

Kollaboratives Indexieren und die Emergenz neuer sozialer Netzwerke

1. Einführung

Soziale Software (Social Software) bezeichnet eine Klasse von vernetzten Anwendungen im World Wide Web (WWW, Web), die es ihren Benutzern ermöglichen, virtuelle soziale Netzwerke, also untereinander über das Internet vernetzte Gruppen und Gemeinschaften zu bilden (Social Networking). Der Nutzer einer Social Software Anwendung gibt durch eine zuvor erforderliche Registrierung dem System gegenüber persönliche Daten, Interessen und Neigungen preis. Soziale Gruppen können nun einerseits implizit über bestehende Gemeinsamkeiten zwischen verschiedenen Nutzern oder aber auch explizit durch die Auswahl „befreundeter“ Nutzer entstehen. Social Networking Anwendungen bieten dem Nutzer interaktive und produktive Partizipation am Medium Internet und haben in den vergangenen Jahren eine enorme Popularität erlangt. So erzielte etwa der Verkauf des im Jahr 2003 gegründeten Social Networking Services Myspace¹, der 2006 bereits mehr als 100 Millionen angemeldete Benutzer vorweisen konnte, an R. Murdochs News Corporation die Rekordsumme von 580 Millionen US-Dollar². Myspace finanziert sich über Werbung und ermöglicht dem Nutzer die kostenlose Einrichtung einer eigenen Webpräsenz. Damit zählt Myspace zu den unter der Bezeichnung Web 2.0 zusammengefassten Webanwendungen einer neuen Generation, die es ihren Nutzern ermöglichen, gemeinsam eigene Inhalte auf einfache Art und Weise im WWW zu publizieren.

Um sich in der neuen Web-Vielfalt besser zurecht zu finden, ergänzen sogenannte Kollaborativen Tagging Systeme (Social/Collaborative Tagging Systems, CTS) traditionelle Web-Suchmaschinen. War es zuvor der Autor selbst oder ein ausgewiesener Experte, der die Verschlagwortung (Tagging) einer Ressource (z.B. einer Webseite) übernahm, so legt ein CTS diese Aufgabe in die Hände der Benutzer. Jedermann kann eine beliebige Ressource – das Spektrum reicht dabei von klassischen (elektronischen) Dokumenten bis hin zu Konsumgütern aus der realen Welt - mit eigenen Etiketten (Schlagworten) versehen. Über diese können die damit etikettierten Ressourcen identifiziert und später wiedergefunden werden. Darüber hinaus können neue, bislang unbekannte interessante Ressourcen entdeckt werden, die von einem anderen Benutzer mit demselben Schlagwort versehen worden sind. Besonders interessant können dabei auch Ressourcen sein, die von ausgewiesenen „Freun-

¹ <http://www.myspace.com>

² http://www.intermix.com/about_press_inthenews.cfm?id=752&startrow=8

den“ verschlagwortet worden sind. Auf diese Weise können soziale Netzwerke mit in die persönliche Suche integriert werden. Andererseits entstehen durch die kollaborative Verschlagwortung selbst neue soziale Netze: zum einen sind elektronische Ressourcen untereinander vernetzt (Hyperlink) und eröffnen auf diese Weise den Weg zu Nutzern mit ähnlichen Interessen, zum anderen können diejenigen Nutzer zusammengefasst werden, die gleiche oder ähnliche Ressourcen verschlagwortet haben, insbesondere dann, wenn sich die von Ihnen verwendete Menge an Schlagworten überschneidet.

Im Zuge dieser Abhandlung soll zunächst der Begriff der Metadaten erläutert werden, der im Rahmen der Verschlagwortung eine zentrale Rolle einnimmt. Im Anschluss daran wird kurz die Entwicklung des WWW von den Anfängen bis hin zum zukünftigen Semantic Web aufgezeigt. Kollaboratives Indexieren und die Emergenz neuer sozialer Netze stehen darauf hin im Mittelpunkt der Betrachtung und führen zu dem Schluss, dass die kollaborative Erstellung von Metadaten einen Punkt von zentraler Bedeutung in der Entwicklung des zukünftigen Semantic Web darstellt.

2. Metadaten – Information über Information

Der Begriff Metadaten leitet sich aus dem Griechischen her und bedeutet soviel wie „Daten über Daten“, d.h. Metadaten beinhalten zusätzliche Informationen über Datenobjekte. In der bibliothekarischen Praxis in Form von Katalogen und Registern mit Autoren, Titeln oder Schlagworten schon seit Jahrhunderten in Gebrauch, erlangten Metadaten in den letzten Jahrzehnten im Rahmen der automatisierten Informationsverarbeitung eine besondere Bedeutung. „Metadaten sind strukturierte, kodierte Daten, die Charakteristika informationstragender Entitäten beschreiben, zum Zweck der Identifikation, Recherche, Beurteilung und der Verwaltung der damit beschriebenen Entitäten.“ (Durell,1985). Insbesondere steht heute der Aspekt der maschinellen Verarbeitung von Metadaten zum Zweck der Identifikation und der Recherche im Vordergrund.

Metadaten können grob klassifiziert werden nach dem Grad ihrer Strukturierung. Strukturierte Metadaten können z.B. durch einfache, vorgegebene Schemata (z.B. Autor, Titel, Verlag, ISBN, usw.) strukturiert werden. Zusätzlich lassen sich Kategorien und einfache Ober- und Unterklassenbeziehungen zwischen den Kategorien über Taxonomien ausdrücken. Komplexere Strukturen, die etwa auch unterschiedliche Beziehungen, Abhängigkeiten und Regeln zwischen den mit Metadaten ausgezeichneten Objekten zulassen, können mit Hilfe von Ontologien formuliert werden. Gänzlich ohne jede Struktur dagegen ist eine Menge von zugeordneten Schlagworten, die selbst keinerlei Regeln unterworfen sind (Davis, Shrobe & Szolovitis, 1993).

Das Auszeichnen von informationstragenden Objekten mit Metadaten ist eine nicht-triviale Aufgabe, die in der Regel über den Urheber (Autor) des informationstragenden Objektes selbst oder durch einen ausgewiesenen Experten zentral erfolgt. Die Angabe von Autor, Titel und anderen Publikationsangaben ist in der Regel sehr

einfach im Vergleich zur Verschlagwortung, die entsprechende Kenntnis des Inhalts des damit ausgezeichneten Objekts voraussetzt.

Für die Informationsrecherche im WWW sind Metadaten von essentieller Bedeutung. Die Semantik, d.h. die Bedeutung des Inhalts von Web-Dokumenten lässt sich aus diesen automatisch nur unzureichend gewinnen. Methoden des Information Retrieval und der Verarbeitung natürlicher Sprache (Natural Language Processing) sind sehr aufwändig und nur innerhalb eines begrenzten Kontextes erfolgreich. Noch schwieriger gestaltet sich die Erschließung der Semantik aus multimedialen Dokumenten wie z.B. Bildern oder Videos. Hier sind Metadaten, die Auskunft über den Inhalt dieser Dokumente liefern, unverzichtbar (Sack & Waitelonis, 2006). Das WWW von heute ist jedoch zur manuellen Nutzung durch den Menschen bestimmt. Maschinelle automatische Recherche von zuverlässigen Informationen ist für den Großteil der im WWW repräsentierten Dokumente nicht möglich, da diesen eine entsprechende Auszeichnung mit Metadaten fehlt. Zudem sieht die derzeitige Standardsprache des WWW, die Hypertext Markup Language (HTML), nur rudimentäre und nicht allgemein standardisierte Möglichkeiten vor, Metadaten direkt in HTML-Dokumente zu integrieren. So wird z.B. das zu diesem Zweck vorgesehene HTML <meta>-Tag von Suchmaschinen nur in begrenztem Umfang genutzt, da die Möglichkeit der Manipulation der Suchmaschine durch den Autor selbst nicht ausgeschlossen werden kann.

3. Das WWW verändert sich – Web 1.0, Web 2.0 und Semantic Web

Seit seiner Geburtsstunde im Jahre 1990 hat sich das WWW sehr verändert. Zunächst nur ein über Hyperlink vernetztes Dokumentenmanagementsystem, das lediglich einer kleinen Zahl von Nutzern zur Verfügung stand, sollte es sich in den folgenden Jahren inflationär zum größten verteilten Informationssystem aller Zeiten entwickeln. Grund dafür war zunächst die freie Verfügbarkeit – Implementierungen für Webserver und Webbrowser als Client-Server-basierte Anwendungen sind kostenfrei erhältlich – sowie die Entwicklung eines Webbrowsers mit einfacher grafischer Benutzerschnittstelle, die es auch einem Nichtfachmann ermöglichte, auf die Ressourcen des WWW zuzugreifen. Ab 1995 entdeckte der eCommerce, der elektronische Handel das WWW: Amazon³ und Ebay⁴ wurden gegründet. Mit dem Aufkommen des eCommerce verschob sich der Fokus des WWW weg von dem persönlichen Kommunikations- und Publikationsmedium für Spezialisten, hin zu einem Medium der Massenkommunikation: Informationsproduktion und Informationskonsumption bleiben strikt getrennt. Lediglich der Spezialist vermag es, eigene Inhalte im WWW online zustellen. Die breite Masse konsumiert das Informationsangebot der gleich einem traditionellen Broadcast-Medium agierenden kommerziellen In-

³ <http://www.amazon.com/>

⁴ <http://www.ebay.com/>

formationsanbieter. Die Interaktion der Nutzer beschränkt sich lediglich auf das Lesen von Webseiten, das Online-Bestellen von Waren und Konsumgütern, sowie das Anklicken von Werbebanner.

Doch das WWW verändert sich weiter. Neue Technologien erscheinen, die es auch dem Laien ermöglichen, Informationsinhalte auf einfache Art und Weise zu publizieren. Weblogs, Chatrooms, Tauschbörsen und Wikis erobern das WWW und eröffnen dem Anwender auf breiter Basis den Weg zu einer echten Interaktion und Partizipation. Als der Medienunternehmer und Netzpionier Tim O'Reilly im Oktober 2004 diese Veränderung des WWW einem reinen Fachpublikum unter dem Namen Web 2.0 vorstellte, war noch nicht abzusehen, dass diese Renaissance des WWW derart um sich greifen würde (O'Reilly, 2005). Das Internet hatte sich von einem reinen Broadcast-Medium zu einem echten Marktplatz gewandelt, der Nutzer ist Informationskonsument und Informationsproduzent zugleich. Diese neue Interaktivität ermöglicht direkt und indirekt die Entstehung sozialer Netzwerke. Es entstanden spezielle webbasierte Systeme, deren Zweck darin besteht, Verbindungen zwischen Nutzern – zu geschäftlichen oder zu privaten Zwecken - zu knüpfen (z.B. openbc⁵). Nutzer eines solchen Systems müssen bei der Anmeldung eine Reihe persönlicher Daten, Interessen und Neigungen hinterlegen. Einerseits kann ein Nutzer einen anderen explizit als Freund markieren, andererseits können Beziehungen auch durch identische Interessenprofile und Gemeinsamkeiten entstehen. Ein Mehrwert dieser so sozialen Netzwerke entsteht mit einer wachsenden Anzahl von Nutzern mit qualitativ guten Kontakten.

Die Vielzahl der neuen Anwendungen des Web 2.0 macht es jedoch nicht leichter, sich in den Milliarden von Webdokumenten zu orientieren. Suchmaschinen wie Google⁶ leiten heute den Nutzer durch das Informationsuniversum WWW. Google verwaltet dazu einen gigantischen Index, der auf die Eingabe eines Suchbegriffes hin einen schnellen Zugriff auf diesbezüglich relevante Webdokumente bietet. Doch stellt sich auch hier das Problem der Verschlagwortung. Die schiere Größe des WWW lässt hier nur automatische Verfahren zum Einsatz kommen, die nach statistischen Verfahren alle Begriffe innerhalb eines Webdokuments auswerten, indizieren und die jeweiligen Dokumente so in eine Rangreihenfolge bringen. Doch bereits die Menge an zurückgelieferter Information ist für den Anwender nicht mehr nachvollziehbar. So liefert eine Anfrage nach dem Suchbegriff „Web 2.0“ mehr als 53 Millionen Ergebnisse (Stand 09/2006). Allerdings finden sich in den Ergebnislisten nur Dokumente, die diesen Begriff auch wortwörtlich enthalten. Umschreibungen und Synonyme können nicht gefunden werden. Ebenso muss ein Dokument, das den gesuchten Begriff enthält, nicht notwendigerweise auch thematisch diesen Begriff in den Mittelpunkt stellen. Vollständigkeit und Genauigkeit der Suchergebnisse können daher alleine aufgrund der problematischen Interpretation natürlicher Sprache nie zu 100 Prozent garantiert werden.

⁵ <http://www.openbc.com/>

⁶ <http://www.google.com/>

An dieser Stelle wäre eine Auszeichnung der betreffenden Webdokumente mit entsprechend aussagekräftigen Metadaten notwendig. Eine derart ausgezeichnete Webseite würde hinter jedem für dieses Dokument relevanten Begriff einen Verweis auf ein diesen Begriff beschreibendes Konzept enthalten. Diese Konzepte können in einer maschinenlesbaren, standardisierten Form als Ontologie hinterlegt werden, die von einer Suchmaschine zusätzlich ausgewertet werden könnten, um so die Trefferquote der präsentierten Suchergebnisse zu erhöhen. Das für die Standardisierung des WWW zuständige WWW-Consortium (W3C⁷) hat bereits die dazu notwendigen Grundlagen in Form von Ontologiebeschreibungssprachen, wie z.B. RDF, RDFS oder OWL geschaffen. Semantisch annotierte Webseiten ermöglichen es autonom agierenden Agenten zielgerichtet Informationen zu sammeln, um selbstständig Entscheidungen im Sinne ihres Auftraggebers zu treffen und Transaktionen über das WWW zu initiieren. Dieses semantische Netzwerk (Semantic Web) stellt die nächste Evolutionsstufe des WWW dar und soll in naher Zukunft Realität werden (Berners-Lee, Hendler & Lassila, 2001).

4. Kollaborative Indexierung

Wie bereits dargestellt, stellt die Verschlagwortung (Indexierung, Tagging) von Dokumenten bereits eine Form der Auszeichnung mit Metadaten dar. Unter Verschlagwortung versteht man im Allgemeinen die Zuweisung von inhaltlich stimmigen und aussagekräftigen Begriffen zu einem Dokument. Dabei können einem Dokument mehrere Schlagworte zugewiesen werden. Im Gegensatz dazu steht die Kategorisierung, d.h. das Einordnen eines Dokuments innerhalb einer bestimmten Begriffshierarchie. In Bibliotheken kommt zu diesem Zweck oft die Universale Dezimalklassifikation (Mcilwayne, 2000), die aus der Dewey-Dezimalklassifikation (Dewey, 1876) entstanden ist und beständig erweitert wird zum Einsatz. Der Unterschied zur Verschlagwortung besteht darin, dass das Dokument einer einzelnen Kategorie zugewiesen werden muss. Das Wiederauffinden eines Dokuments innerhalb des Klassifizierungsschemas wird dadurch erschwert, dass der Klassifizierende und der in der Klassifikation Suchende oft eine unterschiedliche Einschätzung und einen unterschiedlichen Blickwinkel auf das gesuchte Dokument besitzen.

Webbasierte Systeme zur kollaborativen Indexierung (Collaborative Tagging Systems, CTS) bieten die Möglichkeit der gemeinsamen Verschlagwortung von Dokumenten und Ressourcen durch alle Nutzer. Die Aufgabe der Verschlagwortung geht dadurch von einer zentralen Instanz an die Gemeinschaft der Anwender über und erlaubt so auch die Vergabe von Schlagworten rein nach persönlichen Interessen und Neigungen. Man unterscheidet unterschiedliche Varianten von CTS:

CTS, die vordergründig der Verschlagwortung eigener Ressourcen dienen, mit dem Zweck, diese damit zu organisieren und wieder zu finden (eingeschränkte

⁷ <http://www.w3c.org/>

CTS). Die Bilderdatenbank flickr⁸ ist hierzu ein prominentes Beispiel, das es dem Anwender gestattet, eigene Bilddateien in das Webbasierte System zu überspielen und diesen selbst definierte Schlagworte zuzuweisen.

Ein anderes prominentes Beispiel ist das Bookmarkverwaltungssystem del.icio.us⁹. Im Gegensatz zu flickr werden hier fremde Objekte (Webdokumente) mit eigenen Schlagworten ebenfalls zum Zweck der Organisation und der Recherche versehen (offene CTS). Abb. 1 zeigt die vereinfachte, schematische Darstellung eines offenen CTS als Hypergraph $G = (R \cup T \cup U, E)$, wobei $R \cup T \cup U$ die Menge der Knoten darstellt mit $R = (r_1, \dots, r_l)$ als Menge der verschlagworteten Ressourcen, $T = (t_1, \dots, t_m)$ als Menge der vergebenen Schlagworte, $U = (u_1, \dots, u_n)$ als Menge der Nutzer und $E \subseteq R \times T \times U$ die Menge der Kanten. Betrachtet man zusätzlich Verbindungen zwischen den verschlagworteten Ressourcen selbst (also z.B. Hyperlinks zwischen Webdokumenten), $E_R \subseteq R \times R$, und Beziehungen zwischen den Benutzern (z.B. das soziale Netzwerk), $E_U \subseteq U \times U$, erhält man den erweiterten Hypergraphen $G' = (N, E \cup E_U \cup E_R)$.

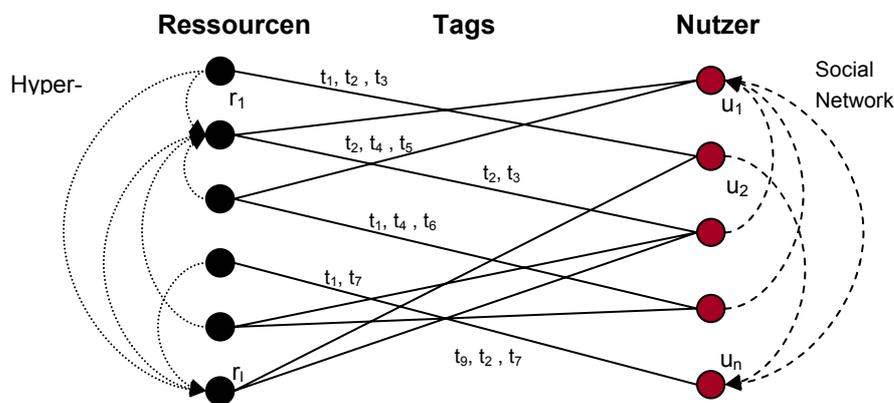


Abb. 1: Schematische Darstellung eines Collaborative Tagging Systems als erweiterter Hypergraph G' .

Im Gegensatz zu traditionellen Suchmaschinen bieten CTS einen anderen Zugang zu den Inhalten des WWW. Zunächst ist es selbstverständlich möglich, Objekte, die man selbst mit Schlagworten ausgezeichnet hat, über eine Suchabfrage, die ebendiese Schlagworte enthält, wieder zu finden. Andererseits bieten offene CTS zusätzlich die Möglichkeit, weitere Ressourcen aufzuspüren, die von anderen Benutzern mit dem eingegebenen Suchbegriff verschlagwortet wurden. Diese Form der Suche wird auch als Tag-Browsing bezeichnet und schafft neue Querverweise zwischen Res-

⁸ <http://www.flickr.com/>

⁹ <http://del.icio.us/>

sourcen, die alleine durch die Verschlagwortung der beteiligten Nutzer erzeugt wurde.

Betrachtet man die Menge aller Tags T_{r_k} , die einer bestimmten Ressource r_k von allen Nutzern zugewiesen wurden, also $T_{r_k} = \{t_i | t_i \in T \wedge (t_i, r_k, u_i) \in E\}$, wobei $r_k \in R, u_i \in U$, erhält man eine Menge mit allen Schlagworten, die einer Ressource zugewiesen wurden. Diese Menge wird auch als Folksonomy bezeichnet¹⁰ und kann als unstrukturierte, flache Ontologie aufgefasst werden, die den Inhalt eines Dokuments beschreibt. Ein populäres Mittel zur Visualisierung von Folksonomies sind so genannte Tag Clouds. Eine Tag Cloud stellt alle einer Ressource zugeordneten Schlagworte grafisch dar, wobei die jeweils für ein Schlagwort gewählte Schriftgröße der Häufigkeit entspricht, mit der dieses Schlagwort von allen Nutzern dieser Ressource zugewiesen wurde. Neben der absoluten Häufigkeit können aber auch andere Kenngrößen, wie z.B. Beziehungen aus sozialen Netzwerken zur Bewertung fremd vergebener Schlagworte in die Gewichtung mit eingehen.

Folksonomies stellen jedoch nicht in allen Fällen eine objektive Auszeichnung einer Ressource mit Metadaten dar. Man unterscheidet deskriptive Schlagworte, also Schlagworte, die eine Ressource unter objektiven Gesichtspunkten beschreiben, und funktionale Schlagworte, also Schlagworte, die für den Schlagwortautor eine bestimmte, in der Regel nur ihn selbst betreffende Funktionalität beinhalten. So kann etwa ein Buch über den Philosophen Platon mit den deskriptiven Schlagworten „Philosophie“, „Platon“ oder „Antike“ versehen werden, während gleichzeitig funktionale Schlagworte wie „klausurrelevant“ oder „Prüfungsstoff“ in der zugehörigen Folksonomy enthalten sein können. Während deskriptive Schlagworte einen Nutzwert für alle Anwender besitzen, bergen funktionale Schlagworte den größten Nutzen jeweils für die Schlagwortautoren selbst. Es konnte jedoch nachgewiesen werden, dass die Verteilung der Schlagworte für eine bestimmte Ressource einem Potenzgesetz entspricht, d.h. mit steigender Anzahl von Nutzern konvergieren die für diese Ressource vergebenen Schlagworte (Golder & Huberman, 2006).

5. Die Emergenz neuer sozialer Netze

Verbindet man die Vorteile von Social Software, also der Bildung sozialer Netzwerke mit CTS, bieten neue Varianten des Tag-Browsing die Möglichkeit der Entdeckung bislang unbekannter, interessanter und relevanter Ressourcen. Die einfachste Möglichkeit bietet dabei die Ausnutzung des „direkten“ sozialen Netzwerks, d.h. diejenigen Ressourcen werden durchsucht, die von Nutzern verschlagwortet wurden, die entweder direkt als Freund gekennzeichnet wurden bzw. deren ähnliches Nutzerprofil diese indirekt als Freund mit ähnlicher Interessenlage auszeichnet. Eine Schlagwortsuche innerhalb dieser eingeschränkten Ressourcenmenge erhöht potenziell die Relevanz der damit erzielten Suchergebnisse.

¹⁰ <http://www.vanderwal.net/random/entrysel.php?blog=1622>

Ein neues „indirektes“ soziales Netzwerk entsteht durch die Betrachtung von Nutzern, die Ressourcen mit identischen – oder etwas abgeschwächt mit bedeutungsähnlichen – Schlagworten auszeichnen. Die Verwendung bedeutungsähnlicher Schlagworte lässt ebenfalls auf einen gewissen Grad an Übereinstimmung der jeweiligen Interessenlage schließen. Auf diese Weise können also Nutzer gefunden werden, die Ähnlichkeiten zu einem gegebenen Nutzer aufweisen und daher als Freund gekennzeichnet werden können.

Abb. 2 zeigt das durch das Tag $t_1 \in T$ induzierte soziale Netzwerk der Nutzer $U_{t_1} = \{u_i | u \in U \wedge (u_i, t_1, r_k) \in E\}$ mit $r_k \in R$ im Graphen G' .

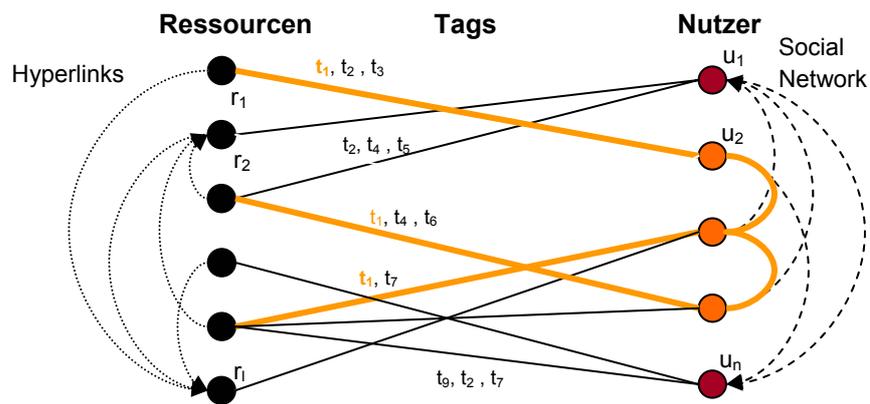


Abb. 2: Graph des durch das Schlagwort $t_1 \in T$ induzierten sozialen Netzwerks

Nicht nur die Verwendung bedeutungsähnlicher Schlagworte, sondern auch die Verschlagwortung identischer Ressourcen kann als Hinweis darauf gewertet werden, dass die betreffenden Nutzer ähnliche Interessen haben und daher ebenfalls als Freunde gekennzeichnet werden können. Interessant sind dabei jeweils durch Freunde ausgezeichnete Ressourcen, die nicht selbst verschlagwortet wurden. Diese Ressourcen sind bislang unentdeckt, aber mit einiger Wahrscheinlichkeit ebenfalls von Interesse. Abb. 3 zeigt das durch die Ressource $r_k \in R$ induzierte soziale Netzwerk der Nutzer $U_{r_k} = \{u_i | u \in U \wedge (u_i, t_m, r_k) \in E\}$ mit $t_m \in T$ im Graphen G' .

Zieht man die z.B. bei Webdokumenten vorhandene Hyperlinkstruktur, die die verschlagworteten Ressourcen untereinander vernetzt, in den Beziehungsgraphen mit ein, entsteht ein indirekt erweitertes soziales Netzwerk. Geht man davon aus, dass miteinander über ein Hyperlink verbundene Ressourcen in einer inhaltlichen Beziehung stehen, d.h. betrachtet man diese als ähnlich, kann man über den Graphen G' diejenigen Nutzer ermitteln, die diese verlinkten Ressourcen selbst verschlagwortet haben. Die auf diese Weise ermittelte Nutzergruppe bildet ein neues soziales Netzwerk, das in Abb. 4 dargestellt ist.

Die mit der Ressource $r_k \in R$ verlinkten Ressourcen R' wobei für alle $r_m \in R'$ gilt $(r_k, r_m) \in E_R$, induzieren ein soziales Netzwerk mit den Nutzern $U' = U_{r_k} \cup U_{R'}$, wobei $U_{R'} = \{u_i | u \in U \wedge (u_i, t_k, r_m) \in E\}$ mit $t_k \in T$.

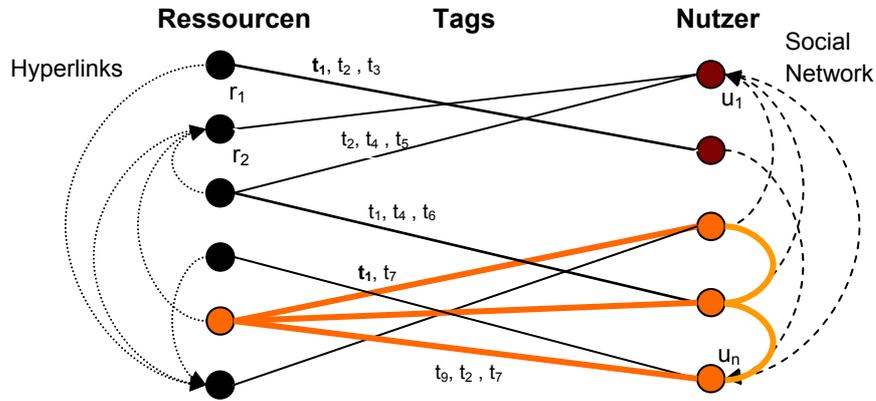


Abb. 3: Graph des durch die Ressource $r_k \in R$ induzierten sozialen Netzwerks

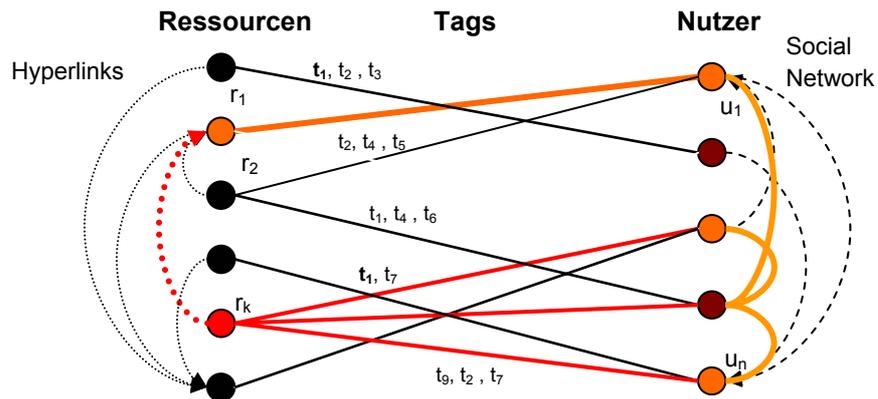


Abb. 4: Graph des durch die mit Ressource $r_k \in R$ verlinkten Ressourcen R' induzierten sozialen Netzwerks

Ausgehend von der gemeinsamen Verwendung von Schlagworten, dem gemeinsamen Verschlagworten von gleichen Ressourcen sowie die Einbeziehung der Vernetzung der Ressourcen untereinander lassen sich im Verschlagwortungsgraphen Gruppen von Nutzern identifizieren, die untereinander Gemeinsamkeiten aufweisen, die sich wiederum für die Recherche von Ressourcen ausnutzen lassen. Dabei können einzelne Verbindungen jeweils auch unterschiedliche Gewichtung erfahren oder auch transitive Verbindungswege verfolgt werden, um das identifizierte soziale Netzwerk noch zu erweitern.

6. Zusammenfassung und Ausblick

Web 2.0 bezeichnet eine neue Generation von WWW-Anwendungen, die den Nutzer in zuvor ungeahnter Weise an der Ausgestaltung des WWW gemeinsam teilhaben lassen. Kollaborative Indexierung als eine der Schlüsselanwendungen im neuen Web 2.0 ermöglicht die gemeinschaftliche Verschlagwortung von Webdokumenten mit deskriptiven und funktionalen Metadaten. Dabei ermöglicht diese Kollaboration die Entstehung neuer sozialer Netzwerke, die ihrerseits dem Nutzer neue Recherchemöglichkeiten im WWW eröffnen.

Das zukünftige Semantic Web kann nur Realität werden, wenn es gelingt, eine ausreichende Zahl von Webdokumenten hinreichend semantisch zu annotieren, so dass automatisch auf die Bedeutung der annotierten Dokumente geschlossen werden kann. Eine zentral gesteuerte, automatische Annotation der vielen Milliarden bereits existierenden Webseiten ist ebenso wie eine manuelle, von menschlichen Experten angeleitete Annotation nicht realistisch. Collaborative Tagging Systeme zeigen einen Weg der gemeinschaftlichen Annotation von Webdokumenten auf. Auch wenn diese Annotation bislang nur Schlagworte, also unstrukturierte, flache Ontologien liefert, zeigt das Prinzip doch, dass auch komplexere Formen der kollaborativen Annotation möglich sind. Dies lässt sich jedoch nur dann auf breiter Basis durchführen, wenn es gelingt, zu diesem Zweck eine intuitive, leicht zu bedienende Benutzerschnittstelle zu schaffen, die es auch Laien ohne tiefere Kenntnisse auf dem Gebiet der Wissensrepräsentation ermöglicht, valide semantische Annotationen zu erstellen.

7. Literaturverzeichnis

- T. Berners-Lee, J. Hendler & O. Lassila (2001): The Semantic Web. In: American Scientific, 284 (5), pp.34-43, 2001.
- R. Davis, H. Shrobe & P. Szolovits (1993): What is a Knowledge Representation. In: AI Magazine, 14 (1), pp.17-33, 1993.
- M. Dewey (1876): A Classification and Subject Index for Cataloguing and Arranging the Books and Pamphlets of a Library. Amherst, Mass (USA), 1876.
- W. R. Durrell (1985): Data Administration: A Practical Guide to Data Administration, McGraw-Hill, 1985.
- S. Golder & B. A. Huberman (2006): Usage Patterns of Collaborative Tagging Systems. In: Journal of Information Science, 32 (2), pp.198-208, 2006.
- I.C. Meilwaine (2000): The Universal Decimal Classification: guide to its use. The Hague: UDC Consortium, 2000.
- T. O'Reilly (2005): What is Web 2.0? Design Patterns and Business Modells for the Next Generation of Software. 30/09/2005, Online verfügbar:
<http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html?page=1>.
(29.09.2006)
- H. Sack & J. Waitelonis (2006): Automated Annotations of Synchronized Multimedia Presentations. In Proc. of the ESWC 2006 Workshop on Mastering the Gap: From Information Extraction to Semantic Representation, CEUR Workshop Proceedings, 2006.