

Implizite Digitale Soziale Netze als Basis für Expertenempfehlungssysteme

Tamara Heck, Isabella Peters

Abteilung für Informationswissenschaft
Heinrich-Heine-Universität Düsseldorf
Universitätsstr. 1
40225 Düsseldorf
{tamara.heck | isabella.peters}@uni-duesseldorf.de

Abstract: Empfehlungssysteme haben sich insbesondere im e-Commerce etabliert, aber die Empfehlung von Experten oder Mitarbeitern in einem firmeninternen Netzwerk oder in wissenschaftlichen Disziplinen wird derzeit noch theoretisch diskutiert. Wir präsentieren einen Ansatz zur Entwicklung von Expertenempfehlungssystemen, der auf Beziehungen in digitalen sozialen Netzwerken wie Social Bookmarking-Systemen mit ihren Folksonomies beruht.

1 Soziale Netzwerke und Social Bookmarking-Systeme (SBS)

Das Web 2.0 bietet eine Vielzahl neuer Möglichkeiten der Kommunikation und Kollaboration. Dabei erfahren insbesondere diese Plattformen einen Boom, die den Nutzern die Bildung von Sozialen Netzwerken ermöglichen. Dabei kann zwischen zwei Arten von Sozialen Netzwerken unterschieden werden: 1) solche, die die Beziehung zwischen den Nutzern explizit machen (z.B. Facebook) und 2) solche, die vorrangig dem Zweck des Aufbaus einer Wissensbasis dienen und hier die Nutzer nur implizit miteinander in Beziehung setzen (z.B. SBS).

SBS wie zum Beispiel BibSonomy, CiteULike oder Connotea erlauben es dem Nutzer, die Favoritenliste des Web-Browsers desktop-unabhängig im World Wide Web abzulegen und damit gefundene Ressourcen von jedem Browser aus wieder zugreifbar und verwaltbar zu machen [Go06]. Beschrieben werden die Bookmarks mit sog. „Tags“, frei wählbaren Schlagworten, die in der Summe eine Folksonomy [Pe09] ergeben. Die Sammlung an Favoriten kann für andere Nutzer des SBS öffentlich gemacht werden. So kann die gesamte Community an den Organisations- und Katalogisierungsaktivitäten des einzelnen Nutzers teilhaben und davon profitieren. Damit wandelt sich das SBS von einem persönlichen Verwaltungsprogramm für URLs zu einem Informationsdienst, der die Grundzüge der Sozialen Netzwerke ausnutzt und der Gemeinschaft dadurch eine verbesserte Informationsbeschaffung ermöglicht. Denn es fungiert nun nicht mehr nur als Ablage, sondern ebenso als Vermittler für interessante oder relevante Web-Ressourcen. Daher haben SBS auch bereits Einzug in unternehmensinterne [MFK06] und wissenschaftliche [BHD09] Kontexte gehalten.

In SBS können Beziehungen zwischen Nutzern ebenfalls auf zweifache Weise erkannt werden. Zum einen können Nutzergruppen explizit über die Verwendung eines „Gruppentags“ (wie z.B. „Wismasys0809“), den sie zur gemeinsamen Literaturverwaltung einführen, sichtbar werden. Zum anderen können Nutzer über ihr Bookmarking- und Taggingverhalten einem sozialen Netz angehören, welches inhärent in der entstandenen Folksonomy (über gemeinsame Bookmarks und gemeinsame Tags) enthalten ist. Dieses implizite soziale digitale Netzwerk lässt sich über die Folksonomy freilegen und beispielsweise für Personenempfehlungssysteme ausnutzen.

Im Folgenden werden Ergebnisse eines laufenden Forschungsprojektes beschrieben, welches ein Expertenempfehlungssystem basierend auf SBS entwickelt. Ziel des Projektes ist es, Mitarbeitern einer Forschungseinrichtung geeignete Personen ihre impliziten sozialen Netzwerke im SBS zur Kommunikation und zum Wissensaustausch vorzuschlagen und das System qualitativ zu evaluieren. Dazu arbeiten wir mit dem Forschungszentrum Jülich (FZJ) zusammen.

2 Folksonomy-basierte Expertenempfehlungssysteme

Empfehlungssysteme finden ihre Grundlage im Kollaborativen Filtern [KH07]. Als Kollaboratives Filtern wird das Einschränken einer Informationsmenge mit Hilfe einer Gruppe von Nutzern bezeichnet: „Collaborative filtering simply means that people collaborate to help one another perform filtering by recording their reactions to documents they read“ [Go92]. Das System ist daher in der Lage, dem Nutzer bei der Informationsbeschaffung zu unterstützen. [KSS97, 65] sprechen einen grundlegenden Vorteil von Kollaborativen Filtersystemen an: „Typically a user is only aware of a portion of the social network to which he or she belongs. By instantiating the larger community, the user can discover connections to people and information that would otherwise lay hidden over the horizon“. Der im SBS hergestellte tri-partite Zusammenhang von Tags, Bookmarks und Nutzern wird in folksonomy-basierten Empfehlungssystemen so ausgenutzt, dass Informationsströme auf Basis verschiedener Profile gefiltert werden, die aus der bibliographischen Kopplung [Ke63] von Nutzern (d.h. Speicherung gleicher Bookmarks) oder der Übereinstimmung von Tags gewonnen werden. Zu beachten ist, dass es sich um eine „Broad Folksonomy“ [Va05] handeln muss, d.h. dass ein gleicher Tag zu einer Ressource mehrmals von verschiedenen Nutzern vergeben werden darf. So kann auf Basis der gleichen genutzten Tags eine Ähnlichkeit zwischen Nutzern hergestellt werden. Die zugrundeliegende Annahme ist hierbei: „Übereinstimmung bedeutet Ähnlichkeit“. Genutzt werden Kollaborative Vorschlagssysteme oftmals bei Einkaufsplattformen im Internet, u.a. bei dem Online-Versandhaus Amazon.

Die Tags nehmen in Folksonomies drei Rollen ein:

- 1) sie fungieren als Links bzw. Verbindungen zwischen Nutzern und Dokumenten,
- 2) sie beschreiben das Dokument inhaltlich, formal oder persönlich und
- 3) sie dienen dem (Wieder-)auffinden von Dokumenten.

Folksonomy-basierte Empfehlungssysteme arbeiten mit genau diesen Informationen und können so grundsätzlich zwei Arten von Ressourcen vorschlagen: „Recommendation systems may, in general, suggest either people or items” [JS06]. [DI06] fassen zusammen, welche Vorschläge auf Basis von Nutzerprofilen und Tags möglich sind:

1. *Objects based on users*: Ressourcen basierend auf anderen Nutzern.
2. *Users based on objects*: Nutzer basierend auf anderen Ressourcen.
3. *Users based on co-tagging*: Nutzer basierend auf gleich getaggtten Ressourcen.
4. *Tags based on users*: Tagempfehlungen basierend auf gleichen Nutzern.
5. *Users based on tags*: Nutzer basierend auf gleichen Tags.

Die vorgeschlagenen Ressourcen ergeben sich dabei durch Ähnlichkeitsberechnungen und clusterbildende Verfahren [SM95].

3 Entwicklung eines Experten Recommender Systems

Kollaborative Informationsdienste wie SBS sind mittels Folksonomies in der Lage, über die ihnen inhärente Netzwerkeigenschaft ähnliche Nutzer sichtbar und dadurch den Informations- und Wissensaustausch effektiver zu machen. Das Motto der Informationsbeschaffung im Web 2.0 lautet dabei: „More like me!“ – finde Nutzer, die mir ähnlich sind, damit ich relevante Informationen von ihnen anschauen kann [Sm08]. Man kann davon ausgehen, dass sich Nutzer dann ähneln oder ähnliche Interessen haben, wenn sie die gleichen Tags zur Indexierung von Ressourcen benutzen („thematische Verbindung“) bzw. durch die gleichen Relationen mit ihnen verbunden sind oder wenn sie die gleichen Ressourcen indexieren, speichern oder bearbeiten („bibliographische Kopplung“). [FPB06] stellen dazu fest: „[Recommender systems] exploit the fact that people with similar tagging behaviour – also known as tag buddies – have related taste“. Bisher gibt es nur sehr wenige empirische Studien zum Einsatz von Vorschlagsystemen bei Folksonomies und ihrem Nutzen während des Information Retrievals und im Wissensmanagement. Erwähnenswert ist die Untersuchung von [Jä07], die die Wirkung von vorgeschlagenen Such-Tags auf die Retrievalleistung (anhand von BibSonomy und dem Musikdienst Last.fm) untersuchen.

3.1 Methode

Zunächst soll das Expertenempfehlungssystem basierend auf dem SBS CiteULike aufgebaut werden. Um das Empfehlungssystem logisch einzugrenzen und eine informetrische Untersuchung der Nutzer durchzuführen, haben wir mit Hilfe der Kollegen des FZJ 45 Journals aus dem Bereich Festkörperphysik ausgewählt und deren Bookmarks (eingestellt in den Jahren 2004 bis 2008) sowie die damit vergebenen Tags und die Nutzer aus dem SBS gefiltert. Die Eingrenzung der Zeitschriften erfolgte, weil die FZJ-Mitarbeiter das entwickelte Tool später evaluieren und die Nützlichkeit eines solchen Systems für sich und ihre Kommunikation untereinander überprüfen sollen. Der erste Schritt bei Empfehlungen von Nutzern besteht darin, mittels eines Koeffizienten (z.B. Dice, Cosinus oder Jaccard-Sneath) die Ähnlichkeiten zwischen dem Ausgangsnutzer und allen anderen Nutzern des SBS zu berechnen. Die Berechnung muss dabei zweifach durchgeführt werden, um zum einen der thematischen Verbundenheit der Nutzer und zum ande-

ren ihrer bibliographischen Kopplung gerecht zu werden. Als Beispiel haben wir zuerst den Dice-Koeffizienten für die Berechnung genommen, wobei a die Anzahl der Tags (oder der Bookmarks) ist, die Nutzer D_i vergeben hat, b die Anzahl der Tags (oder Bookmarks), die Nutzer D_j zum Indexieren heranzog und g die Anzahl derjenigen Tags (oder Bookmarks), die beide Nutzer verwendet haben:

$$S_{D_i, D_j} = \frac{2g}{a+b}$$

Beide ermittelten Werte können nun entweder zusammengefasst oder einzeln betrachtet werden. Der zweite Schritt (folgt man der elaborierteren Variante und vernachlässigt z.B. k-nearest neighbors-Verfahren) umfasst nun die Bildung von Clustern mittels Single-Link-, Complete-Link- oder Group-Average-Link-Verfahren [KSS10]. Die so ermittelten ähnlichsten Nutzer können dann einander vorgeschlagen werden.

3.2 Ergebnisse

Wir haben zunächst Ähnlichkeitswerte zwischen den Nutzern im SBS CiteULike ermittelt. Die Daten der 45 Physik-Zeitschriften wurden anhand der ISSN, des DOI, den Journal-Titeln und deren Abkürzungen gefiltert. Dabei vergaben 1.006 Nutzer 2.861 Bookmarks. Deren Ähnlichkeit wurde mit dem Dice-Koeffizienten zum einem basierend auf Bookmarks und zum anderen basierend auf Tags errechnet. Beide Werte wurden für die Ähnlichkeitsberechnung und die spätere Cluster-Bildung getrennt betrachtet, um Unterschiede bei den errechneten Nutzervorschlägen erkennen zu können. Nutzer, die nur einen Bookmark im SBS gespeichert haben, wurden aus den Berechnungen ausgeschlossen, da sie bei der Ähnlichkeitsberechnung ungerechtfertigterweise das höchste Ergebnis erzielen würden (nämlich einen Ähnlichkeitswert von 1 wenn beide Nutzer jeweils nur den einen Bookmark gespeichert haben). Weitere Berechnungen mit Jaccard-Sneath und Cosinus ergaben sehr ähnliche Ergebnisse.

Abbildung 1a ist das daraus resultierende Complete-Link Cluster (Schwellenwert = 0,1)

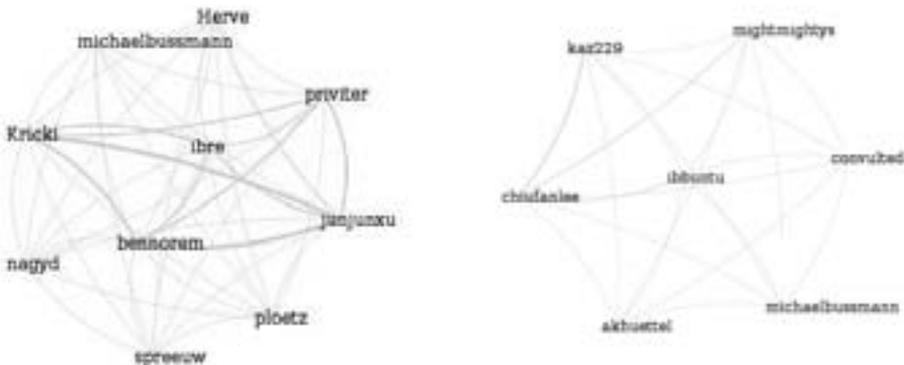


Abbildung 1a/b: Complete-link Cluster basierend auf a) Bookmarks (links) and b) Tags, Schwellenwert =0,1, berechnet nach Dice, Quelle: CiteULike

für den Nutzer „michaelbusmann“, dem ähnliche Nutzer anhand der Bookmarks aufgezeigt werden. Der Nutzer „michaelbusmann“ würde also einer „Gruppe“ aus Personen angehören, bei der die Nutzer „Kricke“, „bennorem“ und „junjunxu“ im Mittelpunkt die höchste Ähnlichkeit hätten (dargestellt durch die Dicke der Verbindungen zwischen den Nutzern). Das Recommender System würde „michaelbusmann“ diese als mögliche Kooperationspartner vorschlagen. In Abbildung 1b resultiert das Complete-Link Cluster (Schwellenwert = 0,1) aus Nutzern mit gleichen Tags. Der Unterschied ist deutlich zu erkennen, „michaelbusmann“ werden nun andere Personen vorgeschlagen. Generell ist festzustellen, dass viele Nutzer kaum oder wenige Tags für ihre Bookmarks vergeben. Häufig wird nur ein Schlagwort genutzt, was die tag-basierte Berechnungen schwieriger macht. Neben dem eigentlichen Nutzerverhalten steht und fällt die Qualität eines Expertenempfehlungssystems auf SBS-Basis mit der Qualität des jeweiligen SBS.

4 Zusammenfassung

Nutzer-Tag-Dokument Relationen in digitalen sozialen Netzwerken wie SBS ermöglichen es, implizite Beziehungsrelationen zwischen Nutzern sichtbar zu machen und Personen-Vorschlagsysteme zur besseren Kommunikation und effektiverem Informationsaustausch zu entwickeln. Zwar werden SBS zurzeit noch von einer kleinen (Wissenschaftler-)Gruppe genutzt, wie Studien u. a. von [BHD09] aufzeigen. Jedoch zeigen andere Untersuchungen, dass Tendenzen zu einer höheren Nutzungsbereitschaft in SBS erkennbar sind [HKG08]. Eine größere Nutzerschaft der SBS ist wünschenswert, um auf Basis der Bookmark- und Tag-Daten ein sinnvolles Vorschlagsystem zu entwickeln. Zudem sollten die Daten in den SBS vollständiger und besser verknüpft sein, damit a) die Nutzer den Dienst effektiver nutzen können und b) die Nutzerbeziehungen auch sinnvoll und realistisch dargestellt werden können.

Zur Diskussion steht noch, welcher Ähnlichkeitskoeffizient und welche Clustermethode am Effektivsten sind und ob eine Verknüpfung zwischen Nutzer-Tags und Nutzer-Bookmarks-Berechnungen erfolgen sollte, beispielsweise mit einer Gewichtung der Faktoren. Des Weiteren wird ein Hauptaspekt des Forschungsprojektes auf der qualitativen Nutzung unseres Systems liegen: Unsere jetzigen Daten aus den SBS beruhen auf 45 einschlägigen Physikzeitschriften, die von FZJ-Mitarbeitern vorrangig genutzt werden. Diese sollen das Empfehlungssystem im weiteren Teil der laufenden Untersuchungen evaluieren, um eindeutig zu ermitteln, ob ein solches Clustertool für die Kommunikation zwischen Wissenschaftlern sinnvoll und unterstützend bzw. überhaupt gewünscht ist. Hierbei stellt sich unter anderem die Frage, ob Wissenschaftler Kollaborationen mit ähnlich interessierten Kollegen eingehen, oder diese wegen des Konkurrenzgedankens meiden. Generell mag es sicherlich auch auf die einzelnen Disziplinen ankommen, weshalb eine Ausweitung der teilnehmenden Probanden angestrebt wird. Sollte sich herausstellen, dass die FZJ-Forscher keine SBS nutzen, werden auf Basis ihrer Literaturlisten und Publikationen „simulierte Nutzer“ erzeugt, denen dann potentielle Kooperationspartner anhand der realen Daten aus dem SBS vorgeschlagen werden. Die daraus entstehenden Nutzer-Cluster werden von den Wissenschaftlern qualitativ geprüft.

Danksagung

Wir danken den Kollegen des Teamprojekts „SoBoCops“ and Wolfgang G. Stock für ihre Unterstützung bei der Erarbeitungen dieses Beitrags. Teile des Forschungsprojektes werden von der DFG (STO 764/4-1) und dem Strategischen Forschungsfonds der Heinrich-Heine-Universität Düsseldorf gefördert.

5 Literaturverzeichnis

- [BHD09] Bernius, S., Hanauske, M., Dugall, B.: Von traditioneller wissenschaftlicher Kommunikation zu „Science 2.0“. In: ABI-Technik, 2009, 29, S. 214-226.
- [DI06] Diederich, J.; Iofciu, T.: Finding Communities of Practice from User Profiles Based on Folksonomies. In (Tomadaki, E.; Scott, P., Hrsg.), Innovative Approaches for Learning and Knowledge Sharing, EC-TEL 2006 Workshops Proceedings, Crete, Greece, S. 288–297.
- [FPB06] Fokker, J.; Pouwelse, J.; Buntine, W.: Tag-Based Navigation for Peer-to-Peer Wikipedia. In Proc. Collaborative Web Tagging Workshop at WWW 2006, Edinburgh, Scotland, 2006.
- [Go92] Goldberg, D.; Nichols, D.; Oki, B. M.; Terry, D.: Using Collaborative Filtering to Weave and Information Tapestry. In: Communications of the ACM, 1992, 35(12), S. 61-70.
- [Go06] Gordon-Murnane, L.: Social Bookmarking, Folksonomies, and Web 2.0 Tools. In: Searcher - The Magazine for Database Professionals, 2006, 14(6), S. 26–38.
- [HKG08] Heymann, P.; Koutrika, G.; Garcia-Molina, H.: Can Social Bookmarking Improve Web Search? In: First ACM Int. Conf. on Web Search and Data Mining, Stanford, CA, 2008.
- [Jä07] Jäschke, R.; Marinho, L.; Hotho, A.; Schmidt-Thieme, L.; Stumme, G.: Tag Recommendations in Folksonomies. In: LNAI, 2007, 4702, S. 506-514.
- [JS06] John, A.; Seligmann, D.: Collaborative Tagging and Expertise in the Enterprise. In: Proc. 15th Int. Conf. on WWW, Edinburgh, Scotland 2006. ACM Press, New York, NY, 2006.
- [KSS97] Kautz, H.; Selman, B.; Shah, M.: Referral Web: Combining Social Networks and Collaborative Filtering. Communications of the ACM, 1997, 40(3), S. 63–65.
- [Ke63] Kessler, M. M.: Bibliographic Coupling between Scientific Papers. In: American Documentation, 1963, 14, S. 10–25.
- [KSS10] Knautz, K.; Soubusta, S.; Stock, W.G.: Tag Clusters as Information Retrieval Interfaces. In: Proc. of HICSS-43, Kauai, Hawaii, 2010. IEEE Computer Society Press.
- [KH07] Kwiatkowski, M.; Höhfeld, S.: Empfehlungssysteme aus informationswissenschaftlicher Sicht - State of the Art. In: Information - Wissenschaft & Praxis, 2007, 58(5), S. 265-276.
- [MFK06] Millen, D. R.; Feinberg, J.; Kerr, B.: Dogear: Social Bookmarking in the Enterprise. In: Proc. Conf. on Human Factors in Computing Systems, Montréal, Canada 2006, S. 111-120.
- [Pe09] Peters, I.: Folksonomies: Indexing and Retrieval in Web 2.0., De Gruyter Saur, Berlin, 2009.
- [SM95] Shardanand, U.; Maes, P.: Social Information Filtering: Algorithms for Automating „Word of Mouth“. In: Proc. Hu. Fact. in Computing Systems, Denver, USA, 1995, S. 210-217.
- [Sm08] Smith, M.; Barash, V.; Getoor, L.; Lauw, H. L.: Leveraging Social Context for Searching Social Media. In (Soboroff, I.; Agichtein, E.; Kumar, R., Hrsg.), Proc. ACM Workshop on Search in Social Media, Napa Valley, California, 2008, S. 91–94.
- [Va05] Vander Wal, T.: Explaining and Showing Broad and Narrow Folksonomies, 2005. Retrieved from <http://www.vanderwal.net/random/entrysel.php?blog=1635>.