angesehen werden.

Systeme zur gemeinsamen Verwaltung von Ressourcen sind webbasierte Systeme, die den Benutzern das Hochladen und Verschlagworten von Ressourcen mit beliebigen Worten, genannt *Tags*, ermöglichen. Die Zuordnung von Tags zu Ressourcen durch den Benutzer stellt eine leichtgewichtige Wissensrepräsentation dar, die man *Folksonomy* nennt. Folksonomies sind Bottom-Up-Ansätze, die die formalen Methoden aus dem Bereich des Semantic Web ergänzen, indem sie sich auf sogenannte „Emergent Semantics“ (Staab et al, 2002a; Steels, 1998) – d. h. entstehende Semantik – stützen. Semantik ergibt sich dabei durch die Konvergenz des verwendeten Vokabulars. Der zentrale Unterschied zum „klassischen“ ontologiebasierten Ansatz besteht in dem Ziel, die Anforderungen der Benutzer, die keine Experten sind, möglichst umfassend zu respektieren, und sie nicht mit dem Overhead eines formalen Ansatzes zu konfrontieren. Dies schließt aber nicht aus, dass intelligente Techniken zur Unterstützung der Nutzer in ein solches System integriert werden, ohne dabei jedoch in den Vordergrund zu treten.

Eine solche benutzergetriebene Verschlagwortung von Ressourcen wird „Folksonomy“ genannt - ein Kunstwort aus „Folk“ (Leute) und „Taxonomy“. Folksonomy-basierte Systeme, wie z.B. Flickr¹ für Bildersammlungen oder del.icio.us² für Bookmarks (Lesezeichen), haben in den letzten zwei Jahren eine große Anzahl an Benutzern angezogen. Unser System *BibSonomy*³ erlaubt das Verwalten und Teilen von Bookmark- und Publikationseinträgen (siehe Abb. 1). Der Grund für den unmittelbaren Erfolg solcher Systeme liegt in der Tatsache, dass keine spezifischen Fähigkeiten benötigt werden, um aktiv mit einem solchen System arbeiten zu können. Der direkte Vorteil für jeden Benutzer, browserunabhängig seine Bookmarks ohne zu großen Mehraufwand zentral organisieren zu können, zieht weitere Benutzer an und macht die Systeme noch attraktiver.⁴ Mit Hilfe solcher Systeme ist es einer großen Nutzergemeinde möglich, in relativ kurzer Zeit eine sehr große Menge an Informationen zusammenzutragen. Die weit verbreitete Nutzung solcher Systeme zeigt deutlich die Fähigkeit, den Flaschenhals zu überwinden, den die Wissensakquisition bisher darstellte und der ein entscheidender Nachteil vieler wissensbasierte Systeme war.

Der vorliegende Artikel fasst die Arbeiten und Ideen aus (Hotho et al, 2006a; Schmitz et al, 2006; Hotho et al, 2006b) zusammen und ist wie folgt strukturiert: Im Abschnitt 2 führen wir ein formales Modell und das System „BibSonomy“ ein. Im Abschnitt 3 präsentieren wir die Anwendung von Assoziationsregeln auf Folksonomies, sowie den FolkRank-Algorithmus – beides unterstützt die Entstehung von Semantik in solchen Strukturen. Der Artikel endet mit einer Diskussion ähnlicher Arbeiten, einer Zusammenfassung und einem Ausblick.

2 *BibSonomy* – ein Folksonomy-basiertes Soziales Lesezeichen-System

Dieser Abschnitt beschreibt das System *BibSonomy*, das in unserer Gruppe entwickelt wurde. *BibSonomy* erlaubt es den Benutzern, Bookmarks und Referenzen auf Publikationen zu verwalten. Das Datenmodell des Publikationsteils basiert auf *BibTeX*, einem populären Literaturverwaltungssystem für *LaTeX*. *BibSonomy* implementiert das formale Modell einer Folksonomy, das in (Hotho et al, 2006a) entworfen wurde. Im Folgenden wird dieses Modell kurz erläutert.

Eine Folksonomy besteht aus Mengen von Benutzern, Ressourcen und Tags sowie einer benutzerspezifischen Zuordnung von Tags zu Ressourcen. Wir formalisieren dies folgendermaßen:

Definition 1 *Eine Folksonomy ist ein Tupel $F:=(U,T,R,Y,<)$, wobei U , T und R endliche Mengen sind, deren Elemente man Benutzer, Tags bzw. Ressourcen nennt.*

¹ <http://www.flickr.com>

² <http://del.icio.us>

³ <http://www.bibsonomy.org>

⁴ Diskussionen auf der del.icio.us Mailingliste kann man entnehmen, dass die Anzahl der Nutzer auf über Fünfhunderttausend geschätzt wird.



Abbildung 1: BibSonomy organisiert gleichzeitig Bookmarks und BibTeX-Referenzen.

Y ist eine ternäre Relation zwischen diesen Mengen, d. h. $Y \subseteq U \times T \times R$, deren Elemente Tag Assignments heißen. Die Menge von Tag Assignments, die ein Benutzer zu einer Ressource herstellt, heißt Eintrag oder Post.

\prec ist eine benutzerspezifische Unter-/Obertag-Relation, d. h. $\prec \subseteq U \times T \times T$, die Is-A-Relation genannt wird.

Die Personomy P_u eines Users $u \in U$ ist die Beschränkung von F auf u , d. h. $P_u := (T_u, R_u, I_u, \prec_u)$ wobei $I_u := \{(t, r) \in T \times R \mid (u, t, r) \in Y\}$, $T_u := \pi_1(I_u)$ und $R_u := \pi_2(I_u)$ gilt; dabei bezeichnet π_i die Projektion auf die i -te Dimension.

Wenn man die Is-A-Relation nicht betrachten möchte, kann man die Folksonomy vereinfacht als ein Quadrupel $F := (U, T, R, Y)$ notieren. Diese Struktur ist in der formalen Begriffsanalyse (Wille, 1982a; Ganter and Wille, 1999) als *triadischer Kontext* bekannt. Eine äquivalente Sicht ist die eines tripartiten (ungerichteten) Hypergraphen $G = (V, E)$, wobei $V = U \cup T \cup R$ die Menge der Knoten und $E = \{(u, t, r) \mid (u, t, r) \in Y\}$ die Menge der Hyperkanten sind.

2.1 Benutzerinterface

Abbildung 1 zeigt eine typische Liste von Bookmark- und Publikationseinträgen für das Tag *web*. Die Seite ist in vier Teile gegliedert: der Kopf enthält Navigations-elemente und Suchfelder. Den Hauptteil der Seite nehmen zwei Listen ein, die Bookmark- und BibTeX-Einträge enthalten, die nach dem Eingabedatum absteigend sortiert sind. Im rechten Teil der Seite befindet sich eine Liste von Tags, die mit den angezeigten Einträgen verwandt sind. Alle Seiten, die Einträge zeigen, sind nach diesem Schema aufgebaut und erlauben so die Navigation entlang aller Dimensionen der Folksonomy.

Abb. 2 und Abb. 3 zeigen eine detaillierte Sicht von Bookmark- und Publikationseinträgen aus Abb. 1. Die erste Zeile des Bookmark-Eintrags stellt in Fettdruck den Titel des Bookmarks dar, hinterlegt mit dem Hyperlink zur entsprechenden URL. Die zweite Zeile ist eine (optionale) Beschreibung, die der Benutzer hinzugefügt hat. Die letzten beiden Zeilen zeigen die Tags, die der User dem Eintrag zugewiesen hat (*web*, *service*, *tutorial*, *guidelines*, *api* und *rest*), sowie den Benutzernamen (*hotho*), gefolgt von einer Angabe, wie viele Benutzer diese Ressource in ihrer Personomy haben.

REST web services

Good intro to the REST "architecture"

to web service tutorial guidelines api rest by hotho and 3 other people on 2006-04-04 16:11:47 copy

Abbildung 2: Ein Bookmark-Eintrag

Semantic Network Analysis of Ontologies

Bettina Hoser and Andreas Hotho and Robert Jäschke and Christoph Schmitz and Gerd Stumme. *Proceedings of the 3rd European Semantic Web Conference* (accepted for publication) (2006)

to web 2006 social ontology myown semantic analysis network sna by hotho and 1 other person on 2006-04-06 21:32:23 pick copy URL BibTeX

Abbildung 3: Ein Publikations-Eintrag

Diese Teile sind jeweils mit Links zu den Tag-Seiten des Benutzers, seiner Überblicksseite bzw. zu einer Seite, die alle vier Einträge (d. h. des Benutzers *hotho* und der drei anderen Benutzer) zu dieser Ressource zeigt, versehen. Der letzte Teil zeigt den Zeitstempel des Eintrags und die Aktionen, die der Benutzer auf diesem Eintrag durchführen kann: *edit* (bearbeiten) und *delete* (löschen) für seine eigenen Einträge, oder *copy* (kopieren) für Einträge anderer Benutzer.

Die Struktur des Publikationseintrags ist sehr ähnlich; er zeigt bibliographische Details statt der Beschreibung. Der Titel des Eintrag verweist auf eine Seite, die alle bibliographischen Informationen über den Eintrag enthält. Die Aktionen für den BibTeX-Eintrag sind die gleichen wie für Bookmarks, zusätzlich gibt es die Möglichkeit, den Eintrag für einen späteren Download vorzumerken (*pick*), die URL des Eintrags direkt anzuspringen oder den BibTeX-Quellcode anzuzeigen.

2.2 Beziehungen zwischen Tags

Folksonomy-Systeme sind u. A. deswegen so populär, weil die Teilnahme einfach ist und keine besonderen Vorkenntnisse erfordert. Andererseits wurden von Benutzern von BibSonomy Möglichkeiten gefordert, die eigenen Tags besser zu strukturieren. Eine benutzerspezifische Relation \prec zwischen Tags, wie im Folksonomy-Modell beschrieben, ist ein Weg, die eigenen Tags zu ordnen. Daher wurde diese Relation in BibSonomy integriert.

Um die Spezifikation der \prec -Relation während der Erstellung eines Eintrags zu ermöglichen, benutzen wir die Zeichenketten \prec und \succ ; d. h. wenn ein Benutzer u als Tag $t_1 \prec t_2$ verwendet, werden die Tags t_1 und t_2 zum jeweiligen Eintrag hinzugefügt und zusätzlich ein Tripel (u, t_1, t_2) in die Relation \prec aufgenommen. Das Tag $t_2 \prec t_1$ wird interpretiert als $t_1 \succ t_2$.

Man kann dies lesen als „ t_1 ist ein t_2 “ oder „ t_1 ist ein *Untertag* des *Obertags* t_2 “. Es gibt weitere Möglichkeiten, die Relation \prec zu bearbeiten, z. B. in einem Relationeneditor.

Die Relation \prec wird in verschiedenen Situationen verwendet. Zum Einen kann der Benutzer seine Tags strukturieren, indem er Untertags eines bestimmten Obertags ein- und ausblendet und so die Tags in einer Hierarchie darstellt. Andererseits bietet BibSonomy die Möglichkeit, auf einer Seite alle Einträge zu einem Tag sowie aller seiner Untertags anzuzeigen.

Die Aufnahme der \prec -Relation in das System wirft allerdings Fragen auf, deren Lösung noch offen ist:

- Wie soll mit Zyklen umgegangen werden, also z.B. $u \in U$ und $t_1, \dots, t_m \in T$ mit $(u, t_i, t_{i+1}) \in \prec$ für $i = 1, \dots, m-1$ und $(u, t_m, t_1) \in \prec$?
- Wie soll Äquivalenz oder Nicht-Äquivalenz von Tags modelliert werden?
- Soll der vollständige transitive Abschluß der Relation \prec betrachtet werden? Wenn ja: wie kann dies effizient geschehen?
- Wie sollen Anfragen folgender Art gestellt werden: „Alle Einträge mit dem Tag *howto* und einem Untertag von *programming*“?

3 Entstehende Semantik in Folksonomien

Im Folgenden wird die Anpassung und Anwendung von Information-Retrieval- und Data-Mining-Algorithmen zur Unterstützung von entstehender Semantik in Folksonomy-Systemen diskutiert. Insbesondere werden Anpassungen von Assoziationsregel-Mining (Schmitz et al, 2006) und PageRank (Hotho et al, 2006b), an die triadische Struktur von Folksonomies vorgestellt. Die Verfahren wurden an Daten aus del.icio.us untersucht und analysiert, da dieses System zum Zeitpunkt unserer Experimente über eine wesentlich größere Datenbasis verfügte. Die Implementierung dieser Verfahren in BibSonomy wird in naher Zukunft angestrebt.

3.1 Mining von Assoziationsregeln in Folksonomies

Diskussionen auf den einschlägigen Mailinglisten wie z. B. *delicious-discuss*⁵ zeigen, dass es einen Bedarf für mehr Struktur in Folksonomies gibt, als es eine unstrukturierte Menge von Tags erlaubt. Diese Struktur kann beispielsweise aus der vorgenannten Relation \prec bestehen. Eine Möglichkeit, eine solche Struktur bereitzustellen, ohne dass der Benutzer diese selbst pflegen muss, ist die Anwendung von Techniken des Ontologie-Lernens (Cimiano et al, 2005).

Benutzer drücken durch ihr Tagging-Verhalten die Bedeutung von Ressourcen aus. Beim Verschlagworten einer gegebenen Ressource werden oft generelle und speziellere Tags vermischt. So könnte z. B. eine Seite über XSLT-Stylesheets mit dem Tag *xslt* sowie mit dem Oberbegriff *xml* versehen werden. Durch die Berechnung von Assoziationsregeln (Agrawal et al, 1993) können solche Beziehungen extrahiert werden.

Assoziationsregeln im herkömmlichen Sinne können nicht direkt auf Folksonomies berechnet werden, da diese eine triadische Struktur tragen. Entweder muss also eine Art triadischer Assoziationsregeln definiert werden, oder man muss vorab die Folksonomy in eine zweidimensionale Form überführen. In dieser Arbeit verfolgen wir den letztgenannten Ansatz. Für eine genauere Definition der Projektionen siehe (Schmitz et al, 2006). Der erstgenannte Ansatz wird in (Jäschke et al, 2006) eingeführt.

Zur Demonstration dieses Ansatzes betrachten wir die Projektion auf den zweidimensionalen Datensatz $K_1 := (U \times R, T, I_1)$ mit $I_1 := \{(u, r, t) \mid (u, t, r) \in Y\}$. Die

⁵ <http://lists.del.icio.us/pipermail/discuss/>

Menge T entspricht hier der Menge von Waren in der klassischen Warenkorbanalyse, und $U \times R$ der Menge von Transaktionen.

Eine Assoziationsregel $A \rightarrow B$ in K_1 ist dabei zu interpretieren als *Benutzer, die Ressourcen die Tags aus der Menge A zuweisen, verwenden auf diesen Ressourcen oft auch die Tags aus B* . Diese Art von Regeln könnte z. B. zum Empfehlen einer Tag-Hierarchie verwendet werden. Wenn also ein Benutzer Tags aus A verwendet, kann das System zusätzlich die Tags aus B vorschlagen.

Wir haben diesen Ansatz auf Daten von del.icio.us evaluiert. Dazu standen uns Daten von $|U|=75.242$ Benutzern, die $|T|=533.191$ Tags auf $|R|=3.158.297$ Ressourcen anwendeten, zur Verfügung; insgesamt werden dabei $|Y|=17.362.212$ Tag Assignments verwendet.

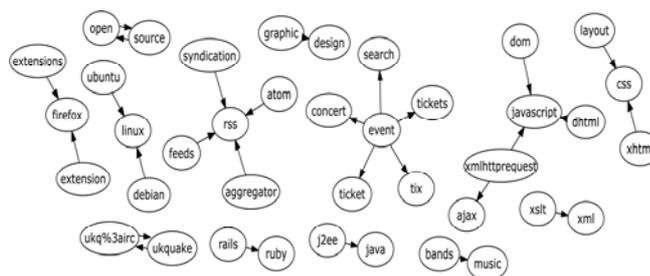


Abbildung 4: Zweielementige Regeln auf der Tag-Menge von del.icio.us mit 0,05 % Support und 50% Confidence

Abbildung 4 zeigt ein Beispiel aus (Schmitz et al, 2006), welches auf den del.icio.us-Daten berechnet wurde. Hierbei betrachten wir, welche Tags besonders oft zusammen mit anderen Tags vorkommen (z. B.: wenn ein Benutzer eine bestimmte Ressource mit *xslt* kennzeichnet, dann wird er oft auch das Tag *xml* verwenden).

Im vorliegenden Fall deutet eine Assoziationsregel auf eine Unter-/Obertag-Beziehung hin, die dem Benutzer für seine \prec -Relation vorgeschlagen werden kann. Dieser Ansatz kann mit volltextbasierten Verfahren wie (Cimiano et al, 2005) kombiniert werden, wenn die Ressourcen Volltextdokumente wie etwa Webseiten sind.

Abbildung 4 zeigt alle Regeln mit jeweils einelementiger Prämisse und Konklusion sowie einem Support von mindestens 0.05 % und einer Konfidenz von mindestens 50 %, die aus K_1 berechnet wurden. Man sieht bspw. dass Benutzer, die eine Webseite mit *debian* taggen, häufig auch das Tag *linux* benutzen, und dass Webseiten über *bands* sich wahrscheinlich auch mit *music* befassen. Diese Resultate können als Empfehlungen benutzt werden, um Tags zu wählen, die beim späteren Auffinden dieser Ressourcen hilfreich sind.

Eine andere Sicht dieser Regeln ist die als Subsumptionsrelation, womit wir das Bestimmen der Assoziationsregeln als einen Schritt zum Lernen einer Taxonomie ansehen. Wenn viele Ressourcen, die das Tag *xslt* tragen, auch mit *xml* versehen sind, so ist dies ein Hinweis, dass *xml* ein Oberbegriff von *xslt* sein kann, der (semi-) automatisch zur Relation \prec hinzugefügt werden kann. Abbildung 4 zeigt auch zwei Paare von Tags, die oft zusammen vorkommen, ohne dass es eine bestimmte Rich-

tung in der Assoziation gibt: *open source* wird als Kompositum meist zusammen verwendet, und *ukquake* and *ukq:irc* werden als Paar automatisch vom Betreiber eines bestimmten Chat-Kanals vergeben.

3.2 FolkRank

Algorithmen wie der bekannte PageRank (Brin and Page, 1998) berechnen Rankings auf Knoten mit der Idee, dass ein Knoten dann wichtig ist, wenn viele andere wichtige Knoten auf diesen zeigen. Wir wenden dasselbe Prinzip auf den tripartiten Hypergraphen der Folksonomy an, um Rankings auf Benutzern, Tags oder Ressourcen zu erstellen. Eine Ressource ist also wichtig, wenn sie mit wichtigen Tags oder Benutzern verbunden ist. Dies gilt symmetrisch auch für Tags und Benutzer. Wie haben den Algorithmus *FolkRank* entwickelt, eine Variante von PageRank, die auf die Struktur einer Folksonomy angepasst ist (Hotho et al, 2006b).

Zunächst wird dabei der tripartite Hypergraph der Folksonomy in einen ungerichteten tripartiten Graphen überführt, indem jeweils zwischen dem Tag, dem Benutzer und der Ressource, die zusammen in einer Hyperkante vorkommen, eine ungerichtete, gewichtete Kante gesetzt wird.

Auf diesen Graphen kann der PageRank-Algorithmus angewandt werden. Mit dem „Random Surfer“-Vektor von PageRank bietet sich die Möglichkeit, durch Höherbewerten einzelner oder mehrerer Tags, Benutzer und/oder Ressourcen ein themenspezifisches Ranking zu erhalten.⁶

Eine direkte Anwendung dieses Ansatzes liefert allerdings Resultate, die stark von den global wichtigen Knoten dominiert werden und nicht oder nur schwach die Präferenzen widerspiegeln. Dies resultiert aus der sehr schiefen Verteilung der Vorkommen verschiedener Elemente in der Folksonomy und aus der Tatsache, dass im ungerichteten Graphen einmal verteiltes Gewicht zum Teil unmittelbar wieder entlang derselben Kanten zurückgegeben wird. Um dies zu kompensieren, wird in FolkRank die Differenz zweier Läufe des adaptierten PageRank betrachtet: zusätzlich zum Lauf mit Präferenzvektor wird ein Lauf ohne Präferenz berechnet, und die Differenz beider Läufe liefert das eigentliche Ergebnis. Das sich daraus ergebende Gewicht eines Knoten bezeichnen wir als *FolkRank* dieses Knotens. Dieser Ansatz hat zu deutlich besseren themenspezifischen Rankings geführt (Hotho et al, 2006b).

3.2.1 Empfehlungen berechnen mittels FolkRank

In der Originalarbeit zu PageRank (Brin and Page, 1998) wird schon auf die Möglichkeit hingewiesen, den Präferenzvektor des Random Surfers als Personalisierungsinstrument für Rankings heranzuziehen. Entsprechend bietet FolkRank die Möglichkeit, z. B. zu einem Tag eine Menge von relevanten Benutzern und Ressourcen zu bestimmen. Damit kann FolkRank benutzt werden, um innerhalb des

⁶ Dieser Ansatz ist nur mit dem Präferenzvektor interessant. Da der Graph ungerichtet ist, entspricht das Ergebnis ohne Präferenzvektor für einen Knoten genau der Summe der Gewichte der berührenden Kanten.

Folksonomy-Systeme Empfehlungen auszusprechen. Diese können an verschiedenen Stellen in der Benutzung des Systems Verwendung finden:

- Dokumente, die für einen Benutzer interessant sein können, können ihm vorgeschlagen werden. Diese Art von Empfehlung ermöglicht es, dem Benutzer ohne dessen Zutun – und ohne das Wissen, dass diese Inhalte überhaupt existieren – interessante Ressourcen zu entdecken.
- Wenn ein Tag benutzt wird, können verwandte Tags vorgeschlagen werden. Dies kann dazu benutzt werden, um die Konsolidierung der verwendeten Nomenklatur voranzutreiben und somit die Bildung eines einheitlichen Vokabulars zu unterstützen. Damit wird die Entstehung einer gemeinsamen Semantik beschleunigt.
- Andere Benutzer, die an verwandten Themen arbeiten, können erkannt und dem Benutzer vorgestellt werden. Damit wird der Wissenstransfer z. B. in Organisationen beschleunigt und die Bildung von Communities gefördert.

4 Andere Anwendungen

Neben dem allgegenwärtigen Gebrauch von Folksonomies als Werkzeug zur Verwaltung von Ressourcen im Web gibt es weitere nahe liegende Anwendungen.

Die Verwaltung von Multimediadaten ist eine davon. Da die Folksonomy den Inhalt von Dokumenten beschreibt, ohne auf diesen selbst zurückgreifen zu müssen, können Folksonomies für alle Arten von Multimedia-Inhalten genutzt werden, ohne aufwendige Extraktion von inhaltsbasierten Merkmalen zu benötigen.

Eine weitere Anwendung zeigt sich in Intranets (Millen et al, 2006). Da Intranets-Inhalte häufig aus Office-Dokumenten bestehen, die i. d. R. keine Navigation zu verwandten Inhalten ermöglichen, können Folksonomies hier als Mittel zur Strukturierung helfen. Außerdem unterstützen sie die Organisation der Inhalte durch die Benutzer selbst, was das Auffinden von benötigten Informationen gegenüber starren Taxonomien erleichtern kann.

Die Vorteile in Bezug auf die Erstellung und insbesondere Wartung von Wissensmanagement-Systemen im Intranet haben auch große Firmen erkannt. So erwägt z. B. IBM die Benutzung von Folksonomies im Intranet⁷ und Microsoft den Einsatz in seinem neuen Betriebssystem.⁸

5 Verwandte Arbeiten

Arbeiten wie (Hammond et al, 2005a; Lund et al, 2005) bieten einen guten Überblick über soziale Bookmarking-Systeme im Allgemeinen, während Golder und Huberman (2005) die Struktur von Folksonomies diskutieren, speziell die von del.icio.us, und sieben verschiedene Arten von Tags identifizieren.

⁷ http://thecommunityengine.com/home/archives/2005/03/ibms_intranet_a.html

⁸ http://thecommunityengine.com/home/archives/2005/03/xfolk_an_xhtml.html

Die Visualisierung der Popularität von Tags über die Zeit sowie ein entsprechender Algorithmus wird in (Dubinko et al, 2006) vorgestellt. Hotho et al (2006c) zeigen einen weiteren Algorithmus zur Entdeckung von Trends, der auch themenspezifische Trends aufzeigen kann. In Bezug auf Blogs untersuchen Paolillo et al (2005) mit Hilfe der Hauptkomponentenanalyse und Clustering-Techniken auf einem Netzwerk von FOAF-Profilen⁹ die zeitliche Veränderung der Benutzerstruktur.

Mika (2005) definiert ein Modell von semantisch-sozialen Netzwerken, um aus Daten von del.icio.us leichtgewichtige Ontologien zu extrahieren. Neben der Berechnung von Maßen wie z. B. dem Clustering-Koeffizienten oder der Zentralität von bestimmten Folksonomy-Elementen zeigt er, wie unter Ausnutzung von Kozitationsbeziehungen die Folksonomy geclustert und eine einfache Ontologie gelernt werden kann.

Weitere verwandte Arbeiten zu den Ansätzen aus Abschnitt 3 werden in den Publikationen (Schmitz et al, 2006) und (Hotho et al, 2006b) vorgestellt.

6 Zusammenfassung und Ausblick

Diese Arbeit beschreibt ein formales Modell für Folksonomies, auf dem unser System BibSonomy basiert, und den Aufbau des Systems selbst. Wir diskutieren Algorithmen, die dazu dienen, die Struktur einer Folksonomy zu untersuchen und Empfehlungen oder Rankings zu erstellen.

Die Struktur der Folksonomy bietet verschiedene Ansatzpunkte zur Untersuchung. Bisher haben wir die Graphenstruktur ausgenutzt und entsprechende Erweiterungen existierender Algorithmen auf diesen Graphen angewendet. Mit wachsenden Benutzerzahlen und Erweiterungen des Modells etwa um Relationen werden neue Verfahren benötigt. BibSonomy dient dabei als Entwicklungs- und Testplattform, wobei die Benutzer direkt von aktuellen Forschungsergebnissen profitieren können.

Wenn die Größe von Folksonomies wächst, werden Methoden zur Handhabung benötigt, die über einfache Anfragen z.B. nach Tags hinausgehen. Die interne Struktur der Folksonomy muss dazu besser organisiert werden; ein gangbarer Ansatz kann die Anwendung von Semantic-Web-Technologien sein. Die zentrale Frage bleibt dabei, wie die Vorteile ausgefeilterer Repräsentationsmechanismen genutzt werden können, ohne den Benutzer mit deren Starrheit und Komplexität zu konfrontieren und zu überfordern. Dies wird eine wichtige Frage für die Forschung im Bereich Semantic Web in den nächsten Jahren sein.

Danksagung. Teile dieser Forschung wurden durch das IST-Projekt „Nepomuk“ der EU finanziert (FP6-027705).

⁹ <http://www.foaf-project.org>

Literatur

- R. Agrawal, T. Imielinski, and A. Swami. Mining association rules between sets of items in large databases. In *Proc. of SIGMOD 1993*, pages 207–216. ACM Press, May 1993.
- S. Brin and L. Page. The Anatomy of a Large-Scale Hypertextual Web Search Engine. *Computer Networks and ISDN Systems* **30**(1-7):107–117, April 1998.
- P. Cimiano, A. Pivk, L. Schmidt-Thieme, and S. Staab. Learning taxonomic relations from heterogeneous sources of evidence. In *Ontology Learning from Text: Methods, Evaluation and Applications*, Frontiers in Artificial Intelligence, pages 59–73. IOS Press.
- M. Dubinko, R. Kumar, J. Magnani, J. Novak, P. Raghavan, and A. Tomkins. Visualizing tags over time. In *Proc. of the 15th International WWW Conference*, 2006.
- B. Ganter and R. Wille. *Formal Concept Analysis: Mathematical Foundations*. Springer, 1999.
- S. Golder and B. A. Huberman. The structure of collaborative tagging systems. Technical report, Information Dynamics Lab, HP Labs, August 2005.
- T. Hammond, T. Hannay, B. Lund, and J. Scott. Social Bookmarking Tools (I): A General Review. *D-Lib Magazine* **11**(4), April 2005.
- A. Hotho, R. Jäschke, C. Schmitz, and G. Stumme. BibSonomy: A social bookmark and publication sharing system. In A. de Moor, S. Polovina, and H. Delugach, editors, *Proc. of the Conceptual Structures Tool Interoperability Workshop at the 14th Int. Conf. on Conceptual Structures*, Aalborg, Denmark, July 2006(a). Aalborg University Press.
- A. Hotho, R. Jäschke, C. Schmitz, and G. Stumme. Information retrieval in folksonomies: Search and ranking. In York Sure and John Domingue, editors, *The Semantic Web: Research and Applications*, volume 4011 of *Lecture Notes in Computer Science*, pages 411–426, Heidelberg, June 2006(b). Springer.
- A. Hotho, R. Jäschke, C. Schmitz, and G. Stumme. Trend detection in folksonomies. In *Proc. First International Conference on Semantics And Digital Media Technology (SAMT) (to appear)*, Athens, Greece, December 2006(c).
- R. Jäschke, A. Hotho, C. Schmitz, B. Ganter, and G. Stumme. TRIAS - an algorithm for mining iceberg tri-lattices. *Proc. ICDM 2006 (to appear)*. 2006.
- B. Lund, T. Hammond, M. Flack, and T. Hannay. Social Bookmarking Tools (II): A Case Study - Connotea. *D-Lib Magazine* **11**(4), April 2005.
- P. Mika. Ontologies Are Us: A Unified Model of Social Networks and Semantics. In Y. Gil, E. Motta, V. R. Benjamins, and M. A. Musen, editors, *ISWC 2005*, volume 3729 of *LNCS*, pages 522–536. Springer-Verlag, November 2005.
- D. R. Millen, J. Feinberg, and B. Kerr. Dogear: Social bookmarking in the enterprise. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 111–120, New York, NY, USA, 2006. ACM Press.
- J. C. Paolillo, S. Mercure, and E. Wright. The social semantics of Livejournal FOAF: Structure and change from 2004 to 2005. In G. Stumme, B. Hoser,

C. Schmitz, and H. Alani, editors, *Proceedings of the 1st Workshop on Semantic Network Analysis at the ISWC 2005 Conference*, pages 69 – 80, 2005.

C. Schmitz, A. Hotho, R. Jäschke, and G. Stumme. Mining association rules in folksonomies. In V. Batagelj, H.-H. Bock, A. Ferligoj, and A. Iberta, editors, *Data Science and Classification: Proc. of the 10th IFCS Conf.*, Studies in Classification, Data Analysis, and Knowledge Organization, pages 261–270. Springer, 2006.

S. Staab, S. Santini, F. Nack, L. Steels, and A. Maedche. Emergent semantics. *Intelligent Systems* **17** (1):78–86, 2002.

L. Steels. The origins of ontologies and communication conventions in multi-agent systems. *Autonomous Agents and Multi-Agent Systems* **1** (2):169–194, October 1998.

R. Wille. Restructuring lattice theory: an approach based on hierarchies of concepts. In Ivan Rival, editor, *Ordered sets*, pages 445–470. Reidel, 1982.