

1.1 Grundlagen des Social Bookmarking

Andreas Hotho

Eine neue Art von Web-Anwendungen, die auf die Beteiligung der Nutzer und auf einfache Benutzerschnittstellen baut, bildet einen Kernbestandteil des entstehenden Web 2.0. Dabei spielen benutzerbezogene Plattformen wie z.B. Wikis, Blogs und kooperative Verschlagwortungssysteme zum Teilen von Bookmarks oder URLs aber auch anderen Ressourcen eine immer zentralere Rolle. Solche Systeme dienen dem Verwalten, Veröffentlichen und Teilen von Informationen und werden demzufolge auch als einfache Wissensmanagementsysteme angesehen. Sie haben leicht verständliche Benutzerschnittstellen und es bedarf keiner spezifischen Fähigkeiten, um aktiv mit einem solchen System arbeiten zu können. Der direkte Vorteil für jeden Benutzer, browser- und rechnerunabhängig seine Daten und Informationen ohne großen Mehraufwand zentral organisieren zu können, zieht weitere Benutzer an und macht die Systeme durch den zusätzlichen Inhalt noch attraktiver. Solche Anwendungen ergänzen die aktuellen Bemühungen im Bereich des Semantic Web, indem es einer großen Nutzergemeinde möglich ist, in relativ kurzer Zeit eine sehr große Menge an schwach strukturierten Informationen zusammenzutragen. Eine Gruppe von Systemen befasst sich mit der Verwaltung, dem Austausch und der Speicherung von Bookmarks, die so genannten *Social Bookmarking Systems*. Sie bieten häufig die Möglichkeit Tags zur Beschreibung der Bookmarks zu vergeben und werden dann auch als *Tagging-Systeme* bezeichnet (siehe Kapitel Tagging-Systeme).

In diesem Abschnitt wird zunächst das Grundprinzip von Bookmarking-Systemen beschrieben, um den gemeinsamen Kern dieser Systeme zu erarbeiten. Anschließend wird ein Überblick über den aktuellen Stand gegeben und die Vorteile einzelner Systeme aufgezeigt. Im letzten Teil werden aktuelle Forschungsfragen diskutiert und diese mit Anwendungen aus der Praxis verknüpft.

1.2 Folksonomies

Erste Bookmarking-Systeme entstanden schon Ende der 90er, hatten aber keine schnelle und einfach zu bedienende Ajax-Oberfläche¹ (siehe Kapitel technische Aspekte) und ein schwaches Geschäftsmodell. Sie konnten sich daher nicht bei der breiten Öffentlichkeit durchsetzen. 2003 veröffentlichte Joshua Schachter das heute wohl bekannteste Social-Bookmarking-System del.icio.us (siehe Abschnitt 1.3.1) und entwickelte es basierend auf Vorschlägen

¹ <http://www.adaptivepath.com/ideas/essays/archives/000385.php>

2Fehler! Kein Text mit angegebener Formatvorlage im Dokument. Fehler! Kein Text mit angegebener Formatvo

seiner Nutzer sehr erfolgreich weiter. 2004 gelang ihm der Durchbruch mit seinem System, und die Nutzerzahl stieg dramatisch an. Ende 2005 verkaufte er das System an Yahoo. Es folgten weitere Systeme wie Connotea, Simpy, Furl, Stumbleupon, Netvouz und BibSonomy im Jahre 2005/2006, um nur einige zu nennen. Mittlerweile gibt es eine sehr große Anzahl an Systemen, die eine sehr breite Nutzerbasis aufweisen und mit ganz unterschiedlichen Features ausgestattet sind.

Der Kern dieser Systeme ist immer gleich aufgebaut. Die zentrale Struktur wird Folksonomy genannt, die im folgenden Abschnitt beschrieben und formal definiert wird. Es wird auf spezielle Eigenschaften von Folksonomies eingegangen. Der Abschnitt endet mit einer Diskussion von prinzipiellen Vor- und Nachteilen von Bookmarksystemen.

1.2.1 Definition

Kooperative Bookmarking-Systeme bieten dem Benutzer die Möglichkeit, Referenzen auf Webseiten im System zusammen mit Schlagworten zu speichern. Diesen Vorgang nennt man auch Tagging, da die beschreibenden Schlagworte - Tags genannt - zusammen mit der Referenz auf die Webseite, dem Bookmark, durch einen Benutzer im System abgelegt werden. Das Resultat ist ein Post. Abbildung 1 stellt diesen Vorgang schematisch dar. Im Zentrum steht der Post – den ein Benutzer beim Taggen erzeugt. Ein Post beschreibt immer eine Ressource/Webseite durch einen oder mehrere Tags, die die Ressource möglichst gut aus Sicht des Benutzers charakterisieren. Mehrere Benutzer können die gleiche Ressource mit zum Teil unterschiedlichen Schlagwörtern taggen. Zentrale Bestandteile von Bookmarking-Systemen sind daher: Benutzer, Tags und Ressourcen (z.B. Bookmarks), die über eine Relation miteinander in Beziehung stehen, siehe Abbildung 1. Die Zuordnung von Tags zu Ressourcen durch den Benutzer wird auch als eine leichtgewichtige Wissensrepräsentation angesehen, die *Folksonomy* genannt wird. Leichtgewichtig deshalb, weil keine formale Semantik die Beziehung zwischen Tags, den Ressourcen und den Benutzern mittels einer Ontologie festlegt. Die Semantik der Tags ergibt sich implizit durch ihre Benutzung. Folksonomy ist ein Kunstwort aus Folks (Leute) und Taxonomy. Die Folksonomy und ihre Elemente bilden den Kern eines jeden kooperativen Bookmarking-Systems.

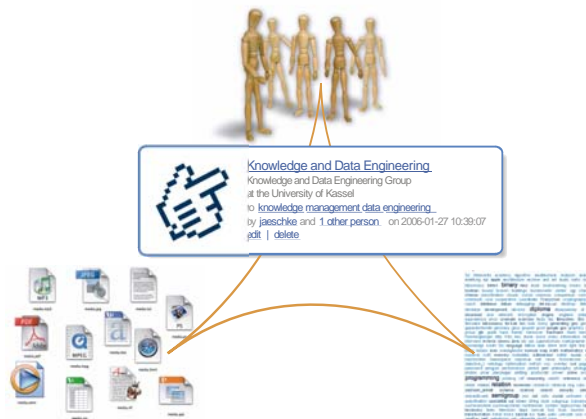


Abbildung 1: Eintrag eines Tagging-Systems am Beispiel eines Bookmarks

Wie eben eingeführt besteht eine Folksonomy 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 $F := (U, T, R, Y, \prec)$, wobei U , T und R endliche Mengen sind, deren Elemente man Benutzer, Tags bzw. Ressourcen nennt.

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 Benutzers $u \in U$ 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 kann man äquivalent als tripartiten (ungerichteten) Hypergraphen $G = (V, E)$ betrachten, 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.

Die benutzerspezifischen Tagrelationen wurden in Anlehnung an Ober- und Unterkonzepte aus dem Semantic Web definiert. Sie erlauben eine benutzerbezogene Strukturierung der Tags, erhöhen die Übersichtlichkeit im System und erweitern die Anfragemöglichkeiten (siehe Abschnitt 1.4.3).

1.2.2 Eigenschaften

Der so entstehende tripartite Graph repräsentiert ein indirektes Beziehungsnetzwerk zwischen den Benutzern des Systems, den Ressourcen und ihren Themen ausgedrückt durch Tags. Diese strukturellen Zusammenhänge (repräsentiert durch den Graphen) lassen sich für unterschiedlichste Aufgaben, z.B. zum Ranking oder für Recommendation-Ansätze etc. ausnutzen (siehe Abschnitt 1.4). Im Folgenden werden weitere Eigenschaften der Folksonomy diskutiert, die Rückschlüsse auf die Benutzung oder das Verhalten der Benutzer im System ermöglichen.

In [Cattuto et al. 2007] wurde die Struktur der Folksonomies von del.icio.us und BibSonomy mit Graphanalyseansätzen untersucht. Es konnte gezeigt werden, dass der Graph Small-World-Eigenschaften aufweist [Newman 2000]. Das heißt, dass es im Schnitt sehr kurze Wege zwischen den Knoten gibt und Gruppen vorhanden sind. Dies erklärt unter anderem, warum auf der einen Seite sehr schnell von einem Thema zu einem anderen gesprungen werden kann, gleichzeitig aber in der Nachbarschaft eines Eintrags viele thematisch verwandte Einträge entdeckt werden können.

Benutzer, Tags und Ressourcen folgen außerdem einer Long-Tail-Verteilung, d.h. es gibt wenige Benutzer mit vielen Einträgen und viele Benutzer mit wenigen Einträgen und gleiches gilt für die Tags und die Ressourcen [Hotho et al. 2006]. Auch Golder und Hubermann diskutieren die Struktur von Folksonomies, speziell die von del.icio.us, und identifizieren sieben verschiedene Arten von Tags [Golder und Huberman 2005]. Neben diesen strukturellen Erkenntnissen bzgl. Tags zeigen sie auch, dass die Verteilung der Tags ein ganz bestimmtes Verhältnis annimmt, welches sich nach einer Einschwingphase einstellt. Eine Ressource wird also im Prinzip durch einen bestimmten Vokabularmix beschrieben. Dieses Phänomen ergänzt Ansätze aus dem Bereich des Semantic Web, indem es sich auf so genannte „Emergent Semantics“ [Staab et al. 2002]; [Steels 1998] – d.h. entstehende Semantik – stützt. Die Semantik ergibt sich dabei durch die Konvergenz des verwendeten Vokabulars im Gegensatz zum ontologiebasierten Ansatz, der auf explizite Absprachen weniger Experten basiert und durch ein formales Modell repräsentiert wird.

Zentraler Bestandteil vieler Systeme sind intelligente Algorithmen, häufig basierend auf Methoden aus den Bereichen Data Mining, Machine Learning und Statistik, die den Anwender bei der täglichen Nutzung des Systems unterstützen. Diese Methoden sind auch Gegenstand aktueller Forschung. Nur so ist eine effektive Unterstützung der Nutzer möglich (z.B. durch Tag-Empfehlungen beim Verschlagworten oder durch die automatische Extraktion von Tag-Beziehungen), die eine Konvergenz von Web 2.0 und Semantic Web erlaubt und auf diese Weise stark und schwach strukturierte Wissensrepräsentationsansätze zusammenführen kann [Mazzocchi 2005].

1.2.3 Vor- und Nachteile

Ein großer Vorteil von Bookmarking-Systemen ist die einfache Benutzbarkeit. Es dauert nur ein paar Sekunden, um einen Post hoch zu laden, wobei die Tags ohne Einschränkung vom

Benutzer gewählt werden können. Positiv fällt auf, dass nicht nur Webseiten, sondern im Prinzip jeder Ressourcetyp getaggt werden kann, was den Zugriff auf z.B. Bilder oder Videos sehr erleichtert. Dies erklärt den großen Erfolg der Plattformen YouTube und Flickr.

Eine große Nutzergemeinschaft ermöglicht, dass schnell und einfach neue thematisch ähnliche Einträge gefunden werden können. Links zwischen den Einträgen verbinden die Nutzer miteinander und zeigen Trends im System auf, die Nutzer sonst erst später entdeckt hätten.

Bookmarking-Systeme stellen ihren Inhalt auf verschiedenste Weisen zur Verfügung - auch über Programmierschnittstellen (API). Dadurch kann der Inhalt auf unvorhergesehene Weise genutzt und mit anderen Inhalten in so genannten Mash-Ups verknüpft werden. Beispielsweise verknüpft der Dienst Plazes² die Karten von Google Maps³ mit den Positionsangaben, die Nutzer im System hinterlassen, und stellt sie übersichtlich dar. Um einen Überblick über aktuelle und archivierte Nachrichten bzw. populäre Bookmarks zu bekommen, fasst feed-Mashr⁴ die RSS-Feeds verschiedenster Dienste übersichtlich zusammen. Häufig binden Blogger ihre Bookmarklisten per RSS-Feed aus dem Social Bookmarksystem im Blog ein und erhalten so automatisch aktuelle Link-Listen zu bestimmten Themen.

Die meisten Nachteile entstehen durch die uneingeschränkte Nutzung von Tags durch nicht trainierte Nutzer, wodurch Ungenauigkeiten und Mehrdeutigkeiten entstehen. Die aus der Sprache bekannten Phänomene wie Schreibfehler, Synonyme und Homonyme - also der Mehrdeutigkeit von Worten - treten auch hier auf. Auch der Abstraktionsgrad von verwendeten Tags ist häufig nicht eindeutig bestimmbar. Ein spezielles Phänomen stellen Begriffe dar, die durch Mehrwort-Lexeme bezeichnet werden. Die Schreibweise variiert von Nutzer zu Nutzer im Bookmarksystem. Beispielsweise kann man für Europäische Union die folgenden Schreibweisen beobachten: EuropäischeUnion, Europäische_Union, Europäische-Union und „Europäische Union“.

Neben diesen sprachlichen Problemen besteht bei einer steigenden Nutzerzahl das Problem des Rankings von Einträgen beim Suchen in Bookmarksystemen. Häufig wird auch nach erweiterten Strukturierungsmöglichkeiten gefragt. Erste Lösungen werden in Abschnitt 1.4 vorgestellt.

1.3 Bookmarking-Systeme

Einen aktuellen Überblick über die wichtigsten und bekanntesten Bookmarking-Systeme befindet sich bei 3spots Blog⁵. Dabei werden die Systeme in Social Bookmark Managers,

² <http://plazes.com/>

³ <http://maps.google.com/>

⁴ <http://www.feedmashr.com/>

⁵ <http://3spots.blogspot.com/2006/01/all-social-that-can-bookmark.html>

Social Bookmarks, Semi-Social, Search-Engines Bookmarking, Public Bookmarks, Directories, Network/Communities, Format/Subject specific/Shopping, Annotations, Clippings und Special Bookmarking eingeteilt. Bei genauerer Durchsicht der Kategorien kann schnell festgestellt werden, dass die Einteilung der Systeme in diese Kategorien nicht immer einfach ist. Alternative Übersichten sind z.B. Web2List⁶ oder die Bookmark-Manager-Liste im Open Directory⁷.

Eine alternative Klassifikation von Bookmarking-Systemen findet sich unter anderen in [Marlow et al. 2006]. Es werden sieben Hauptdimensionen bei solchen Systemen unterschieden, die auch Einfluss auf die Qualität der Tags und die Benutzbarkeit des Systems haben. Diese Dimensionen sind: Tagging-Rechte, die Unterstützung beim Taggen, die Art der Zusammenfassung von Inhalten, der Gegenstand des Taggens, die Quelle des Rohmaterials, die Möglichkeit, die Ressourcen in Beziehung zu setzen oder soziale Aspekte durch das System auszunutzen. Die verschiedenen Dimensionen dieses Systems haben Einfluss auf Nutzbarkeit und Gestaltung der Anwendungen. Sie stehen orthogonal zur ersten Einteilung, die eher auf verschiedene Systemtypen abzielt.

Arbeiten von HAMMOND [Hammond et.al. 2005] und LUND [Lund et.al. 2005] bieten einen guten Überblick über soziale Bookmarking-Systeme im Allgemeinen, wobei bei weitem nicht so viele Systeme wie im 3spot Blog besprochen werden. Im Folgenden wird eine kleine Auswahl der bekanntesten Bookmarking-Systeme vorgestellt. Um Systeme mit einer um Literaturlisten erweiterten Funktionalität geht es im anschließenden Abschnitt. Systeme, die alternative Ressourcen zum Gegenstand des Taggings machen, werden in dem Kapitel „Grundlagen des Social Taggings“ eingeführt und ausführlich diskutiert.

1.3.1 Taggen von Webseiten

Del.icio.us⁸: Das bekannteste Social-Bookmarking-System ist Del.icio.us. Es ging 2003 online und gehört heute Yahoo. Die Nutzung des Systems ist kostenlos. Seit der Übernahme durch Yahoo wird auf einigen Seiten Werbung eingeblendet. Die Anzahl der Nutzer wird auf knapp 2 Mio. geschätzt. Im Kern folgt es der zuvor beschriebenen Folksonomy-Idee (siehe Abschnitt 1.2) und erlaubt das Taggen von Bookmarks. Es hat ein schlichtes Design und unterstützt den Nutzer bei der Navigation durch das System, indem Aggregate und Links zu anderen Nutzern aufgezeigt werden. Die Tags werden durch Leerzeichen getrennt. Es gibt ein klares URL-Schema, was den direkten Zugriff auf Informationen über die URL erlaubt (z.B. Tag-Seiten fangen immer mit /tag/ and und populäre Seiten zu einem Tag immer mit /popular/, siehe Abbildung 2). Neben den üblichen Zusammenfassungen wie der globalen oder nutzerbezogenen Tag-Wolke (Tag-Cloud) oder den aktuell populären Bookmarks (siehe Abbildung 2) bietet das System Netzwerkfunktionen: Die erste und bekannteste Funktion ist die Inbox. Sie erlaubt es, anderen Benutzern interessante Links mit einem speziellen Tag zu

⁶ <http://web2list.com/>

⁷ http://www.dmoz.org/Computers/Internet/On_the_Web/Web_Applications/Bookmark_Managers/

⁸ <http://del.icio.us/>

1.3 Bookmarking-Systeme Fehler! Kein Text mit angegebener Formatvorlage im Dokument.

7

schicken. Über die „your network“-Funktion werden einem die Bookmarks anderer ausgewählter Nutzer in einer Liste zur Verfügung gestellt. Außerdem kann geprüft werden, wer sich für die eigenen Bookmarks interessiert. Auf Tag-Ebene wird eine ähnliche Funktion per „subscription“ angeboten. Einen großen Vorteil bietet die REST-API. Sie wird häufig zur Integration der Daten in andere Systeme eingesetzt. Auch das Firefox-Plugin setzt hier auf. Es wurde eine Vielzahl von Tools⁹ entwickelt, die die Arbeit mit del.icio.us erheblich erleichtern und es in viele andere Systeme integrieren.

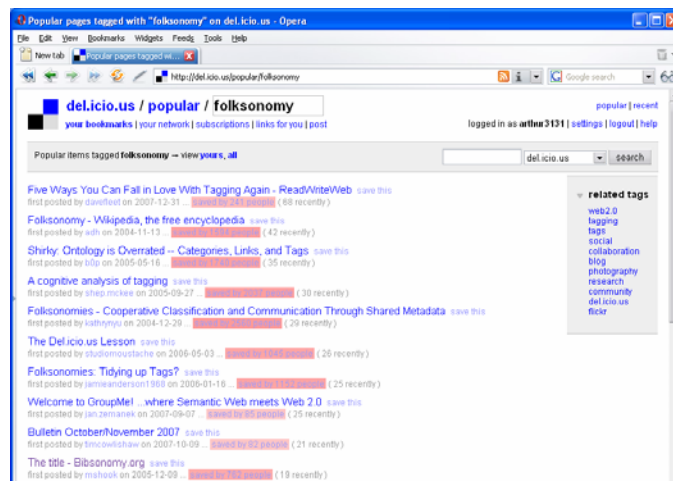


Abbildung 2: Populäre Webseiten zum Thema Folksonomy in del.icio.us

Simpy¹⁰: Ein alternatives Bookmarking-System bietet Simpy an. Auch hier ist die Nutzung frei. Die Benutzeroberfläche ist übersichtlicher und ansprechender gestaltet als bei del.icio.us. Neben der Möglichkeit, die Bookmarks anderer Nutzer zu abonnieren, bietet Simpy (im Gegensatz zu del.icio.us) Gruppen an. Eine Bookmark-History-Funktion gibt Einblick in das Taggingverhalten der Nutzer.

Mister Wong¹¹: Eines der bekanntesten deutschsprachigen Tagging-Systeme ist Mister Wong. Neben den üblichen Funktionen enthält jeder Post einen Link auf andere „empfohlene“ Webseiten.

Es gibt eine Reihe weiterer Dienste. Einige bieten Rating-Funktionen (FURL¹² oder blinklist¹³). Bei Feed Me Links¹⁴ können Bookmarks auch per Mail hinzugefügt werden. Eine

⁹ <http://del.icio.us/help/thirdpartytools>

¹⁰ <http://www.simpy.com/>

¹¹ <http://www.mister-wong.de/>

¹² <http://www.furl.net/>

¹³ <http://www.blinklist.com/>

8Fehler! Kein Text mit angegebener Formatvorlage im Dokument. Fehler! Kein Text mit angegebener Formatvo

automatische erzeugte Hierarchie bietet RawSugar¹⁵. Tags werden dabei auf der Basis von Beziehungsinformationen anderer Nutzer zu Gruppen zusammengefasst. Bei den meisten Systemen basiert die Annotation der Bookmarks auf Tags. backflip¹⁶ oder AllMyFavorites.net¹⁷ setzen hingegen auf Ordner und Chipmark¹⁸, Spurl¹⁹ oder Netvouz²⁰ auf eine Kombination von Tags und Ordnern. Viele Systeme ermöglichen den Zugriff für externe Software über eine API und bieten Firefox-Erweiterungen für den leichten Zugriff auf die Bookmarks an.

Neben dem allgegenwärtigen Gebrauch von Folksonomies als Werkzeug zur Verwaltung von Ressourcen im Web gibt es weitere nahe liegende Anwendungen. Eine Anwendung zeigt sich in Intranets [Millen et al. 2006]. Da Intranetinhalte 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 Intranet²¹ und Microsoft deren Einsatz in seinem neuen Betriebssystem.²²

¹⁴ <http://feedmelinks.com/portal>

¹⁵ <http://www.rawsugar.com/>

¹⁶ <http://www.backflip.com/>

¹⁷ <http://www.allmyfavorites.net/>

¹⁸ <https://www.chipmark.com/Main>

¹⁹ <http://www.spurl.net/>

²⁰ <http://www.netvouz.com/>

²¹ http://thecommunityengine.com/home/archives/2005/03/ibms_intranet_a.html

²² http://thecommunityengine.com/home/archives/2005/03/xfolk_an_xhtml.html



Abbildung 3: BibSonomy organisiert gleichzeitig Bookmarks und BibTeX-Referenzen.

1.3.2 Wissenschaftliche Publikationen und Bücher

BibSonomy²³ ist ein webbasierter Dienst zur Verwaltung von Webseiten und Publikationen der seit Anfang 2006 online ist und vom Fachgebiet Wissensverarbeitung²⁴ der Universität Kassel betrieben wird. Dieses für jeden frei nutzbare System erlaubt es, Lesezeichen (Favoriten, Bookmarks) für Webseiten zentral auf dem BibSonomy-Server abzuspeichern und zu verschlagworten. Als zweite Kern-Komponente wurde in BibSonomy eine kollektive Literaturverwaltung eingebaut (siehe Abbildung 3). Für interessierte Communities (z. B. Universitätsinstitute oder Projektteams) bietet BibSonomy die Einrichtung von Anwendergruppen im System an, mit der sowohl der interne als auch der externe Literaturausaustausch organisiert werden kann. BibSonomy verwendet das BibTeX-Format zur Speicherung der Publikationsdaten. Seine Publikationsverwaltung ist somit direkt in das Satzsystem LaTeX integriert, mit dem Forscher (insbesondere aus den Naturwissenschaften) ihre wissenschaftlichen Veröffentlichungen gleich druckfertig gestalten. Das System erzeugt automatisch Literaturlisten in weiteren Formaten (z.B. in RTF, EndNote, XML, RDF), so dass die einmal eingegebenen Daten in verschiedenen Kontexten genutzt werden können.

BibSonomy ist derzeit das einzige System, das die Verwaltung von Lesezeichen und Publikationen verbindet. Es wurde mit Fokussierung auf die Anwendbarkeit im akademischen Bereich entwickelt; die Rückmeldungen von vielen Forscher-Kollegen sind in die Entwicklung des Systems eingeflossen. Das System hat 20.540 registrierte Anwender, die sich 117.711 Bookmarks und 49.438 Publikationen teilen (Stand 11.9.2007). Hierzu kommen 907.506 Publikationen und 17.741 Tagungs- und Personen-Homepages, die automatisch von der Computer Science Library DBLP²⁵ übernommen werden. BibSonomy verzeichnet derzeit im Schnitt pro Tag 50.000 Besuche mit 500.000 Seitenzugriffen; mit steigender Tendenz.

²³ <http://www.bibsonomy.org/>

²⁴ <http://www.kde.cs.uni-kassel.de/>

²⁵ <http://dblp.uni-trier.de/>

Während BibSonomy sowohl Bookmarks also auch Referenzinformationen abspeichern kann, zielen CiteULike²⁶ als auch Connotea²⁷ nur auf das Verwalten von wissenschaftlichen Büchern und Publikationen. Alle Systeme unterstützen die Nutzer durch die automatische Übernahme von Metadaten aus bekannten Digitalen Bibliotheken oder Onlineangeboten von Verlagen. Obwohl das alternative System LibraryThing²⁸ auch Bücher verwalten kann, hat es eine andere Zielgruppe. Es folgt der Idee eines digitalen Bücherregals, das im Internet abgelegt ist. Wissenschaftliche Abhandlungen stehen nicht im Vordergrund.

BibSonomy ist mit seinen vielfältigen Funktionen nicht nur für Forscher und Privatanwender eine interessante, im Web stets verfügbare Plattform, sondern kann auch die typischerweise in Unternehmens-Intranets beobachteten Probleme beim Wiederfinden von Dokumenten/Informationen überwinden. Es ist in der Lage, Verbindungen zwischen ganz unterschiedlichen Systemen im Unternehmen herzustellen und kann auf diese Weise als Wissensmanagementanwendung in einem kommerziellen Umfeld eingesetzt werden. Eine zentrale Rolle nimmt dabei der in Abschnitt 1.4.1 beschriebene Rankingansatz ein, der mittelbar auf Basis der durch die Mitarbeiter gesammelten Informationen des Unternehmens die Bedeutung einzelner interner Webseiten/Ressourcen bezüglich gegebener Suchbegriffe berechnet. Auf diese Information können bisherige unternehmensinterne Suchmaschinen nicht zurückgreifen. Dieses deutlich verbesserte Ranking ist nicht auf die Anwendung von Webseiten beschränkt, da jegliche Art von Dokumenttypen mittels BibSonomy verschlagwortet werden kann. Wichtig ist nur, dass die Klassifikationen vieler Mitarbeiter als kollektive Leistung ins System einfließen und so eine gemeinsame Sicht auf die internen Informationen entsteht.

1.4 Aktuelle Forschungsansätze

Die im letzten Abschnitt beschriebenen Systeme entwickeln sich permanent weiter. Neue Funktionen sollen die Benutzbarkeit des Systems weiter steigern und eine zunehmende Verwendung in Unternehmen fördern. Gleichzeitig bieten sie auch Raum für neue algorithmische Ansätze, die im Verborgenen dem Benutzer die tägliche Arbeit erleichtern, gleichzeitig aber Funktionen bereitstellen, die eine Konvergenz und Interaktion verschiedenster Systeme erlauben.

1.4.1 Ranking und Trend-Entwicklung

Die Popularität von Social-Bookmarking-System hat in den letzten Jahren stark zugenommen. Dadurch sammeln immer mehr Nutzer Informationen in diesen Systemen. Die große Menge lockt weitere Nutzer an, schafft aber auch Anreize, neue Informationen zu entdecken.

²⁶ <http://www.citeulike.org/>

²⁷ <http://www.connotea.org/>

²⁸ <http://www.librarything.com/>

Liefert das System aber zu viele Informationen auf Suchanfragen, so steht der Nutzer vor dem Problem, dass er häufig die relevanten Treffer in der großen Menge nicht identifizieren kann. Rankingverfahren, wie sie bei Suchmaschinen zum Einsatz kommen, sollen den Anwender an dieser Stelle unterstützen.

Heutige Systeme setzen auf die Popularität der Einträge oder die Wichtigkeit der Nutzer und nutzen dies zum Ranking. [Michail 2005] und [Szekely und Torres 2005] bestimmen die Rangfolge der Treffer für Suchanfragen in zwei Schritten. Erst ermitteln Sie wichtige Benutzer bzgl. des gesuchten Themas basierend auf den Daten des Systems. Im zweiten Schritt wird das Ranking auf Basis der Nutzergewichtung berechnet, d.h. Ressourcen wichtiger Nutzer erzielen typischerweise eine bessere Position bei entsprechenden Suchanfragen.

In HOTHO ET AL. [Hotho et al. 2006b] wird der FolkRank-Algorithmus vorgestellt. Diese Weiterentwicklung des PageRank-Algorithmus für Folksonomies erlaubt es, in Abhängigkeit von vorgegebenen Anwenderpräferenzen ein Ranking über Tags, Benutzer und Ressourcen in einem Schritt zu erstellen, so dass Elemente der Folksonomy, die ähnlich zur vorgegebenen Präferenz sind, ein hohes Gewicht erhalten. Damit ist es möglich, auf der Basis der Benutzer des Systems beliebige Ressourcen in eine Reihenfolge zu bringen. Unser Ansatz basiert auf der PageRank-Idee von Google. Dabei nutzt Google bei der Berechnung des PageRanks im Internet eine ähnliche Information durch die Auswertung der Linksstruktur zwischen den Webseiten [Page et al. 1998]. Zusätzlich können mit dem gleichen Algorithmus auch Nutzer als Experten zu einem bestimmten Thema identifiziert werden. Einen zum FolkRank sehr ähnlichen Ansatz stellt BAO vor [Bao 2007].

Können aufkommende Trends automatisch in solchen Systemen entdeckt werden, dann kann dies genutzt werden, um Anwender gezielt auf diese Trends aufmerksam zu machen. Die Hauptseite von Del.icio.us macht dies, indem sie auf aktuell häufig gespeicherte Bookmarks hinweist. Die Visualisierung der Popularität von Flickr-Tags über die Zeit auf einer speziell dafür vorbereiteten Webseite²⁹ sowie ein entsprechender Algorithmus wird in DUBINKO [Dubinko et al, 2006] vorgestellt. Ebenfalls in HOTHO ET AL. [Hotho et al. 2006c] werden Rankingverfahren für die Entdeckung von Trends in del.icio.us Daten angewendet.

1.4.2 Empfehlungssysteme

Die häufigste Anwendung von Empfehlungssystemen (Recommendern) im Bereich von Social Bookmarking ist die Empfehlung von Tags für aktuell zu taggende Webseiten. Das System schlägt mehrere Tags vor, aus denen der Anwender auswählen kann, mit dem Ziel, dem Nutzer das Taggen so einfach wie möglich zu machen. Dabei steht es ihm frei, die Vorschläge zu ignorieren oder aus ihnen auszuwählen. Solche Recommender-Ansätze erleichtern dem Nutzer den Umgang mit Bookmarking-Systemen. Durch die Präsentation der Tags während des Verschlagwortens kann der Nutzer leicht vorgeschlagene Tags per Mausklick oder durch die automatische Vervollständigung auswählen bzw. übernehmen. Tippfehler und unterschiedliche Schreibweisen können auf diese Weise vermieden werden und führen zur

²⁹ <http://www.research.yahoo.com/taglines/>

Vereinheitlichung und der Konvergenz des verwendeten Vokabulars. Vorteilhaft ist auch, dass das System nicht nur Tags des Benutzers, sondern auch Tags von anderen Nutzern vorschlägt. Auf diese Weise macht es den Anwender auf Tags aufmerksam, auf die er in dieser Situation wahrscheinlich nicht gekommen wäre.

Ein erster Collaborative-Filtering-Ansatz (CF) für kollaborative Verschlagwortungssysteme wird in BENZ [Benz et al 2007] vorgestellt. Hier wird eine Kombination von Nächste-Nachbarn-Klassifikatoren verwendet, um Bookmarks zu klassifizieren. Das Nächste-Nachbarn-Verfahren sucht unter allen Nutzern des Systems die, die dem aktuellen Nutzer am ähnlichsten sind. Die Annahme ist, dass sie das zu klassifizierende Bookmark ähnlich einsortiert haben, wie es auch der aktuelle Nutzer machen wird. Auf diese Weise wird eine gute Vorhersage der zu wählenden Tags berechnet.

XU ET AL. [Xu et al. 2006] nutzt die Graphstruktur und stellt einen Algorithmus auf der Basis von HITS [Kleinberg 1999] vor. In JÄSCHKE ET AL. [Jäschke et al. 2007] wird der FolkRank-Algorithmus für die Vorhersage von Tags verwendet. Es konnte empirisch eine deutliche Verbesserung von Precision (der Güte der vorhergesagten Tags) und Recall (der Menge der Tags, die das System korrekt vorgeschlagen hat) gegenüber inhaltsbasierten und CF-Ansätzen nachgewiesen werden.

1.4.3 Begriffshierarchien

Die Extraktion von Begriffshierarchien aus Folksonomies wäre ein wesentlicher Schritt zur Konvergenz mit dem Semantic Web. Dabei besteht der erste Schritt in der Identifikation von Synonym-Beziehungen zwischen Tags gefolgt von der Identifikation von allgemeinen Ober-/Untertag-Beziehungen, die später verfeinert werden können. Die Tag-Beziehungen dienen im Wesentlichen der Strukturierung der Begriffswelt, die zum Taggen verwendet wurde.

Die Modellierung von Begriffshierarchien - bzw. in diesem Fall Taghierarchien - erlaubt dem Nutzer die Strukturierung der verwendeten Tagsammlung. Dies ist notwendig, wenn die Tagwolke des Anwenders eine gewisse Größe erreicht hat und die Übersichtlichkeit nicht mehr gewährleistet ist. Erste Ansätze in dieser Richtung findet man bei del.icio.us unter der Bezeichnung *Bundels*, in die man Tags gruppieren kann. Ansätze zur automatischen Extraktion solcher Gruppen oder auch ganzer Hierarchien würden den Anwender bei der Organisation seines abgelegten Wissens zusätzlich unterstützen, sind aber in den bekannten Systemen noch nicht vorhanden.

MIKA [Mika 2005] definiert ein Modell von semantisch-sozialen Netzwerken, um aus Daten von del.icio.us leichtgewichtige Ontologien zu extrahieren. Er zeigt die Berechnung von Maßen wie z. B. dem Clustering-Koeffizienten oder der Zentralität von bestimmten Folksonomy-Elementen, die unter Ausnutzung von Kozitationsbeziehungen die Folksonomy clustert. Dadurch kann eine einfache Ontologie gelernt werden. HEYMANN UND GARCIA-MOLINA [Heymann und Garcia-Molina 2006] und SCHMITZ [Schmitz 2006] schlagen alternative Clusteransätze zur Identifikation der Beziehungen vor. BEGELMAN ET AL. [Begelman et al. 2006] basieren ihren Ansatz auf dem Tag-Tag-Co-Occurrence-Graphen und verwenden einen Graphclusteringansatz, um die Cluster zu berechnen. GRAHL ET AL. [Grahl et al. 2007] haben einen begrifflichen, hierarchischen Clusteransatz für Folksonomies

entwickelt. Diese Ansätze können verwendet werden, um Hierarchien von Tags für eine bessere Strukturierung der Tag-Menge zu lernen. SCHMITZ ET AL. [Schmitz et al 2006] adaptierten Assoziationsregeln auf die Folksonomy-Strukturen und lieferten interessante Einblicke in die von den Nutzern der Bookmarking-Systeme zusammengetragenen Daten.

1.4.4 Entdeckung von Nutzergruppen

Neben organisatorischen Gruppen entstehen in den Systemen mittelbar thematische Gruppen. Eine Identifikation solcher Gruppen böte die Möglichkeit, themenspezifisch auf andere Nutzer zu verweisen, und so das Navigieren der Benutzer durch die Systeme zu lenken oder auf spannende Inhalte aufmerksam zu machen. Gruppen werden in der Regel automatisch mittels Clusterverfahren identifiziert, die aber in diesem Fall die tripartite Struktur der Daten (siehe Abschnitt 1.2.1) verarbeiten müssen. JÄSCHKE ET AL. [Jäschke et al 2006] erweitern die Formale Begriffsanalyse auf die triadische Struktur von Folksonomies, um synchron Cluster von Benutzern, Tags und Ressourcen berechnen zu können. Damit steht ein erstes Verfahren zur Identifikation von Nutzergruppen zur Verfügung, das gleichzeitig auch eine Beschreibung der Gruppen durch Mengen von Tags bietet und die passenden Ressourcen identifiziert. Dadurch kann man jedem Nutzer thematische Links zu Nutzern mit ähnlichen Interessen anbieten und so den Wissensaustausch und die Interaktion zwischen Nutzern erhöhen.

1.4.5 Spambekämpfung

Der Missbrauch von Social-Bookmarking-Systemen durch Spammer nimmt stetig zu. Wie auch bei E-Mail oder Web Spam stehen die Systembetreiber vor der Aufgabe, Spammer schnell und effektiv erkennen zu können. KOUTRIKA ET AL. [Koutrika et.al. 2007] nutzen die Vertrauenswürdigkeit der Anwender zum Ranken von Dokumenten und sind so in der Lage, auch Spam-Attacken zu entdecken. Die effektive Spambekämpfung in Social Bookmarking Systemen ist noch ein offenes Feld, und es wird an dem Transfer bekannter Ansätze aus den Bereichen Suchmaschinen-Spam und E-Mail-Spam auf Social Bookmarking Systeme gearbeitet.

Social Bookmarking Systeme bieten auch dem unerfahrenen Nutzer direkt Vorteile und erfreuen sich unter anderem wegen ihrer Einfachheit großer Beliebtheit. Die weitere Unterstützung der Nutzer durch die Analyse der zusammengetragenen Daten stellt nicht nur für die Forschung sondern auch für die Systementwickler eine große Herausforderung dar. Sie bietet aber auch die Chance in Zukunft die Informationsorganisation mit Hilfe der Gemeinschaft effektiv zu gestalten und mit Hilfe von neuen Methoden die Ausnutzung der gemeinschaftlichen Arbeit durch z.B. Spammer zu unterbinden.