

In diesem Kapitel werden die technischen Grundlagen des Social Bookmarkings, dass durch das Buch hindurch als Beispielanwendung des Web 2.0 dienen wird, eingeführt. Ausgehend vom Web und der Einordnung und Definition von Begriffen wie Web 2.0 und Social Web werden in Abschn. 2.2 Social Bookmarking-Systeme als Teil des Web 2.0 eingeführt. Abschn. 2.3 beschreibt typische Funktionalitäten solcher Systeme, während Abschn. 2.4 die zu grundlegende Datenstruktur, die Folksonomy formal definiert. Abschn. 2.5 geht kurz auf die Eigenschaften von Folksonomies ein und die Vor- und Nachteile werden in Abschn. 2.6 diskutiert.

2.1 Vom Web zum Web 2.0

Das World Wide Web (WWW) hat sich seit seiner Geburtsstunde im Jahr 1989 rasant entwickelt. Stand am Anfang der einfache Informationsaustausch zwischen Wissenschaftler im Vordergrund, werden heute vielfältige Dienste angeboten. Eine große Revolution im Web gab es vor ca. 10 Jahren, als das Web 2.0 entstand.

Das Web 2.0 ist im Wesentlichen durch zwei Aspekte gekennzeichnet. Auf der einen Seite wurde das Web interaktiver, d.h. Nutzer können Inhalte nicht nur konsumieren, sondern auch selber zur Verfügung stellen. Auch wurden technische Neuerungen eingeführt, die interaktive Webseiten deutlich einfacher möglich machten. Inhalte lassen sich nun durch neue Dienste miteinander verknüpfen, so dass den Nutzern eine ganz neue Art des Webs zur Verfügung steht. In Anlehnung an die Weiterentwicklung von Software, die mit fortlaufenden Versionsnummern dokumentiert wird, wird diese neue Version des Webs auch als Web 2.0 bezeichnet. Heute hat sich neben dem Web 2.0 auch der Begriff Social Media etabliert, der im Grunde alle aktuellen Dienste und Angebote des heutigen Webs umfasst.

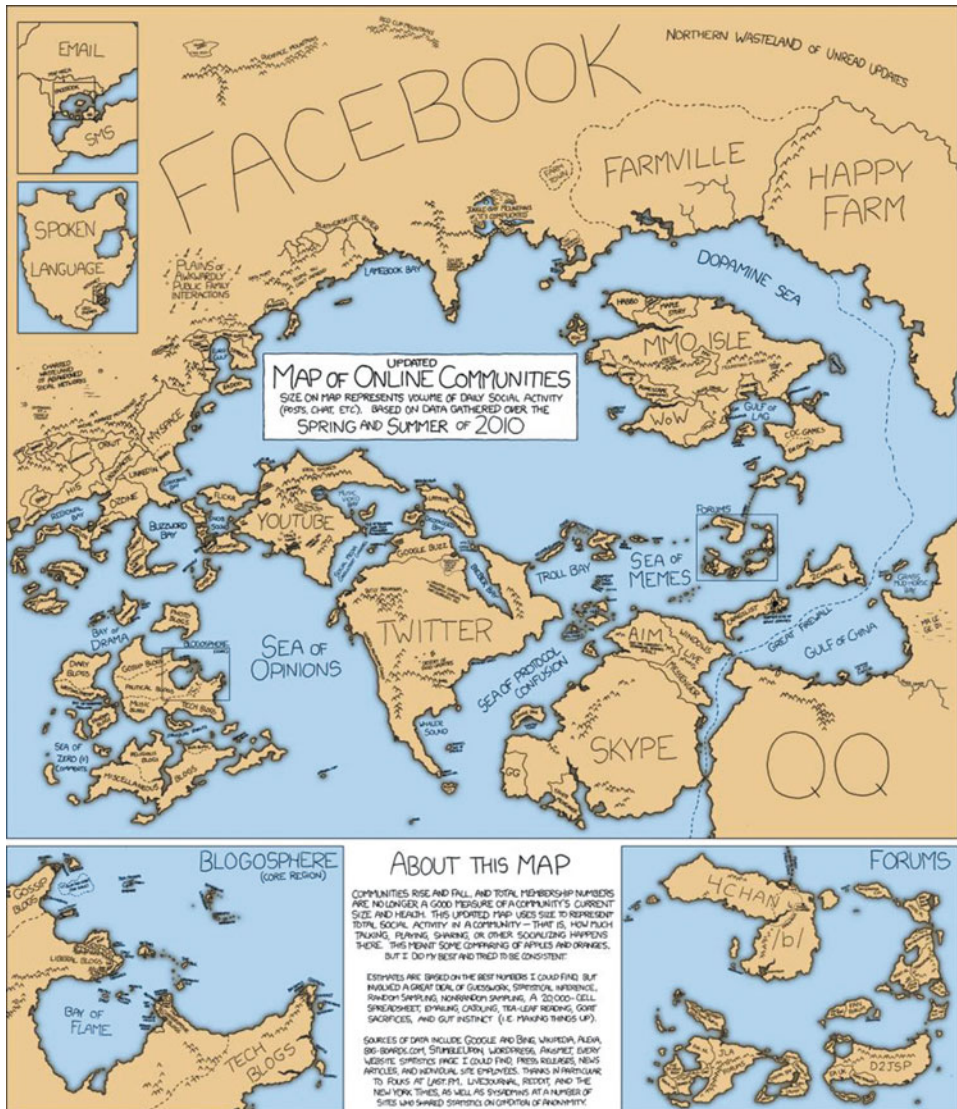


Abb. 2.1 Illustration der aktuellen Web 2.0 Dienste. Größen spiegeln die Wichtigkeit der Dienste wider. (Quelle: <http://xkcd.com/802/>)

Einen Überblick über die aktuelle Dienste des Web 2.0 gibt Abb. 2.1. Man erkennt eine deutliche Dominanz der Sozialen Netze wie Facebook¹ (gesamter oberer Bereich) oder des Kurznachrichtendienstes Twitter² (zentral in der Karte). Zentrale Bestandteile des

¹ <https://www.facebook.com/>.

² <https://twitter.com/>.

Web 2.0 sind nach wie vor Kommunikationsdienste wie Skype³ oder auch ICQ.⁴ Wichtig sind auch Dienste wie Blogs oder Foren, die zum Informationsaustausch genutzt werden. Aus Sicht der Kommunikation nehmen Social Bookmarking-Dienste wie Delicious⁵ oder auch die Wikipedia⁶ nur einen kleinen Teil ein. Sie erlauben aber das Speichern von zentralen Informationen und stellen aus diesem Grund einen wichtigen Teil des Web 2.0 dar. Die gesamten Kommunikationssysteme aller Nutzer im Web sind in der Karte in der linken oberen Ecke dargestellt. Dabei wird klar, dass Email und SMS die gesamte Onlinekommunikation deutlich dominieren.

Ein weiteres Phänomen der letzten Jahre im Web ist das „Semantische Web“ oder auch Semantic Web. Ziel des Semantic Web ist die Unterstützung des intelligenten Datenaustauschs durch Maschinen im Web, um so die nächste Stufe des Webs zu erreichen. Daten sollen in standardisierter Form z. B. in RDF ausgetauscht werden. Ein zentrales Element des Semantic Web sind Ontologien, die das Wissen in formaler Form repräsentieren und so eine Art gemeinsame Basis schaffen (siehe [2]). Das Web 2.0 und das Semantische Web haben klare Berührungspunkte, ergänzen sich aber auch hervorragend. So werden die im weiteren Verlauf dieses Kapitels eingeführten Folksonomies als leichtgewichtige Ontologien angesehen. Es gibt auch eine Reihe von Arbeiten, die die semantischen Beziehungen von Tags in Folksonomies untersuchen und hier eine Anreicherung der Strukturen in Richtung mehr Semantik vorstellen (z. B. [14]). Auf der anderen Seite kommen semantische Technologien wie z. B. XML, RDF oder OWL immer häufiger im Web 2.0 zum Einsatz (z. B. Atom oder FOAF).

Die Begriffe Web 2.0 und Social Web werden häufig auch als Synonym verwendet. Der Begriff „Social Web“ stammt aus dem Englischen und stellt die sozialen Beziehungen, die durch die neuen Dienste des Web 2.0 modelliert werden, in den Vordergrund. Hauptvertreter dieser Art von Systemen sind Facebook und Twitter. Bei beiden kann man in unterschiedlicher Art mit Freunden in Kontakt treten oder diesen Zugriff auf persönliche Informationen geben. Der Begriff „social“ ist aber im Web 2.0 noch weiter gefasst. So bezeichnet man auch Bookmarking-Systeme als social, da man die Daten nicht nur für sich selbst sammelt, sondern diese auch anderen Nutzern im System zugänglich macht. Diese indirekte soziale Interaktion ist neu, und erlaubt den Nutzern neuartige Einblicke in die Interessen anderer – zum Teil sogar völlig unbekannter – Nutzer. Ähnliches kann man auch bei Wikis und in der Blogosphäre beobachten. Diese Systeme werden im Folgenden kurz angesprochen.

Wiki: Ein Wiki ist eine Websoftware zum einfachen, schnellen und gemeinsamen Schreiben von Text. Jeder Nutzer kann den gemeinsam verfassten Text ändern. Meist wird dieser Prozess durch eine Versionsverwaltung unterstützt. Wikisysteme nutzen eine Markup-sprache zum Auszeichnen von Formatierungen, Links und anderen wichtigen Elementen

³ <http://www.skype.com/>.

⁴ <http://www.icq.com/>.

⁵ <https://delicious.com/>.

⁶ <http://www.wikipedia.org/>.

im Text. Es gibt eine sehr große Anzahl von Wikisoftwareprodukten⁷, die in unterschiedlichen Programmiersprachen erstellt wurden und für eine Vielzahl von Anwendungsfällen genutzt werden. Der bekannteste Vertreter der Wikisoftware ist das Mediawiki. Es dient als Basis für die gemeinsam editierte Enzyklopädie Wikipedia, an der mehrere tausend Editoren und eine noch viel größere Anzahl an Autoren mitgewirkt haben. Die Weisheit der Vielen führt bei der Wikipedia zu einem Lexikon, das sich qualitativ mit der Encyclopaedia Britannica⁸ messen kann und aus der heutigen Welt nicht mehr wegzudenken ist. Verwandt sind online Editier- oder Officesysteme wie z. B. Priate Pad⁹ oder Google Drive¹⁰, die auch das gemeinsame Editieren von Text im Web erlauben.

Blog: Ein Blog ist im Grunde ein im Web geführtes, öffentliches Tagebuch. Jeder kann ohne technische Vorkenntnisse ein solches Blog erstellen, und auch jeder im Web kann es lesen. Blogs sind wie Tagebücher auch chronologisch sortiert und werden durch eine Person, die man „Blogger“ nennt, gepflegt. Ausgehend von der ursprünglichen Idee hat sich eine große Nutzergemeinschaft rund um die Blogs entwickelt, die Blogosphäre. Blogger schreiben nicht nur über persönliche Dinge – es gibt mittlerweile Blogs für jedes Thema. Interaktionen und Feedback zu den Blogbeiträgen finden über die Kommentarfunktion statt, bei der die Leser Stellung zum Thema beziehen können und dort auch Links auf weiterführende Informationen hinterlassen. Es besteht auch die Möglichkeit, dass Blogger Themen Anderer aufgreifen und einen eigenen Beitrag verfassen. Durch die Verlinkung fremder Beiträge entsteht ein Netzwerk von Diskussionen, und damit auch eine Vernetzung der Blogger in der Blogosphäre. Zugang zu den verschiedenen Beiträgen und Blogs erhält man über spezielle Suchmaschinen wie z. B. Technorati.¹¹ Ein interessanter Aspekt von Technorati ist die Nutzung eines linkbasierten Rankingverfahrens (vergleichbar zu PageRank), um die Wichtigkeit eines Blogs (und so auch eines Bloggers) zu bestimmen.

Soziale Netzwerke: Social Networks wie Facebook¹², Twitter¹³ und Google +¹⁴ sind wohl die bekanntesten Vertreter von Systemen, die im Web die Kommunikation und Interaktion mit Freunden unterstützen. Neben dem Profil mit Interessen und Bildern ist es möglich, Kontakte von Freunden zu pflegen und mit Freunden öffentlich, aber auch privat zu kommunizieren. Neuerdings gibt es auch Chat-Systeme, die immer stärker in aktuelle Smartphones Einzug halten. Neben den Systemen für jedermann gibt es auch eine Reihe

⁷ http://de.wikipedia.org/wiki/Liste_von_Wiki-Software.

⁸ <http://blog.wikimedia.org/2012/08/02/seven-years-after-nature-pilot-study-compares-wikipedia-favorably-to-other-encyclopedias-in-three-languages/>.

⁹ <http://piratepad.net/>.

¹⁰ <https://drive.google.com/>.

¹¹ <http://technorati.com/>.

¹² <https://www.facebook.com/>.

¹³ <https://twitter.com/>.

¹⁴ <https://plus.google.com/>.

von Systemen für Geschäftskunden, wie z. B. LinkedIn,¹⁵ eine weltweit bekannte Plattform, oder der in Deutschland häufig genutzte Dienst Xing.¹⁶ Neben der Kommunikation mit Geschäftspartnern und der Pflege von Kontaktinformationen bieten solche Systeme auch die Möglichkeit, Fähigkeiten von Nutzern zu bewerten oder nach Arbeit zu suchen. In Gruppen können verschiedenste Themen diskutiert werden. Gemeinsam ist allen Systemen die Abbildung von sozialen Netzwerken im Web. Diese werden nicht nur von Wissenschaftlern untersucht (z. B. [137]), sondern sie bieten auch die Geschäftsgrundlage der betreibenden Unternehmen, da sie zielgerichtete, personalisierte Werbung erlauben.

Dieser Überblick über die Web 2.0-Anwendungen ist bei weitem nicht vollständig, und es gibt eine Reihe wichtiger Dienste, die schon vor der Web 2.0-Revolution wichtig waren, wie z. B. Suchmaschinen, und die noch immer eine zentrale Rolle spielen. Ihre Beschreibung würde den Rahmen dieser Einleitung sprengen, dennoch sollte man ihren Einfluss auf das Web als Ganzes nicht vernachlässigen.

2.2 Social Bookmarking-Systeme

Im weiteren Verlauf des Kapitels wird ausführlich auf die wesentlichen Aspekte von Social Bookmarking-Systemen eingegangen. Diese werden im Buch immer wieder stellvertretend als Beispiel für Web 2.0-Systeme herangezogen. In den allermeisten Fällen lassen sich die Erkenntnisse auf andere Web 2.0-Dienste verallgemeinern.

Social Bookmarking-Systeme (oder allgemeiner Taggingssysteme) werden auch als kooperative oder kollaborative Verschlagwortungssysteme bezeichnet und erlauben das Teilen von Bookmarks, URLs und anderen Ressourcen. Sie dienen dem Verwalten, Veröffentlichenden und Teilen von Informationen und werden demzufolge auch als einfache Wissensmanagementsysteme angesehen. Sie haben leicht verständliche Nutzerschnittstellen und es bedarf keiner spezifischen Fähigkeiten, um aktiv mit ihnen arbeiten zu können. Der direkte Vorteil für jeden Nutzer, browser- und rechnerunabhängig seine Daten und Informationen ohne großen Mehraufwand zentral organisieren zu können, zieht weitere Nutzer an und macht die Systeme durch den zusätzlichen Inhalt noch attraktiver. Solche Anwendungen ermöglichen es einer großen Nutzergemeinde, 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 (auch Favoriten oder Web-Lesezeichen genannt), die so genannten Social Bookmarking-Systeme. Sie bieten häufig die Möglichkeit, frei wählbare Schlagwörter (*Tags*) zur Beschreibung der Bookmarks zu vergeben und werden dann auch als Taggingssysteme bezeichnet. Dieses Prinzip findet man nicht nur für Bookmarks, sondern auch für andere Arten von Ressourcen wie z. B. Bildern in Flickr¹⁷ oder Videos in YouTube.¹⁸

¹⁵ <http://www.linkedin.com/>.

¹⁶ <http://www.xing.com/>.

¹⁷ <http://www.flickr.com/>.

¹⁸ <http://www.youtube.com/>.

Erste Bookmarking-Systeme entstanden schon Ende der 90er, hatten aber keine schnelle und einfach zu bedienende Ajax-Oberfläche¹⁹ und ein schwaches Geschäftsmodell. Sie konnten sich daher nicht bei der breiten Öffentlichkeit durchsetzen. 2003 veröffentlichte Joshua Schachter das wohl bekannteste Social Bookmarking-System *del.icio.us*²⁰ (inzwischen Delicious) und entwickelte es basierend auf Vorschlägen seiner Nutzer sehr erfolgreich weiter. 2004 gelang ihm der Durchbruch mit seinem System, und die Nutzerzahl stieg dramatisch an. Es folgten eine Reihe weitere Systeme wie Connotea, Simpy, Furl, sowie Stumbleupon,²¹ Netvouz²² 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. In den letzten Jahren kam es zu einer Konsolidierung, und eine Reihe von Bookmarking-Systemen stellte den Dienst ein. Trotzdem führt die Webseite „Ingenious Talk“ eine Liste mit den Top 500²³ Social Bookmarking Seiten in 2013 und auch das Open Directory führt noch 39 Bookmark-Manager²⁴ auf.

Eine alternative Klassifikation von Bookmarking-Systemen findet sich unter anderen in [77]. 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 Verschlagworten, die Art der Zusammenfassung von Inhalten, der Gegenstand des Taggings, 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. Darüber hinaus bieten die Arbeiten [31] und [72] einen guten Überblick über soziale Bookmarking-Systeme im Allgemeinen.

2.2.1 Tagging von Webseiten

Im Folgenden wird eine kleine Auswahl der bekanntesten Bookmarking-Systeme vorgestellt. Um Systeme mit einer um Literaturdaten erweiterten Funktionalität geht es im anschließenden Abschnitt. Es gibt eine Reihe von Systemen, die alternative Ressourcen zum Gegenstand des Taggings machen. Details findet man z. B. in [4].

¹⁹ <http://www.adaptivepath.com/ideas/essays/archives/000385.php>.

²⁰ <https://delicious.com/>.

²¹ <http://www.stumbleupon.com/>.

²² <http://www.netvouz.com/>.

²³ <http://ingenioustalk.com/top-500-pr-2-dofollow-social-bookmarking-sites-2013/>.

²⁴ http://www.dmoz.org/Computers/Internet/On_the_Web/Web_Applications/Bookmark_Managers/.

Der Pionier unter den Social-Bookmarking-Systemen ist das wohl bekannteste System **Delicious**²⁵. Es ging 2003 online (noch als Del.icio.us), wurde dann 2005 an Yahoo²⁶ verkauft und gehört seit 2011 AVOS²⁷. Die Nutzung des Systems ist kostenlos. Die Anzahl der Nutzer wurde zwischenzeitlich auf mehr als 5 Mio. in 2008 geschätzt²⁸. Mittlerweile gibt es ein Relaunch der Seite und viele der alten Funktionen stehen nicht mehr zur Verfügung. Im Kern folgt es der Folksonomy-Idee (siehe Abschn. 2.4) und erlaubt das Tagging 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, wurde aber in den letzten Jahren durch eine Reihe von Multimediaelementen angereichert.

Die Tags wurden im alten System durch Leerzeichen getrennt; aktuell durch Kommas. Es gab ein klares URL-Schema, was den direkten Zugriff auf Informationen über die URL erlaubte (z. B. fingen Tag-Seiten immer mit/tag/an). Neben den üblichen Zusammenfassungen wie einer globalen oder nutzerbezogenen Tag-Wolken (Tag-Clouds) bot das System Networking-Funktionen: Die bekannteste Funktion war die Inbox. Sie erlaubte es, anderen Nutzern interessante Links mit einem speziellen Tag zu schicken. Über die „your network“-Funktion wurden einem die Bookmarks anderer ausgewählter Nutzer in einer Liste zur Verfügung gestellt.

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 [81]. Da Intranet-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 Nutzer 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. Daher listet Wikipedia eine Reihe von Enterprise Bookmarking Plattformen²⁹.

2.2.2 Wissenschaftliche Publikationen und Bücher

BibSonomy³⁰ ist ein webbasierter Dienst zur Verwaltung von Webseiten und Publikationen der seit Anfang 2006 online ist und vom Fachgebiet Wissensverarbeitung³¹ der Universität Kassel betrieben wird. Dieses, für jeden frei nutzbare, System erlaubt es, Lesezeichen

²⁵ <https://delicious.com/>.

²⁶ <http://www.yahoo.com/>.

²⁷ <http://avos.com/>.

²⁸ <http://www.searchenginejournal.com/deliciouscom-relaunches-enhanced-speed-search-design-with-no-dots/7403/>.

²⁹ http://en.wikipedia.org/wiki/Comparison_of_enterprise_bookmarking_platforms.

³⁰ <http://www.bibsonomy.org/>.

³¹ <http://www.kde.cs.uni-kassel.de/>.

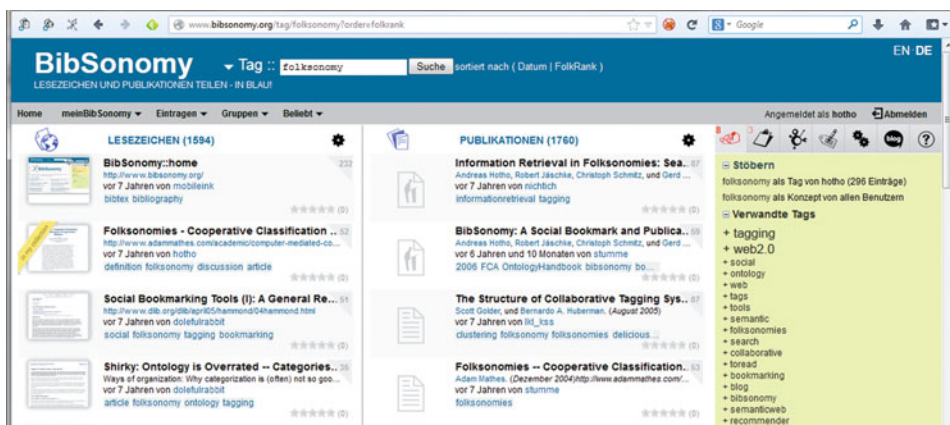


Abb. 2.2 BibSonomy organisiert gleichzeitig Bookmarks und BibTeX-Referenzen

(Favoriten, Bookmarks) für Webseiten zentral auf dem BibSonomy-Server abzuspeichern und zu verschlagworten. Als zweite Kern-Komponente wurde in BibSonomy eine kollektive Literaturverwaltung eingebaut. Details zum System findet man in Kap. 4. Abbildung 2.2 zeigt einen Ausschnitt der Home-Seite von BibSonomy.

Während BibSonomy sowohl Bookmarks also auch Referenzinformationen abspeichern kann, zielt z. B. CiteULike³² 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 LibraryThing³³ auch Bücher verwalten kann, hat es eine andere Zielgruppe. Es folgt der Idee eines digitalen Bücherregals, das im Internet abgelegt ist. Wissenschaftliche Abhandlungen stehen bei LibraryThing nicht im Vordergrund.

In folgenden werden zunächst zentrale Funktionalitäten und Eigenschaften von Social Bookmarking-Systemen herausgearbeitet, bevor dann auf den Kern von Bookmarking- bzw. Taggingssystemen, den sogenannten *Folksonomies*, eingegangen wird. Nach einer formalen Definition werden wir die Eigenschaften von Folksonomies vorstellen und ihre Vor- und Nachteile diskutieren. Die Ausführungen nächsten Abschnitte folgen dem Artikel [45].

2.3 Funktionalitäten

Social Bookmarking-Systeme bieten eine Reihe von Standardfunktionen. Daten und Struktur einer Folksonomy ermöglichen verschiedene Funktionalitäten, die Social Bookmarking-Systeme von einfachen Literaturverwaltungsprogrammen, aber auch von

³² <http://www.citeulike.org/>.

³³ <http://www.librarything.com/>.

digitalen Bibliotheken oder Suchmaschinen unterscheiden. Das grundsätzliche Ziel eines solchen Systems ist es, den Nutzer bei der Organisation von Ressourcen z. B. Bookmarks, Fotos oder Videos zu unterstützen und diesen Prozess so leicht wie möglich zu machen. Hierzu haben die Nutzer die Möglichkeit, Ressourcen mit frei wählbaren Schlagwörtern – sogenannten *Tags* – zu beschreiben. Dabei kann der Nutzer seine Ressourcen öffentlich oder privat einstellen und diese dann auch mit geeigneten Funktionen wieder finden. Nutzer, Tags und Ressourcen bilden also die Basis des Systems und ihre Verknüpfung stellt die zentrale Struktur für alle Funktionen wie Browsing, Suche, Empfehlungssysteme, Werbung und Spam-Bekämpfung dar.

In Social Bookmarking-Systemen kooperieren die Nutzer nicht zwangsweise aktiv miteinander, aber durch die Art der Systeme unterstützen diese das Sammeln von Bookmarks dadurch, dass Nutzer die Posts anderer sehen und übernehmen können. Diese indirekte Interaktion der Nutzer miteinander stellt eine Art der sozialen Kommunikation dar, was auch der Grund dafür ist, dass man diese Systeme mit dem Zusatz „social“ versteht. Die soziale Interaktion mit anderen Nutzern wird durch Systemfunktionen wie Freundeslisten oder Gruppen weiter verstärkt. Diese Funktionen stellen eine wichtige Eigenschaft dieser Art von Systemen dar.³⁴

2.3.1 Suche

Nutzer können durch die Eingabe von Suchbegriffen gezielt nach Posts suchen. Das System zeigt daraufhin Einträge an, die den gesuchten Begriff enthalten oder damit verschlagwortet wurden. Neben einer chronologischen Reihenfolge, bietet sich das Sortieren der Einträge nach Relevanz wie in Suchmaschinen an. Hierfür werden Techniken des Rankings wie in Suchmaschinen genutzt, die auf die Beziehungen zwischen Nutzern, Ressourcen und Tags zurückgreifen. Neben den Einträgen selbst lassen sich weitere Daten für ein Ranking nutzen, wie die Merkmale der Ressource selbst (Webseitentext, Titel eines Musikstückes) oder Daten über die mitprotokollierte Interaktion mit dem System. Bei der Suche kann neben der Volltextsuche auch einfach pro Nutzer oder systemweit ein Tag als Suchkriterium genutzt werden. Die aus klassischen Suchmaschinen bekannte Suche wird in Bookmarking-Systemen ergänzt durch die Suche nach Tags, aber auch durch das Browsen im System.

2.3.2 Browsing

Tags werden zur Strukturierung der Ressourcen von Nutzern eingesetzt und bieten damit die Möglichkeit, Posts anderer Nutzer explorativ zu erkunden. Jeder Nutzer kann also die Daten anderer Nutzer zur Navigation durch das System verwenden. Hierbei beobachtet

³⁴ Folgende Ausführungen orientieren sich an [69].

man häufig den Effekt, dass man unerwartet auf interessante Ressourcen stößt. Ruft ein Nutzer die Seite eines Tags³⁵ auf, werden aus einer Datenbank alle Ressourcen ausgelesen, die mit diesem Tag von irgendeinem Nutzer versehen wurden. Das System verlinkt alle Elemente der Struktur, also nicht nur die Ressourcen, sondern auch alle verwendeten Tags und die Nutzer, die die Ressourcen gepostet haben. Bei einem Klick auf den Nutzernamen werden alle Posts des entsprechenden Nutzers angezeigt, und beim Klick auf das Tag alle Ressourcen dieses Nutzers mit dem entsprechenden Tag. Durch diese Verlinkung können mit wenigen Navigationsschritten völlig andere Themengebiete betrachtet und neue Ressourcen in unerwarteter Art und Weise gefunden werden. Der Nutzer folgt den angebotenen Links und kann so über die gebotene Navigationsstruktur andere Einträge erkunden und weitere interessante Informationen finden. Zusätzlich werden Übersichtsseiten zur Verfügung gestellt, auf denen z. B. die populärsten Ressourcen oder die am häufigsten genutzten Tags angezeigt werden. Der Erfolg dieser einfachen selbstbestimmten Suche (Browsing) beruht darauf, dass viele Nutzer ihre Posts der Öffentlichkeit zur Verfügung stellen.

2.3.3 Empfehlungssysteme

Empfehlungssysteme (Recommender) unterstützen den Nutzer bei der Beschreibung, Klassifikation und Verwaltung von Ressourcen. Häufig werden sie als Hilfe bei der Auswahl von Tags für eine neue Ressource eingesetzt oder weisen auf neue, unbekannte Ressourcen hin. Inhaltsbasierte Methoden suchen in der schon vorhandenen Menge an Ressourcen inhaltlich ähnliche Ressourcen. Dafür nutzt man inhaltliche Merkmale wie die Worte einer Webseite, die Farben bei Fotos oder die Tonhöhe eines Songs zur Beschreibung der Ressourcen. Findet man ähnliche Ressourcen, die schon Tags von diesem oder anderen Nutzern bekommen haben, so empfiehlt man diese Tags. Beim kollaborativen Filtern dagegen wird nach ähnlichen Nutzern im System gesucht. Haben z. B. zwei Nutzer die gleichen Ressourcen abgelegt, sind sich diese Nutzer sehr ähnlich. Liegen Bewertungen vor, können die Bewertungsprofile zusätzlich zu den gespeicherten Ressourcen der Nutzer verglichen werden. Wir betrachten Empfehlungssysteme genauer in Kap. 6.

2.3.4 Werbung

Die steigenden Nutzerzahlen von Social Bookmarking-Systemen macht die Schaltung von Werbung für Systembetreiber und Drittanbieter attraktiv. Neben Werbeanzeigen, die spezifisch für den Inhalt der jeweiligen Seite ausgesucht werden, kann man Werbeanzeigen auch auf die individuellen Bedürfnisse eines Nutzers anpassen. Dafür müssen Nutzerprofile gebildet werden. Wie bei den Empfehlungssystemen können neben den demographischen

³⁵ Z. B. im System BibSonomy für das Tag Recht die URL: <http://www.bibsonomy.org/tag/recht>.

Informationen auch die Inhalte der gespeicherten Ressourcen oder Tags genutzt werden. Für nicht angemeldete Nutzer ist es möglich, zumindest einen möglichen Herkunftsraum mit Hilfe der IP-Adresse einzugrenzen sowie das Navigationsverhalten innerhalb einer Sitzung zu verfolgen. Anhand der aufgerufenen Seiten sowie der Aufenthaltsdauer auf einer Seite lassen sich ebenfalls Interessen erkennen, die zu einem differenzierten Werbeangebot führen können.

2.3.5 Spam-Bekämpfung

Neben den positiven Effekten für soziale Interaktion, Suche und Navigation versuchen einige Nutzer die Sichtbarkeit Ihrer Posts im System für geschäftliche Zwecke auszunutzen. Sie fluten ein Social Bookmark-System mit Links auf z. B. Shopping-Seiten und versuchen so den Traffic auf ihren Seiten zu steigern. Diesem Missbrauch, der auch als Bookmarking-Spam bezeichnet wird, muss der Betreiber häufig aktiv entgegenwirken, z. B. durch ein Spamerkennungsframework.

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. In [59] wird die Vertrauenswürdigkeit der Anwender zum Ranken von Dokumenten genutzt und man ist so in der Lage, auch Spam-Attacken zu entdecken. Die effektive Spambekämpfung wird detaillierter in Kap. 7 in Social Bookmarking Systemen diskutiert. Dies ist noch ein offenes Feld, und es wird derzeit an dem Transfer bekannter Ansätze für die Entdeckung von Suchmaschinen- und E-Mail-Spam auf Social Bookmarking-Systeme gearbeitet.

2.4 Folksonomies

Social Bookmarking-Systeme bieten dem Nutzer die Möglichkeit, Bookmarks oder andere Ressourcen im System zusammen mit Schlagwörtern zu speichern. Diesen Vorgang nennt man auch Tagging, da die beschreibenden Schlagwörter – *Tags* genannt – zusammen mit der Ressource (d. h. der Referenz auf eine Webseite, eine Publikation, ein Bild, usw.) durch einen Nutzer im System abgelegt werden. Das Resultat ist ein *Post*. Abbildung 2.3 stellt diesen Vorgang schematisch dar. Im Zentrum steht der Post, den ein Nutzer beim Tagging erzeugt. Ein Post beschreibt immer eine Ressource durch einen oder mehrere Tags, die die Ressource möglichst gut aus Sicht des Nutzers charakterisieren. Mehrere Nutzer können die gleiche Ressource mit zum Teil unterschiedlichen Schlagwörtern versehen. Zentrale Bestandteile von Bookmarking-Systemen sind daher: Nutzer, Tags und Ressourcen, die über eine Relation miteinander in Beziehung stehen, siehe Abb. 2.3. Die Zuordnung von Tags zu Ressourcen durch den Nutzer wird auch als eine leichtgewichtige Wissensrepräsentation angesehen, die *Folksonomy* genannt wird. Leichtgewichtig deshalb,



Abb. 2.3 Eintrag eines Tagging-Systems am Beispiel eines Bookmarks

weil keine formale Semantik die Beziehung zwischen Tags, den Ressourcen und den Nutzern mittels einer Ontologie festlegt. Die Semantik der Tags ergibt sich implizit durch ihre Benutzung. Folksonomy ist ein Kunstwort aus Folks (Leute) und Taxonomy (Taxonomie). Die Folksonomy und ihre Elemente bilden den Kern eines jeden kooperativen Bookmarking-Systems.

Wie eben eingeführt besteht eine Folksonomy aus Mengen von Nutzern, Ressourcen und Tags sowie einer nutzerspezifischen 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 Nutzer, Tags bzw. Ressourcen nennt.

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

Die Personomy P_u eines Nutzers $u \in U$ ist die Beschränkung von F auf u , d. h. $P_u := (T_u, R_u, I_u)$, wobei, $I_u := \{(t, r) \in T \times R | (u, t, r) \in Y\}$, $T_u := \pi_T[I_u]$ und $R_u := \pi_R[I_u]$ gilt. Dabei bezeichnen π_T und π_R die Projektionen auf die Tag- bzw. die Ressourcendimension von I_u .

Diese Struktur kann man äquivalent als tripartiten (ungerichteten) Hypergraphen $G_F = (V, E)$ betrachten, wobei $V = U \cup T \cup R$ die Menge der Knoten und $E = \{\{u, t, r\} | (u, t, r) \in Y\}$ die Menge der Hyperkanten sind.

2.5 Eigenschaften von Folksonomies

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

In [15] wurde die Struktur der Folksonomies von del.icio.us und BibSonomy mit Graphanalyseansätzen untersucht. Es konnte gezeigt werden, dass der Graph Small-World-Eigenschaften [88] aufweist. 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.

Nutzer, Tags und Ressourcen folgen außerdem einer Long-Tail-Verteilung, d. h. es gibt wenige Nutzer mit vielen Einträgen und viele Nutzer mit wenigen Einträgen, und gleiches gilt für die Tags und die Ressourcen [47]. Auch Golder und Hubermann diskutieren die Struktur von Folksonomies (speziell die von del.icio.us) und identifizieren sieben verschiedene Arten von Tags [27]. 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“ [129, 131] – d. h. entstehende Semantik – stützt. Die Semantik ergibt sich dabei durch die Konvergenz des verwendeten Vokabulars – im Gegensatz zum ontologiebasierten Ansatz, der auf expliziten 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 [79].

Folksonomies können in diese Richtung erweitert werden, indem zusätzlich nutzerspezifische Tagrelationen in Anlehnung an Ober- und Unterkonzepte aus dem Semantic Web definiert werden (vgl. [47]). Sie erlauben eine nutzerbezogene Strukturierung der Tags, erhöhen die Übersichtlichkeit im System und erweitern die Anfragemöglichkeiten.

2.6 Vor- und Nachteile von Folksonomies

Ein großer Vorteil von Bookmarking-Systemen ist die einfache Benutzbarkeit. Es dauert nur ein paar Sekunden, um einen Post anzulegen, wobei die Tags ohne Einschränkung vom Nutzer gewählt werden können. Positiv fällt auf, dass nicht nur Webseiten, sondern im Prinzip jeder Ressourcentyp gepostet 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 Posts 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 verschiedene Weisen zur Verfügung – insbesondere auch über Programmierschnittstellen (API). Dadurch kann der Inhalt auf eine Weise, die von den Systemanbietern ursprünglich nicht geplant war, genutzt und mit anderen Inhalten in so genannten Mash-Ups verknüpft werden. Um einen Überblick über aktuelle und archivierte Nachrichten bzw. populäre Bookmarks zu bekommen, fasst man z. B. mit Yahoo Pipes³⁶ die RSS-Feeds verschiedenster Dienste übersichtlich zusammen. Häufig binden Blogger ihre Bookmarklisten per RSS-Feed aus dem Social Bookmarking-System im Blog ein und erhalten so automatisch aktuelle Link-Listen zu bestimmten Themen. Noch weiter geht IFTTT³⁷, indem es mit Hilfe von Bedingungen neue Ereignisse auslösen kann. Z. B. kann man prüfen lassen, ob man bei einem Dienst mit einem Foto verknüpft wurde um in solch einem Fall dann eine passende Mail zu erhalten.

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 Wörtern – 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 Bookmarking-System. Beispielsweise kann man für Europäische Union die folgenden Schreibweisen beobachten: Europäische Union, Europäische_Union, Europäische-Union und „Europäische Union“.

³⁶ <http://pipes.yahoo.com/pipes/>.

³⁷ <https://ifttt.com/wtf>.

Neben diesen sprachlichen Problemen besteht bei einer steigenden Nutzerzahl das Problem des Rankings von Posts beim Suchen in Bookmarking-Systemen. Häufig wird auch nach erweiterten Strukturierungsmöglichkeiten gefragt, z. B. das Zusammenfassen von Tags mittels Konzepten.

Social Bookmarking-Systeme sind ein wichtiger Bestandteil des Web 2.0 und werden von vielen genutzt. Das Verschlagworten von Bookmarks und anderen Ressourcen erlaubt deren einfache Organisation, hat aber auch Nachteile durch das Fehlen einer zentralen Organisation und Kontrolle. Im folgenden Kapitel werden die Grundlagen des Datenschutzes eingeführt, die die zweite Säule des Buches darstellen.

Informationelle Selbstbestimmung im Web 2.0

Chancen und Risiken sozialer

Verschlagwortungssysteme

Doerfel, S.; Hotho, A.; Kartal-Aydemir, A.; Roßnagel, A.;

Stumme, G.

2013, X, 189 S. 15 Abb., Hardcover

ISBN: 978-3-642-38055-6