
Kollaboratives Wissensmanagement

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

¹ Fachgebiet Wissensverarbeitung, FB 17, Universität Kassel
{schmitz,hotho,jaeschke,stumme}@cs.uni-kassel.de

² Forschungszentrum L3S, Universität Hannover

Zusammenfassung: Wissensmanagement in zentralisierten Wissensbasen erfordert einen hohen Aufwand für Erstellung und Wartung, und es entspricht nicht immer den Anforderungen der Benutzer.

Wir geben in diesem Kapitel einen Überblick über zwei aktuelle Ansätze, die durch kollaboratives Wissensmanagement diese Probleme lösen können. Im *Peer-to-Peer-Wissensmanagement* unterhalten Benutzer dezentrale Wissensbasen, die dann vernetzt werden können, um andere Benutzer eigene Inhalte nutzen zu lassen. *Folksonomies* versprechen, die Wissensakquisition so einfach wie möglich zu gestalten und so viele Benutzer in den Aufbau und die Pflege einer gemeinsamen Wissensbasis einzubeziehen.

1 Einführung

1.1 Probleme im Wissensmanagement

Dass das Entdecken, Bewahren und Nutzen von Wissen heute einer der wesentlichen Wettbewerbsfaktoren in Unternehmen ist, steht außer Frage. Technische Lösungen zum Wissensmanagement zielen bisher allerdings hauptsächlich darauf ab, eine zentrale Wissensbasis zu pflegen, in der Dokumente mehr oder weniger strukturiert abzulegen sind; siehe dazu z. B. die Fallbeispiele in [22].

Verschiedenartige Werkzeuge wurden zu diesem Zweck entwickelt; Tsui [31] benennt z. B. Expertensysteme, Intranets (Enterprise Information Portals, Enterprise Search Engines), Taxonomien, Information Repositories, Case Based Reasoning und Question-Answering-Systeme. Diese Hilfsmittel führen allerdings bisher nicht immer zum gewünschten Erfolg:

The majority of today's organizational knowledge still exists outside of organizational information repositories, and often only in people's heads. While organizations are eager to capture this knowledge formally, existing acquisition methods are not up to the task.

(Christian Wagner [32])

Die Bereitstellung von Wissen in Organisationen und die Wissensakquisition – das Überführen des vorhandenen Wissens in eine persistente, explizite Form – stellen also ein wesentliches Problem von zentralisierten Wissensmanagement-Projekten dar. Wagner [32] benennt eine Reihe von Problemen, die solchen Vorhaben im Wege stehen:

Bandbreite der Akquisition: Das Erstellen von formalisierten Wissensrepräsentationen ist zeitaufwendig. Es kann daher in der Regel nur ein geringer Teil des vorhandenen Wissens formalisiert werden.

Latenz der Akquisition: Aus denselben Gründen wird neues Wissen oft erst mit einer Verzögerung in zentrale Systeme eingepflegt.

Ungenauigkeiten: Sowohl menschliche *Knowledge Engineers* als auch Data-Mining-Techniken können Wissen nur mit einer gewissen Fehlerrate erfassen und repräsentieren.

Wartungsaufwand: Diese Fehler zu finden und zu korrigieren und neues, sich änderndes Wissen zu erfassen bedeutet einen ebenso großen Aufwand wie die Erfassung selbst.

Insbesondere in dynamischen Tätigkeitsfeldern mit schnell wechselnden Anforderungen ist es dabei für Mitarbeiter ein Problem, aktuelle und relevante Informationen zu finden. Von den Mitarbeitern in der täglichen Arbeit neu erworbenes Wissen müsste eigentlich zeitnah wieder zurück in die zentrale Wissensbasis fließen. Aus den oben genannten Gründen ist dies aber zu aufwendig und ohne direkten, für den Mitarbeiter unmittelbar ersichtlichen Nutzen. Daher wird neues Wissen oft nicht in zentrale Wissenbasen eingepflegt.

1.2 Vom persönlichen zum kollaborativen Wissensmanagement

Neben dem zentral organisierten Wissensmanagement mit dedizierten Werkzeugen organisieren sich daher Benutzer einen Großteil der benötigten Informationen selber: typischerweise wird ein *Knowledge Worker* mit verschiedenen Werkzeugen sein persönliches Wissen strukturieren. Zu diesen gehören z. B. E-Mail-Ordner, Dokumentensammlungen auf dem eigenen Rechner samt zugehöriger Ordnerstruktur und Desktop-Suchmaschine, PIM-Tools³, die Termine und Aufgaben strukturieren, oder auch Mindmaps [31]. Die Gründe für die Benutzung solcher persönlicher Werkzeuge anstelle zentraler Lösungen sind vielfältig [31]:

- Benutzer müssen in der Lage sein, “on the job” erlangtes Wissen aufzubewahren. Dazu muß der technische Aufwand für die Speicherung gering sein – geringer beispielsweise als Upload, Verschlagwortung usw. in einem zentralen System.

³ Personal Information Manager

- Die Anforderungen an den Benutzer sind unstrukturiert und ändern sich schnell, so dass starre Strukturen in einem zentralen System nicht zum Arbeitsalltag passen.
- Benutzer fordern Autonomie über Art der Speicherung und Verwendung von Informationen und benutzte Schlagworte und Taxonomien.

Auf der anderen Seite führt die freie Verwendung verschiedener Wissensmanagementtechniken und -werkzeuge zu Problemen, wenn vorhandenes Wissen gefunden, wiederverwertet und integriert werden soll:

- Verschiedene Werkzeuge sind in der Regel nicht integriert. Selbst einfache Zusammenhänge, z. B. die inhaltliche Verknüpfung zwischen einer EMail, dem darin vereinbarten Termin im Terminkalender, und dem als Anhang mitgesendeten Dokument, das im Dateisystem des Benutzers abgelegt wird, gehen verloren.
- Jedes der Werkzeuge benutzt ein anderes Datenmodell, bietet andere Suchmöglichkeiten, und präsentiert die Informationen auf andere Art und Weise.
- Dadurch, dass jeder Benutzer seine eigene Wissensbasis pflegt, wird Redundanz erzeugt: dieselbe Information wird von mehreren Personen modelliert und in vielen Wissensbasen abgelegt, ohne dass Synergieeffekte ausgenutzt werden.

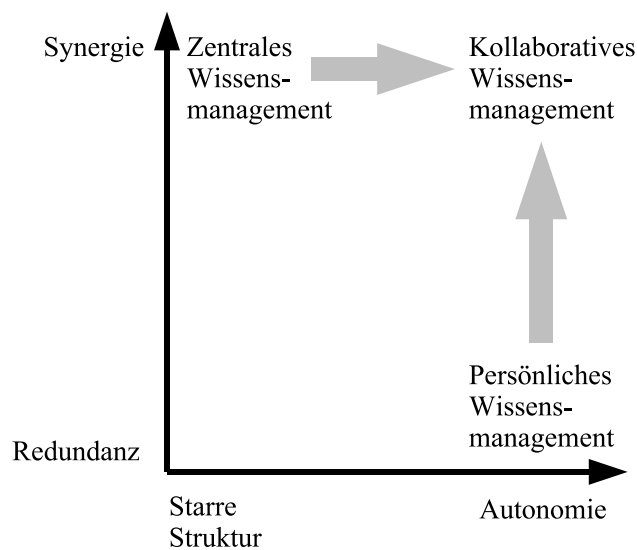


Abbildung 1. Vom persönlichen zum kollaborativen Wissensmanagement

Abbildung 1 stellt zwei Dimensionen dar, in denen sich die bisher genannten Lösungen bewegen. Auf der einen Seite gibt es zentral organisier-

te Lösungen, die zwar Redundanz vermeiden und Wiederverwertung fördern sollen, aber starr, kostenintensiv und aufwendig zu warten sind. Andererseits betreiben die Benutzer in der täglichen Praxis eigene Lösungen, die flexibel und mit wenig Zusatzaufwand Wissen verwalten, aber durch die mangelnde Vernetzung Redundanz und Mehrfacharbeit mit sich bringen.

Im Folgenden werden wir zwei Ansätze zum *kollaborativen Wissensmanagement* diskutieren, die derzeit vorangetrieben werden, um einerseits den Benutzern zu ermöglichen, schnell, autonom und mit geringem Mehraufwand ihr eigenes Wissen zu verwalten, und trotzdem Synergieeffekte durch Mitbenutzen der Inhalte anderer zu erzielen.

Abschnitt 2 stellt *Peer-to-Peer-Wissensmanagement* vor, bei dem die persönlichen Wissensbasen Einzelner in einem dezentralen Netzwerk zur Verfügung gestellt werden. In Abschnitt 3 werden *Folksonomies* eingeführt, die darauf abzielen, das Erstellen, Warten und Bereitstellen persönlicher Informationen mit möglichst geringem Aufwand, ohne lange Einarbeitung und nahtlos in die täglichen Abläufe integriert zu ermöglichen.

2 P2P-Wissensmanagement

Ein Lösungsansatz, um beide zuvor genannten Vorteile – Autonomie des Benutzers und geringer Aufwand einerseits, Synergieeffekte andererseits – zu realisieren, ist *Peer-to-Peer-Wissensmanagement* (P2P Knowledge Management; P2PKM).

In einem P2PKM-System pflegt zunächst jeder Benutzer seine eigene Wissensbasis. Darüber hinaus kann die Wissensbasis anderen Benutzern zugänglich gemacht werden, so dass Informationen nicht mehrfach modelliert werden müssen, sondern man sich die Wissensbasis anderer zu Nutze machen kann.

Das Ziel ist es dabei, den Datenaustausch völlig dezentral – also ohne zentralen Koordinator oder Server – zu gestalten. Dadurch soll ein skalierbares, robustes Netzwerk von Wissensbasen entstehen, das jedem Benutzer einen globalen Zugriff auf das Wissen ermöglicht.

2.1 Definition/Einführung

Ein P2PKM-System ist gekennzeichnet durch folgende Eigenschaften [21]:

- Jeder Benutzer wird durch einen *Peer* repräsentiert, d. h. ein Programm, das auf seinem Rechner läuft und in einem P2P-Netzwerk für den Benutzer Informationen anbietet, entgegennimmt und weiterleitet.
- Der Peer greift auf eine *Wissensbasis* zu. In dieser werden die vom Benutzer zur Verfügung gestellten oder aus dem Netz gesammelten Informationen gespeichert.
- Die Wissensbasis integriert verschiedene Quellen in ein gemeinsames *Datenmodell* und lässt sich mit Hilfe einer *Anfragesprache* lokal oder von anderen Peers anfragen.

- *Mediatoren (Wrapper)* oder *Konverter* dienen dazu, aus Quellen wie Mailprogrammen, Dateisystemen, Volltextdokumenten oder Datenbanken eine in der Wissensbasis integrierte Sicht der Informationen eines Benutzers zu erstellen.
- Es existiert ein Satz von *Protokollen*, die die verschiedenen Aspekte des P2PKM-Netzes, wie z. B. Routing von Nachrichten, Zugangskontrolle und Beurteilung von Reputation, realisieren.

2.2 Bestehende Systeme

In den vergangenen Jahren wurden eine Reihe von derartigen P2PKM-Projekten durchgeführt, die verschiedene der oben genannten Teilaspekte bearbeiteten.

In *Edutella* [21] wurde eine P2PKM-Infrastruktur für den ELearning-Bereich zur Verfügung gestellt. Besonderes Augenmerk wurde dabei auf eine Integration verschiedenster Datenquellen gelegt, die in einem logischen Datenmodell basierend auf Datalog zusammengefasst werden. Von den unterliegenden Anfragesprachen und ihren Möglichkeiten wird abstrahiert. Dadurch können insbesondere bereits bestehende Systeme leichter integriert werden. Auf Basis von Edutella wurden prototypisch Endbenutzer-Applikationen realisiert, wie z. B. der *Courseware Watchdog* [30], die die Organisation von ELearning-Ressourcen auf dem Desktop des Endbenutzers sowie die Vernetzung über Edutella ermöglichen.

Ähnliche Ziele wurden mit *Bibster* verfolgt. Bibster [12] erlaubt es, ein typisches Problem von Wissenschaftlern – die Literaturrecherche und Verwaltung von Literaturreferenzen – mit einem P2PKM-Netzwerk zu erleichtern. Dabei können $\text{BIB}_{\text{T}_{\text{E}}\text{X}}$ -Einträge in einer Wissensbasis verwaltet, mit weiteren Metadaten (z. B. thematischen Kategorien, was in $\text{BIB}_{\text{T}_{\text{E}}\text{X}}$ nicht vorgesehen ist) versehen und über das P2P-Netz getauscht werden. Etliche Erweiterungen zu Bibster, wie z.B. das Empfehlen von unbekannter, relevanter Literatur mittels neuer $\text{BIB}_{\text{T}_{\text{E}}\text{X}}$ -Einträgen aus dem P2P-Netz [9] oder die Unterstützung von personalisierten Ontologiestrukturen zur Organisation der Wissensbasis [10] wurden entwickelt.

Etliche weitere Projekte, wie z. B. *Edamok*⁴, *InfoQuilt*⁵ und *BookQuest*⁶, haben vergleichbare Ziele verfolgt.

2.3 Probleme und Lösungsansätze

Um ein P2PKM-System in der genannten Art realisieren zu können, ist eine Reihe von Problemen zu lösen, von denen viele im zentral organisierten Wissensmanagement keine Rolle spielen.

⁴ <http://edamok.itc.it/>

⁵ <http://lsdis.cs.uga.edu/projects/past/InfoQuilt/>

⁶ <http://mybook.uc.edu/>

Routing

Kann eine Anfrage eines Benutzers nicht von der lokalen Wissensbasis beantwortet werden, so müssen andere Peers gefunden werden, die diese Anfrage beantworten können. Dieses Problem wird als *Routing* von Anfragen bezeichnet.

Da triviale Lösungen – etwa das Fluten des gesamten Netzwerkes mit jeder Anfrage – nicht skalieren, wurden im P2PKM-Zusammenhang zweierlei Strategien verfolgt:

Verteilte Hashtabellen

Ein weit verbreiteter Ansatz ist die Nutzung sogenannter *verteilter Hashtabellen* (Distributed Hash Tables, DHT). Dabei wird eine Datenstruktur zwischen den Peers aufgebaut, die es erlaubt, effizient den Peer zu bestimmen, der ein bestimmtes Datum enthält [3, 28, 23]. Um DHTs in einem P2PKM-Szenario anzuwenden, bedarf es allerdings Anpassungen. Zum einen organisieren DHTs die Daten anhand binärer Schlüssel, was semantisch modellierte Informationen oft nur unzureichend wiedergeben kann. Zum anderen wird die Autonomie der Benutzer in Frage gestellt, da jedem Peer vom System Inhalte zugewiesen werden, die dieser bereitstellen muss, damit die DHT funktionieren kann.

Semantische Topologien

Semantische Topologien basieren auf der Idee, die Semantik der auf den Peers befindlichen Informationen zur Bildung geeigneter Netzwerktopologien und zum Routing zu nutzen. Hierbei wird das P2P-Netz analog zu thematischen Zusammenhängen strukturiert, d. h. Peers, die sich mit ähnlichen Themen beschäftigen, liegen im Netzwerk nahe beieinander. Dadurch können Anfragen mit Hilfe von sozialen Metaphern, die das Verhalten von Menschen beim Suchen nach Informationen nachbilden, zu Erfolg versprechenden Peers geleitet werden. Dies können z. B. Peers sein, die sich mit einem zur Anfrage verwandten Thema befassen, oder die in der Vergangenheit bereits gute Antworten geliefert haben.

Verschiedene Arbeiten [24, 16, 11] haben gezeigt, dass man auf diese Weise das P2P-Netz thematisch clustern und somit effiziente Anfragen ermöglichen kann. Zudem erlaubt es eine solche thematische Gruppierung, weitere Dienste anzubieten, etwa die Unterstützung bei der Bildung von Communities oder eine Browsing-Funktionalität, die die Suche durch Anfragen ergänzt.

Mediation

Sollen verschiedene Benutzer in einem großen verteilten System mit Hilfe von Semantic-Web-Technologie Wissen austauschen, ergibt sich das Problem, dass sie im Allgemeinen ihr Wissen auf verschiedene Art modellieren werden, d. h. verschiedene Ontologien benutzen. Um dennoch Nutzen aus den Wissensbasen

Anderer ziehen zu können, muss zwischen verschiedenen Ontologien vermittelt werden.

Lösungsansätze in diesem Bereich zielen in der Regel darauf ab, lokale Abbildungen, die paarweise zwischen Peers hergestellt werden, zu verketteten und somit eine Kommunikation auch zwischen solchen Peers zu ermöglichen, die kein unmittelbares Ontologiemapping zur Verfügung haben [8, 2].

Zugangskontrolle

Auch wenn Benutzer in einem P2PKM-Szenario weit gehend gewillt sein werden, ihre Inhalte an andere Peers weiterzugeben, so gibt es doch Szenarien, in denen der Zugang zur eigenen Wissensbasis mit Zugangsbeschränkungen versehen werden muss.

Die semantische Repräsentation der bekannten Abstraktionen wie Rollen, Berechtigungen, Zertifikate usw. aus dem Bereich Public-Key-Infrastrukturen [20] erlaubt eine regelbasierte Spezifikation von Zugangskontrollen, die mit Hilfe von logischen Schlüssen die Schutzanforderungen des Peers durchsetzen kann, ohne dass ein Benutzer im Vorhinein jeden möglichen Zugriff auf seine Wissensbasis regeln müsste.

Vertrauen und Reputation

In einem offenen P2PKM-System werden sich im Allgemeinen nicht alle Peers jederzeit gutartig verhalten. Manche Peers werden möglicherweise keine oder falsche Antworten geben, oder aktiv versuchen, durch ihr Verhalten andere zu schädigen.

Um solches Verhalten erkennen und die betreffenden Peers meiden zu können, ist es zum einen notwendig, dass Peers ihr Vertrauen (Trust) in andere Peers – z. B. anhand vergangener Transaktionen – dokumentieren. Zum anderen müssen die so erhaltenen Werte effizient verfügbar gemacht und aggregiert werden, um so die Reputation eines jeden Peers bestimmen zu können.

Trust-Bewertungen werden in der Regel in DHTs vorgehalten. Damit sind sie effizient auffindbar und werden nicht vom betreffenden Peer selber gespeichert, der ein Interesse daran hätte, sie zu fälschen. Zur Aggregation wird die Idee eines *web of trust* aufgenommen, in dem sich Vertrauen transitiv fortsetzt. So kann Vertrauen (oder analog Misstrauen) zu einem unbekanntem Peer aufgebaut werden, indem eine vertrauenswürdige Kette von Peers gefunden wird, die dessen Reputation bestätigen [1, 14, 27].

3 Folksonomies

In diesem Kapitel gehen wir auf einen weiteren Ansatz des kollaborativen Wissensmanagements, die sogenannten Folksonomies, ein. Nach einer kurzen

Einführung in Abschnitt 3.1 diskutieren wir in Abschnitt 3.2 bestehende Folksonomy-Systeme. In Abschnitt 3.3 führen wir unser formales Folksonomy-Modell ein und zeigen erste Erweiterungen entlang zweier Beispiele in Abschnitt 3.4.

3.1 Einführung

Soziale Resource-Sharing-Systeme, also Systeme zum gemeinsamen Verschlagworten von Bookmarks, Bildern, Musikstücken oder $\text{BIB}_{\text{T}}\text{E}_\text{X}$ -Einträgen, sind webbasierte Systeme, die es dem Benutzer erlauben, seine Ressourcen auf einem Server zu speichern und sie dort mit einem *Tag* (Schlagwort) zu versehen. Der Vorgang des Taggens, also des Kategorisierens von Ressourcen mit frei wählbaren Schlagworten, führt zu einem Beziehungsnetzwerk aus Tags, Benutzern und Ressourcen, das man auch als *Folksonomy* bezeichnet. Nach Thomas Vander Wal⁷ ist Folksonomy ein Kofferwort, das die Worte “folk” und “taxonomy” kombiniert (siehe [18]) und die Bottom-up-Entwicklung einer Taxonomie zum Ausdruck bringt. Benutzer eines Folksonomy-Systems werden angehalten, Ressourcen mit frei wählbaren Schlagworten/Tags zu versehen, wodurch man ein breites Vokabular erhält und dem Benutzer viel Freiheit erlaubt. Das dahinterliegende gemeinsame Taggen/Klassifizieren von Ressourcen ist eine Art Wissensakquisition, die durch die Nutzung von populären Tags belohnt werden kann, indem Ressource und Benutzer auf der Seite der populären Einträge angezeigt werden, und so zur Profilierung des Wissensgebers beitragen. Trotzdem kann der Benutzer unabhängig von anderen Benutzern seine Ressourcen verwalten und klassifizieren.

Es gibt verschiedene Systeme, die man nach der Art der unterstützten Ressource unterscheidet. Während Flickr⁸ die Nutzung von Bildern unterstützt, zielt del.icio.us⁹ auf Bookmarks, CiteULike¹⁰ und Connotea¹¹ auf bibliographische Einträge, last.fm¹² auf Musik und 43Things¹³ auf gute Vorsätze ab. Das von uns entwickelte System *BibSonomy*¹⁴ unterstützt sowohl Bookmarks als auch $\text{BIB}_{\text{T}}\text{E}_\text{X}$ -Einträge (siehe Abb. 2). Vander Wal unterscheidet weiterhin *umfassende* und *schmale* Folksonomies.¹⁵ Während bei umfassenden Folksonomies wie del.icio.us viele Benutzer eine Ressource klassifizieren, wird bei einer schmalen Folksonomy wie Flickr jede Ressource nur von einem Benutzer getaggt.

Im Kern sind sich die verschiedenen Systeme sehr ähnlich. Wenn ein Benutzer angemeldet ist, kann er Ressourcen zum System hinzufügen und diesen beliebige Tags zuordnen. Wir nennen die Sammlung aller Ressourcen eines Benutzers und die jeweils zugeordneten Tags seine *Personomy* und die Sammlung aller Personomies *Folksonomy*. Ein Benutzer kann die Personomies aller anderen Benutzer entlang aller Dimensionen des Systems einsehen: So kann

⁷ http://atomiq.org/archives/2004/08/folksonomy_social_classification.html

⁸ <http://www.flickr.com/> ⁹ <http://del.icio.us/> ¹⁰ <http://www.citeulike.org/>

¹¹ <http://www.connotea.org/> ¹² <http://www.last.fm/> ¹³ <http://www.43things.com/>

¹⁴ <http://www.bibsonomy.org>¹⁵ <http://www.iawiki.net/FolksOnomy>



Abbildung 2. Bibsonomy zeigt gleichzeitig Bookmarks und (BIB_TE_X-basierte) bibliographische Referenzen an.

man für einen gegebenen Benutzer alle ins System eingestellten Ressourcen sowie die vom Benutzer zugewiesenen Tags einsehen (siehe Abb. 2). Klickt man auf eine Ressource, so sieht man alle anderen Benutzer des Systems, die diese Ressource ebenfalls ins System eingestellt haben, sowie deren Tags. Klickt man auf ein Tag, so erhält man alle Ressourcen, die ein solches Tag bekommen haben.

Die meisten Systeme bieten weitere Funktionen an, wie z.B. das Kopieren von vorhandenen Ressourcen anderer Benutzer und das Taggen mit eigenen Schlagworten. Im allgemeinen stellen solche Systeme eine sehr intuitive Navigation durch die Daten zur Verfügung.

3.2 State of the Art

Zur Zeit gibt es praktisch keine wissenschaftlichen Publikationen über Folksonomy-Systeme. Unter den seltenen Ausnahmen findet man die Arbeiten [13] und [17], die einen guten Überblick über soziale Bookmarksysteme geben sowie [18], die Stärken und Schwächen von Folksonomies diskutiert. Die Diskussion über Folksonomies und verwandte Themen findet zur Zeit hauptsächlich auf Mailinglisten statt, z.B. [5]. Kürzlich veröffentlicht wurde ein Ansatz zur Extraktion von leichtgewichtigen Ontologien aus Folksonomies [19], der Co-Occurrence-Techniken zum Clustern des Begriffsnetzes verwendet, und an einem Ausschnitt von del.icio.us demonstriert wurde.

Weiterhin existieren einige Systeme, die auf del.icio.us aufbauend die darunterliegende Folksonomy untersuchen. CollaborativeRank¹⁶ stellt sortierte Suchergebnisse auf Basis der del.icio.us-Bookmarks zur Verfügung. Die Sortierung der Suchergebnisse ergibt sich aus der Popularität der Bookmarks und dem Zeitpunkt, an dem man ein Bookmark getaggt hat. Andere System zeigen populäre Seiten (Populicious¹⁷) oder fokussieren auf eine bessere graphische Präsentation (Cloudalicious¹⁸, Grafolicious¹⁹) der Statistiken über del.icio.us.

¹⁶ <http://collabrank.org/> ¹⁷ <http://populicio.us/> ¹⁸ <http://cloudalicio.us/>

¹⁹ <http://www.neuroticweb.com/recursos/del.icio.us-graphs/>

Confoto,²⁰ der Gewinner der Semantic Web Challenge 2005, erlaubt das Annotieren und Browsen von Konferenzfotos und bietet neben einer reichen Semantik auch Taggingdienste zur Annotation an. Damit geht dieses System über die bisher diskutierten Social-Resource-Sharing-Systeme hinaus, indem es Semantic-Web-Techniken und Tagging-Ideen miteinander verbindet. Ungeklärt ist, ob das System auf große Datenmengen und viele Benutzer skalieren kann.

3.3 Formales Folksonomy-Modell

User (Benutzer), Ressourcen (Webseiten, BIB_TEX -Einträge o.ä.) und Tags (Schlagworte) bilden die Basis einer Folksonomy, wobei das Folksonomy-System den Benutzern die Möglichkeit bietet, frei wählbare Tags mit einer Ressource zu verknüpfen. Im folgenden präsentieren wir die formale Definition einer Folksonomy, wie sie auch dem BibSonomy-System zugrunde liegt.

Definition 1. Eine Folksonomy ist ein Tupel $\mathbb{F} := (U, T, R, Y, \prec)$ wobei

- U, T und R endliche Mengen sind, deren Elemente man Users, Tags und Ressourcen nennt,
- Y eine ternäre Beziehung zwischen diesen ist, d.h. $Y \subseteq U \times T \times R$ gilt, und
- \prec eine userspezifische Untertag/Obertag-Beziehung ist, d.h. $\prec \subseteq U \times ((T \times T) \setminus \{(t, t) \mid t \in T\})$ gilt.

Die Personomy \mathbb{P}_u eines Users $u \in U$ ist die Einschränkung \mathbb{F} auf u , d.h. $\mathbb{P}_u := (T_u, R_u, I_u, \prec_u)$ mit $I_u := \{(t, r) \in T \times R \mid (u, t, r) \in Y\}$, $T_u := \pi_1(I_u)$, $R_u := \pi_2(I_u)$ und $\prec_u := \{(t_1, t_2) \in T \times T \mid (u, t_1, t_2) \in \prec\}$.

Benutzer werden typischerweise durch eine User-ID und Tags durch beliebige Strings beschrieben. Was man als Ressourcen betrachtet, hängt vom Systemtyp ab. Ressourcen sind in del.icio.us zum Beispiel Bookmarks und in Flickr Bilder. Unser System BibSonomy enthält zwei Arten von Ressourcen, Bookmarks und BIB_TEX -Einträge. Verschiedene Typen von Ressourcen unterscheiden sich strukturell nicht, da sie unabhängig vom Typ intern durch eine ID repräsentiert werden. Nur die Art der Bildschirmanzeige hängt von der Art der Ressource ab.

Wenn man die Untertag/Obertag-Beziehung beiseite lässt (d.h. $\prec := \emptyset$ setzt), kann man eine Folksonomy vereinfacht als Quadrupel $\mathbb{F} := (U, T, R, Y)$ beschreiben. Diese Struktur entspricht einem *triadischen Kontext* [15, 29] in der Formalen Begriffsanalyse [33, 7]. Eine äquivalente Sicht auf Folksonomy-Daten ist ein tripartiter (ungerichteter) Hypergraph $G = (V, E)$, wobei $V := U \dot{\cup} T \dot{\cup} R$ die Menge der Knoten und $E := \{\{u, t, r\} \mid (u, t, r) \in Y\}$ die Menge der Hyperkanten ist.

²⁰ <http://www.confoto.org/>

3.4 Erweiterung bestehender Folksonomy-Tools

Mit der Weiterentwicklung von Folksonomy-Tools auf Basis aktueller Forschungsergebnisse beschäftigen sich die folgenden zwei Abschnitte.

Emergent Semantics in Folksonomies

Die Diskussionen auf einschlägigen Mailinglisten, etwa *delicious-discuss*²¹, zeigen, dass neben einfachen Tags Bedarf nach mehr Struktur in Folksonomies besteht, z. B. in Form der oben genannten \prec -Relation. Eine Möglichkeit, den Benutzern solche Strukturen zur Verfügung zu stellen, ohne dass sie sie vollständig selber pflegen müssen, ist die Anwendung von Techniken des Ontologie-Lernens.

Durch ihr Verhalten beim Taggen von Ressourcen drücken Benutzer die Bedeutung der Ressourcen aus. Dabei werden oft allgemeinere und speziellere Tags vermischt – so sind etwa viele Webseiten über XSLT-Stylesheets mit dem Tag *xslt* und zusätzlich mit dem Oberbegriff *xml* versehen. Durch das Berechnen von Assoziationsregeln können aus der Folksonomy solche Beziehungen zwischen Tags extrahiert werden. Da Folksonomies inhärent dreidimensional sind, müssen sie zuerst auf zwei Dimensionen projiziert werden, damit man die klassischen Assoziationsregel-Verfahren anwenden kann [26].

Wir haben diesen Ansatz auf einem der bekanntesten Folksonomy-Systeme, *del.icio.us*, analysiert. Zu diesem Zweck haben wir $|U| = 75.242$ User, $|T| = 533.191$ Tags und $|R| = 3.158.297$ Ressourcen betrachtet, die in *del.icio.us* über insgesamt $|Y| = 17.362.212$ Tripel miteinander verbunden sind.

Abbildung 3 zeigt ein Beispiel aus [26], welches auf den *del.icio.us*-Daten berechnet wurde. Hierbei wird untersucht, welche Tags bevorzugt im Zusammenhang mit anderen auftreten (also etwa: wenn von einem Benutzer für eine bestimmte Ressource *xslt* vergeben wird, dann oft auch *xml*). In der Assoziationsregel-Nomenklatur sind also Tags die Produkte und User-Ressource-Kombinationen die Warenkörbe. In unserem konkreten Fall deutet eine solche Regel auf eine Unterbegriff-Overbegriff-Beziehung hin, die dem Benutzer zur Aufnahme in seine \prec -Relation vorgeschlagen werden kann. Dieser Ansatz kann mit Volltext-basierten Methoden (z. B. [4]) kombiniert werden, wenn die Ressourcen Webseiten oder andere Textdokumente sind.

Entdecken von Communities

Eine weitere Aufgabe des Wissensmanagements ist das Entdecken sogenannter *Communities of Interest* in Folksonomies. Damit sind zunächst Gruppen von Benutzern gemeint, die gemeinsame Interessen zeigen. Zur Beschreibung der Communities können zusätzlich die wichtigsten Tags und Ressourcen herangezogen werden.

²¹ <http://lists.del.icio.us/pipermail/discuss/>

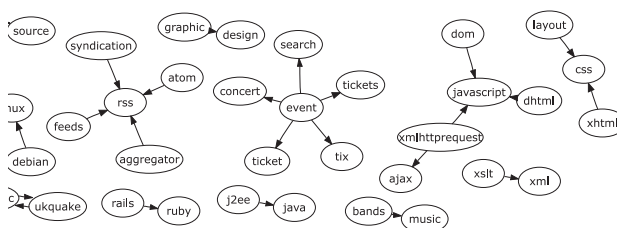


Abbildung 3. Zweielementige Regeln zwischen den del.icio.us-Tags mit 0,05 % Support und 50 % Konfidenz

Konkret können Communities z. B. mit dem Algorithmus *FolkRank* [25] bestimmt werden. Diese Weiterentwicklung des PageRank-Algorithmus für Folksonomies erlaubt es, in Abhängigkeit von vorgegebenen Benutzerpräferenzen ein Ranking über Tags, Benutzer und Ressourcen zu erstellen, so dass verwandte Elemente der Folksonomy ein hohes Gewicht erhalten. Erreicht wird dies durch eine differentielle Betrachtung des größten Eigenvektors eines Graphen, der aus der Folksonomy gewonnen wird.

Tabelle 1 zeigt z. B. die Community, die sich in del.icio.us rund um <http://www.semanticweb.org/> bildet. Durch das Explizitmachen solcher Communities können Benutzer schneller relevantes Material finden, Kontakte zu anderen Benutzern knüpfen und das Vokabular der Community kennenlernen.

4 Zusammenfassung und Ausblick

P2P Knowledge Management und Folksonomies treten an, um durch *kollaboratives Wissensmanagement* die Probleme des zentral organisierten Top-Down-Wissensmanagements zu beheben, die in der starren Struktur und der aufwendigen Pflege begründet sind. Beide tragen dazu bei, dass Benutzer einfacher, praxisnäher und mit weniger Aufwand Informationen sammeln, ihr Wissen strukturieren und es mit Anderen teilen können.

Eine Integration der beiden Ideen mit den vorhandenen Desktop-Tools wäre wünschenswert. Tags an einen Termin im Kalender zu schreiben, so dass automatisch das passende Dokument als Anhang vorgeschlagen werden kann, und diesen Termin dann auch noch mit Projektpartnern im P2P-Netz teilen zu können – solche Möglichkeiten würden das Leben vieler Knowledge Worker erheblich erleichtern. Diese Vision eines *Social Semantic Desktop* [6] ist derzeit eine der meistdiskutierten Forschungsrichtungen im Semantic-Web-Bereich und wird sich sicherlich Ideen aus Folksonomies und P2PKM zu Eigen machen, um dieses Ziel zu verwirklichen.

Tabelle 1. Die wichtigsten User, Tags und Ressourcen rund um <http://www.semanticweb.org/>

User	Tags	Ressourcen
up4	semanticweb	http://www.semanticweb.org/
awenger	semantic	http://flink.semanticweb.org/
j.deville	web	http://simile.mit.edu/piggy-bank/
chaizzilla	semantic_web	http://www.w3.org/2001/sw/
elektron	rdf	http://infomesh.net/2001/swintro/
captsolo	semweb	http://del.icio.us/register
dissipative	resources	http://mspace.ecs.soton.ac.uk/
stevag	community	http://www.adaptivepath.com/pu...
krudd	portal	http://www.ontoweb.org/
williamteo	xml	http://www.aaai.org/AITopics/html/ontol.html
stevecassidy	research	http://simile.mit.edu/
pmika	imported-bo...	http://itip.evcc.jp/itipwiki/
millette	en	http://www.google.be/
myren	.idate2005-04-11	http://www.letterjames.de/index.html
morningboat	newfurl	http://www.daml.org/
philip.fennell	tosort	http://shirky.com/writings/ontology_overrated.html
webb.	cs	http://jena.sourceforge.net/
dnaboy76	academe	http://www.alistapart.com/
mote	rfid	http://www.federalconcierge.com/...
alphajuliet	sem-web	http://pchere.blogspot.com/2005/02/...
nymetbarton	w3c	http://www.shirky.com/writings/...

Literatur

1. K. Aberer und Z. Despotovic. Managing trust in a peer-2- peer information system. In *Proc. 10th International Conference on Information and Knowledge Management (2001 ACM CIKM)*. Atlanta, 2001.
2. K. Aberer und M. Hauswirth. Semantic gossiping. In *Database and Information Systems Research for Semantic Web and Enterprises, Invitational Workshop*. University of Georgia, Amicalola Falls and State Park, Georgia, Apr. 2002.
3. K. Aberer, M. Puceva, M. Hauswirth und R. Schmidt. Improving data access in P2P systems. *IEEE Internet Computing Journal*, 6(1):58–67, 2002.
4. P. Cimiano, A. Pivk, L. Schmidt-Thieme und S. Staab. Learning taxonomic relations from heterogeneous sources of evidence. In *Ontology Learning from Text: Methods, Evaluation and Applications*, Frontiers in Artificial Intelligence, S. 59–73. IOS Press, 2005.
5. Connotea. Connotea Mailing List, 2005. <https://lists.sourceforge.net/lists/listinfo/connotea-discuss>.
6. S. Decker und M. R. Frank. The networked semantic desktop. In *Proc. WWW Workshop on Application Design, Development and Implementation Issues in the Semantic Web*. Net York City, NY, Mai 2004.
7. B. Ganter und R. Wille. *Formal Concept Analysis : Mathematical foundations*. Springer, 1999.

8. F. Giunchiglia und I. Zaihrayeu. Coordinating mobile databases. In *Proc. MobiQuitous Workshop on Peer-to-Peer Knowledge Management (P2PKM 2004)*. Boston, MA, USA, Aug. 2004.
9. P. Haase, M. Ehrig, A. Hotho und B. Schnizler. Personalized information access in a bibliographic peer-to-peer system. In S. Staab und H. Stuckenschmidt (Hg.), *Peer-to-Peer and SemanticWeb, Decentralized Management and Exchange of Knowledge and Information*, S. 143–158. Springer, 2006.
10. P. Haase, A. Hotho, L. Schmidt-Thieme und Y. Sure. Collaborative and usage-driven evolution of personal ontologies. In A. Gómez-Pérez und J. Euzenat (Hg.), *ESWC*, Bd. 3532 von *Lecture Notes in Computer Science*, S. 486–499. Springer, 2005.
11. P. Haase, R. Siebes und F. van Harmelen. Peer selection in peer-to-peer networks with semantic topologies. In *International Conference on Semantics of a Networked World: Semantics for Grid Databases*. Paris, 2004.
12. P. Haase, N. Stojanovic, Y. Sure und J. Völker. Personalized information retrieval in Bibster, a semantics-based bibliographic peer-to-peer system. In *Proceedings of the 5th International Conference on Knowledge Management (I-KNOW 05)*. JUL 2005.
13. T. Hammond, T. Hannay, B. Lund und J. Scott. Social Bookmarking Tools (I): A General Review. *D-Lib Magazine*, 11(4), April 2005.
14. S. D. Kamvar, M. T. Schlosser und H. Garcia-Molina. The eigentrust algorithm for reputation management in p2p networks. In *Proc. 12th International World Wide Web Conference*. Budapest, Hungary, Mai 2003.
15. F. Lehmann und R. Wille. A triadic approach to formal concept analysis. In G. Ellis, R. Levinson, W. Rich und J. F. Sowa (Hg.), *Conceptual Structures: Applications, Implementation and Theory*, Bd. 954 von *Lecture Notes in Computer Science*. Springer, 1995.
16. A. Löser, C. Tempich, B. Quilitz, S. Staab, W. T. Balke und W. Nejdl. Searching dynamic communities with personal indexes. In Y. Gil, E. Motta, V. R. Benjamins und M. A. Musen (Hg.), *Proc. 4th International Semantic Web Conference, ISWC 2005*, Bd. 3729 von *LNCS*, S. 491 – 505. Springer-Verlag GmbH, Galway, Ireland, Nov. 2005.
17. B. Lund, T. Hammond, M. Flack und T. Hannay. Social Bookmarking Tools (II): A Case Study - Connotea. *D-Lib Magazine*, 11(4), April 2005.
18. A. Mathes. Folksonomies – Cooperative Classification and Communication Through Shared Metadata, December 2004. <http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html>.
19. P. Mika. Ontologies Are Us: A Unified Model of Social Networks and Semantics. In Y. Gil, E. Motta, V. R. Benjamins und M. A. Musen (Hg.), *ISWC 2005*, Bd. 3729 von *LNCS*, S. 522–536. Springer-Verlag, Berlin Heidelberg, November 2005.
20. W. Nejdl, D. Olmedilla, M. Winslett und C. C. Zhang. Ontology-based policy specification and management. In *2nd European Semantic Web Conference (ESWC)*, Bd. 3532 von *Lecture Notes in Computer Science*, S. 290–302. Springer, Heraklion, Crete, Greece, Mai 2005.
21. W. Nejdl, B. Wolf, C. Qu, S. Decker, M. Sintek, A. Naeve, M. Nilsson, M. Palmér und T. Risch. EDUTELLA: A P2P networking infrastructure based on RDF. In *Proc. 11th International World Wide Web Conference (WWW 2002)*. Honolulu, Mai 2002.
22. G. Probst, S. Raub und K. Romhardt. *Wissen managen: Wie Unternehmen ihre wertvollste Ressource optimal nutzen*. Gabler, Wiesbaden, 4. Aufl., 2003.

23. A. Rowstron und P. Druschel. Pastry: scalable, decentraized object location and routing for large-scale peer-to-peer systems. In *Proc. 18th IFIP/ACM International Conference on Distributed Systems Platforms (Middleware)*. Nov. 2001.
24. C. Schmitz. Self-organization of a small world by topic. In *Proc. 1st International Workshop on Peer-to-Peer Knowledge Management*. Boston, MA, Aug. 2004.
25. C. Schmitz, A. Hotho, R. Jäschke und G. Stumme. Information retrieval in folksonomies: Search and ranking (*eingereicht*). In *Proc. 3rd European Semantic Web Conference*. Budva, Montenegro, Juni 2006.
26. C. Schmitz, A. Hotho, R. Jäschke und G. Stumme. Mining association rules in folksonomies. In *Proc. IFCS 2006 Conference Data Science and Classification (eingereicht)*. Ljubljana, Juli 2006.
27. S. Song, K. Hwang, R. Zhou und Y.-K. Kwok. Trusted p2p transactions with fuzzy reputation aggregation. *IEEE Internet Computing*, 9(6):24–34, 2005.
28. I. Stoica, R. Morris, D. Karger, M. F. Kaashoek und H. Balakrishnan. Chord: A scalable peer-to-peer lookup service for internet applications. In *Proc. 2001 conference on applications, technologies, architectures, and protocols for computer communications (ACM SIGCOMM)*. ACM Press New York, NY, USA, 2001.
29. G. Stumme. A finite state model for on-line analytical processing in triadic contexts. In B. Ganter und R. Godin (Hg.), *ICFCA*, Bd. 3403 von *Lecture Notes in Computer Science*, S. 315–328. Springer, 2005.
30. J. Tane, C. Schmitz und G. Stumme. Semantic resource management for the web: An elearning application. In *Proc. 13th International World Wide Web Conference (WWW 2004)*. New York, Mai 2004.
31. E. Tsui. Technologies for personal and peer-to-peer knowledge management. Techn. Ber., CSC Leading Edge Forum, Juli 2002.
32. C. Wagner. Collaborative knowledge management: Breaking the knowledge acquisition bottleneck. In *Proc. Information Resources Management Association International Conference*. New Orleans, Mai 2004.
33. R. Wille. Restructuring lattices theory : An approach based on hierarchies of concepts. In I. Rival (Hg.), *Ordered Sets*, S. 445–470. Reidel, Dordrecht-Boston, 1982.