

Kollaboratives Tagging zur inhaltlichen Beschreibung von Lern- und Wissensressourcen¹

Ronald Maier, Stefan Thalmann

Institut für Wirtschaftsinformatik, Produktionswirtschaft und Logistik
Leopold Franzens Universität Innsbruck
A-6020 Innsbruck, Österreich
ronald.maier@uibk.ac.at
stefan.thalmann@uibk.ac.at

Abstract: Für die effektive Verwaltung von Lern- und Wissens-Ressourcen in unternehmensweiten Wissensinfrastrukturen sowie deren Verwendung in fortgeschrittenen Lern- und Wissensdiensten werden aussagekräftige Metadaten benötigt. Mit der starken Zunahme von Lern- und Wissensressourcen unterschiedlicher Qualität, Reife und Granularität in Unternehmen und Organisationen wird deren inhaltliche Beschreibung zunehmend herausfordernd, da einerseits die vollautomatische Extraktion keine befriedigenden Ergebnisse liefert und professionelle Metadatenautoren überlastet sind. In diesem Beitrag wird der Einsatz des kollaborativen Tagging zur inhaltlichen Beschreibung von Ressourcen im organisatorischen Umfeld diskutiert. Dazu werden Lern- und Wissensressourcen, Metadaten und der Ansatz des kollaborativen Taggings reflektiert. Im Rahmen eines Mehrunden-Laborexperiments werden Fragen der Handhabung dieses Ansatzes in Organisationen untersucht. Dies betrifft insbesondere die gemeinschaftliche Anerkennung von Schlagworten (commitment), die Dynamik zugeordneter Schlagworte (convergence) und die Beeinflussbarkeit des Verschlagwortungsprozesses (coordination).

1 Einleitung

Organisationen verwalten neben strukturierten Daten in (meist) relationalen Datenbanksystemen in zunehmendem Maße sogenannte semi-strukturierte digitale Ressourcen, die über eine Vielzahl von Systemen verteilt sind. Ein Teil der Ressourcen kann gezielt für Lern- und Wissensprozesse eingesetzt werden und wird daher hier als Lern- und Wissensressourcen (LWR) bezeichnet. Ihre systematische inhaltliche Beschreibung mit Hilfe von Metadaten und ihre semantische Integration lassen erhebliche Potenziale für die Steigerung der Produktivität des Umgangs mit diesen erhoffen.

Technisch bestehen die Ressourcen, die in zunehmend wissensintensiven Prozessen genutzt werden, primär aus semi-strukturierten Dokumenten bzw. Dateien. Für deren Austausch und Verwaltung werden bedeutungsvolle Metadaten benötigt. Die Erzeugung dieser Metadaten ist besonders für den Bereich der LWR, die eine besondere Beschreibung

¹ Maier, R.; Thalmann, S. (2007): Kollaboratives Tagging zur inhaltlichen Beschreibung von Lern- und Wissensressourcen. In: Tolksdorf, R.; Freytag, J. (Eds.): Proceedings of XML Tage 2007 in Berlin, Freie Universität Berlin, 75-86.

ihres semantischen Kontextes benötigen [Gr03], ein zeitaufwendiger und kostenintensiver Prozess. Diese Beschreibung sollte durch Fachexperten oder Themenverantwortliche durchgeführt werden, die jedoch aufgrund knapper zeitlicher Ressourcen oftmals nicht verfügbar sind, um diese wahrzunehmen [Du04]. Als Konsequenz werden LWR vielfach nur mit unvollständigen Metadaten ausgestattet [MD05]. Zwar werden einerseits in der Literatur viele Ansätze zur vollautomatischen Generierung von Metadaten beschrieben und andererseits in der Praxis insbesondere in Organisationen wie z.B. Bibliotheken und Museen professionelle Metadatenautoren mit der Verschlagwortung betraut [Gr03]. Keiner der beiden Ansätze hat sich jedoch außerhalb spezifischer Einsatzgebiete etabliert. Das im Internet bereits verbreitete Tagging könnte, in kollaborativer Form angewendet, hier Abhilfe schaffen.

Beim Tagging könnten alle Mitglieder einer Arbeitsgruppe, eines Projektes oder einer Community bei der Beschreibung von Ressourcen miteinbezogen werden. Dadurch könnten von der Expertise und den Assoziationen vieler, wenn nicht aller Wissensarbeiter profitiert werden. Die hohe Belastung der Fachexperten könnte so reduziert, die Last der Verschlagwortung auf die Schultern vieler verteilt sowie ein gruppenweiter Konsens über die Beschreibung von Ressourcen erzielt werden.

Ausgehend von dieser zentralen Hypothese werden im Rahmen dieses Artikels einige wesentliche Aspekte der Ausgestaltung eines kollaborativen Taggings bei der Verwendung in Unternehmen und Organisationen zur Beschreibung von LWR diskutiert und im Rahmen einer empirischen Studie untersucht. Dazu wird im nachfolgenden Teil die Bedeutung von Metadaten für den Bereich der LWR und deren Erstellung beschrieben. Anschließend wird Tagging als eine Möglichkeit zur Beschreibung von Ressourcen in Unternehmen und Organisationen reflektiert und im vierten Teil im Rahmen eines Experiments untersucht. Abschließend werden die zentralen Ergebnisse zusammengefasst und es wird ein Ausblick gegeben.

2 Metadaten für Lern- und Wissensressourcen

LWR liegen in unterschiedlichen Qualitätsstufen, Reifegraden [MS07] sowie Granularität vor. Kleinste Einheiten sind dabei Lern- und Wissenselemente. Diese können auf Basis des Begriffs Dokument (z.B. [GS04]) definiert werden als atomare Einheiten expliziter, formal definierter, durch Externalisierung oder Kombination von Wissen [NT95] (Wissenselement) und dessen didaktische Aufbereitung (Lernelement) erzielter Inhalte. Diese spezielle Form von Datensätzen wird sowohl von einer konzeptionellen als auch von einer technischen Perspektive als Einheit gesehen. Sie bestehen aus einer Gruppe zusammengehöriger Datenobjekte, die ohne substantiellen Bedeutungsverlust nicht voneinander getrennt werden können, zusammen mit zugehörigen beschreibenden Metadaten [Ma07]. Beispiele sind Lernobjekte, Web-Based Trainings, interne Studien, Lessons learned oder Best Practices.

Metadaten werden im Bereich der Datenbanken als reine Strukturbeschreibung von Daten aufgefasst [HS00]. Im Bereich von LWR wird jedoch ein weitergefasstes Verständnis über Metadaten notwendig und deshalb unter Metadaten eine detaillierte Beschreibung der

Charakteristika und des Formates einer Dateninstanz in Abhängigkeit vom Metadatenempfänger [Ta02] verstanden.

Die Verarbeitung von Metadaten erfordert ein standardisiertes Format [MHP05], denn nur mit diesem kann die Integrität der technischen Systeme gewährleistet werden. Metadatenstandards stellen dieses Format bereit und liefern eine syntaktische und semantische Beschreibung für die Erfassung von Metadaten. Metadaten für LWR werden zurzeit meist in XML-basierten Metadatenstandards, wie beispielsweise LOM², Dublin Core³ oder MPEG-7⁴, gespeichert [Bu04]. Diese bieten neben der breiten Verfügbarkeit von Werkzeugen und der damit verbundenen universellen Verwendbarkeit auch Vorteile durch eine leicht realisierbare Transformation, z.B. mit Hilfe von XSLT. Neben XML bietet RDF stärkere semantische Beschreibungsmöglichkeiten.

Metadaten, ihre formale Beschreibung und semantische Integration sind Kern von Systemen zur Unterstützung des Wissensmanagements, auch Wissensmanagementsysteme oder unternehmensweite Wissensinfrastrukturen (enterprise knowledge infrastructures) genannt [MHP05]. Diese stellen ihre Funktionen als Plattform zur Verfügung und treiben damit die Integration der in Unternehmen und Organisationen eingeführten Lern- und Wissensinstrumente sowie diese unterstützender Anwendungssysteme voran. Beispiele sind Dokumentenmanagementsysteme, Contentmanagementsysteme, in jüngerer Zeit vor allem einfach zu nutzende Systeme wie Wikis und Weblogs, Systeme zur Zusammenarbeit, ergänzt um sogenannte social software, aber auch betriebswirtschaftliche unternehmensweite Anwendungssysteme (enterprise systems).

Diese Infrastrukturen postulieren eine Integration zum einen unternehmensweit für alle für das Wissensmanagement relevanten Applikationen und zum anderen bieten sie auf dieser Basis eine Reihe von fortgeschrittenen Diensten an. Damit wird die bei der Gestaltung von Wissensportalen ([Co03], [Fi03], [Sa05]) eingeleitete Idee einer Integration von Applikationen für das Wissensmanagement in eine Infrastrukturlösung für das Gesamtunternehmen überführt. Eine Ontologie hilft dabei, Wissens Elemente, die aus verschiedenen Quellen stammen, zu analysieren und miteinander zu vernetzen. Die Integration heterogener Daten- und Wissensquellen erfolgt dabei auf Basis von Infrastrukturdiensten über die in den Metadaten beschriebene Semantik der Daten, über standardisierte Funktionsaufrufe und die Anordnung von Funktionen zu Prozessen sowie über eine einheitliche Verwaltung von Daten über die Teilnehmer. Die Kern-Wissensprozesse – (1) Entdecken, also Suchen und Präsentieren, (2) Publizieren, also Erfassen, Organisieren, Speichern, Portionieren, (3) Zusammenarbeit, also Verteilen, Verwenden, Revidieren, Kommentieren und (4) der Meta-Prozess des Lernens, der hinter allen externalisierenden, internalisierenden, sozialisierenden und kombinierenden Prozessen der Wissensumwandlung [NT95] steht, werden auf der Basis der Integrationsdienste miteinander verknüpft [Ma07]. Inhaltlich liegen diesen Prozessen LWR zugrunde.

In der Regel handelt es sich bei LWR nicht um rein textbasierte Elemente, sondern um Mischformen aus Text- und Multimediainhalten. Viele Beschreibungen der Metadaten beziehen sich dabei auf den Verwendungskontext und bei den Lernressourcen besonders

² http://tsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf

³ <http://dublincore.org/>

⁴ <http://www.mpegif.com/>

auf pädagogische und didaktische Beschreibungen [MBP06]. Außerdem handelt es sich bei diesen Metadaten um keine objektiven Informationen über Daten, da diese stark durch den Kontext ihrer Erstellung beeinflusst werden [NPN02].

Aus diesen Gründen erscheint eine automatische Generierung von Metadaten praktisch (noch) nicht umsetzbar [Pi04], [MBP06]. Auf Grund der Komplexität, des hohen Zeitaufwandes und der Probleme bei der Formalisierung und Strukturierung erscheint eine komplette Realisierung durch einzelne Menschen praktisch jedoch ebenso nicht realisierbar [CMD05], [MBP06]. Daher ist die Annotation von LWR eine oft vernachlässigte Aufgabe, was in der Konsequenz zu einer unvollständigen und damit mangelhaften Beschreibung der vorhandenen Ressourcen führt [MD05].

Insgesamt könnte eine Kombination der automatischen und der manuellen Annotation eine mögliche Lösung darstellen. So könnten technische Metadaten, wie z.B. Größe oder Format, die sich durch Extraktion zuverlässig automatisch erheben lassen [Gr04], durch automatische Verfahren und die weichen und kontextabhängigen Metadatenelemente, wie z.B. Schlagworte oder Schwierigkeitsniveaus, durch manuelle Eingaben erhoben werden. Für den zweiten Teil könnte social software im Allgemeinen [Bo03] und Tagging im Speziellen eine Lösung für die gemeinschaftliche Beschreibung von Ressourcen sein [KI06]. Wie das Tagging diese Erstellung unterstützen kann und welche Vorteile und Risiken es dabei gibt, soll im folgenden Abschnitt betrachtet werden.

3 Tagging zur Annotierung von Lern- und Wissensressourcen

Beim traditionellen Vorgehen der Annotierung katalogisiert und indiziert eine kleine Gruppe von Experten Ressourcen und schafft damit eine gemeinsame Basis für deren Nutzung innerhalb eines Unternehmens oder einer Organisation [MM06]. Mit der stetig ansteigenden Menge an Ressourcen, die innerhalb einer Organisation erstellt werden oder von außen in diese aufgenommen werden, übersteigen die mit diesem Ansatz verbundenen Kosten und der Zeitaufwand den erwarteten Nutzen. Des Weiteren fällt es den Experten schwer die Ressourcen für heterogene Anwendungsdomänen zu beschreiben [SM94], was durch die zunehmende Spezialisierung der Experten, die Fragmentierung von Wissen sowie die Vielfalt der Einsatzmöglichkeiten verstärkt wird. So kann es neben ökonomischen, Kapazitäts- und Motivationsproblemen bei der Annotation von Ressourcen auch zu Akzeptanzproblemen im Nutzerkreis kommen.

Zur Überwindung dieser Herausforderungen könnte das Tagging eine adäquate Lösung darstellen. Einerseits könnte das gemeinschaftliche Annotieren von Ressourcen eine Entlastung der Autoren, Experten und Themenverantwortlichen [MD05] ermöglichen und andererseits könnte die Akzeptanz innerhalb der Nutzerschaft verbessert werden, da diese eine Gruppenübereinkunft über die Beschreibung erzielt.

Ein „Tag“ ist dabei ein Begriff, der vom Benutzer frei Ressourcen zugeordnet werden kann, also nicht wie traditionell aus einem vorgegebenen Vokabular ausgewählt wird [Ro07]. Die Zuordnung erfolgt dabei in einem [Benutzer, Ressource, Schlagwort]-Tripel [CLP07]. Diese Tripel können durch statistische Auswertungen in Tag-Wolken, Tag-Netzwerken oder Tag-Clustern dargestellt werden [Ro07]. Tags dienen dabei im Allge-

meinen primär dazu den persönlichen Arbeitsbereich zu strukturieren und zu organisieren und in gemeinschaftlichen Bereichen die Navigation und die Suche nach Ressourcen zu verbessern [MM06]. Mit einer großen Anzahl von Nutzern, die Ressourcen annotieren, kann durch das Tagging ein weites Spektrum von Beschreibungen für verschiedene Anwendungsbereiche gewonnen werden. Im Sinne der ökonomischen Netzwerktheorie [Ho01] steigt der erwartete Nutzen dieses Ansatzes mit der Zahl der Benutzer. Durch die einfache Handhabung von Werkzeugen für das Tagging können die Barrieren zur Teilnahme und die Kosten für die Organisation niedrig gehalten werden.

Vorgegebene Vokabulare stellen oft eine Barriere für die Teilnahme dar, da die Benutzer ihre eigenen Sichtweisen bzw. Beschreibungen, die nicht im vorgegebenen Vokabular vorhanden sind, nicht eingeben können [SM04]. Außerdem lässt sich das primär im Bibliothekswesen verwendete Verfahren nicht immer auf digitale Ressourcen übertragen [MM06]. Das Tagging mit seinen unbeschränkten Beschreibungsmöglichkeiten kann diese Barrieren reduzieren und die Motivation zur Partizipation erhöhen. Erfolgreich angewendet wird das Tagging bereits im Internet, z.B. <http://del.icio.us/> oder <http://www.annotea.org/>. Neben Freitextbeschreibungen kann der Benutzer meist auch aus einer Anzahl von häufig verwendeten Tags auswählen und somit eine Webresource beschreiben.

Beim Tagging entsteht eine Menge von gemeinschaftlich anerkannten Beschreibungen, die von einem großen Teil der Gemeinschaft getragen wird, und eine Menge von spezifischen Beschreibungen, die nur von relativ wenigen Mitgliedern der Gemeinschaft getragen wird [GH05], [GT06]. Für Organisationen erscheint die erste Menge bedeutend, da diese die gemeinschaftliche Übereinkunft über die Beschreibung einer Ressource darstellt und somit für die Annotation und breite Wiederverwendung besser geeignet erscheint. In diesem Sinne könnten die Tags mit der höchsten Übereinstimmung als „semantische Breite“ einer Ressource betrachtet werden [CLP07].

Für eine Anwendung dieses Ansatzes in Organisationen ergeben sich einige Fragen bezüglich dessen Ausgestaltung. Die Nutzbarkeit der anerkannten Tags erfordert (1) die Konvergenz gegen eine stabile Menge von Tags (convergence) und (2) eine Verteilung, bei der sich anerkannte Tags zuverlässig von spezifischen Tags trennen lassen (commitment). Für einen Einsatz in einem organisatorischen Umfeld ist (3) die Beeinflussbarkeit der Beschreibungsrichtung durch organisatorische Maßnahmen bedeutsam (coordination). Aufgrund der relativen Neuartigkeit dieses Ansatzes liegen bisher noch keine gesicherten Erkenntnisse über den Einsatz von Tagging in Organisationen und die Effekte unterschiedlicher Gestaltungsalternativen vor.

4 Experiment

Mit dem nachfolgend beschriebenen Mehrunden-Laborexperiment wurden folgende Fragen untersucht, die als bedeutsam für einen Einsatz des kollaborativen Taggings in einem organisatorischen Umfeld erachtet werden:

1. Convergence: Konvergiert die Menge an vergebenen Tags?

2. Commitment: Kann die Menge von gemeinschaftlich anerkannten und die Menge an spezifischen Tags zuverlässig separiert werden?
3. Coordination: Kann die Beschreibungsrichtung durch organisatorische Maßnahmen beeinflusst werden?

Das Experiment wurde von den Autoren im Mai/Juni 2007 an der Universität Innsbruck mit Studierenden der Wirtschaftswissenschaften im ersten Studienabschnitt im Rahmen einer Lehrveranstaltung zur Einführung in die Wirtschaftsinformatik durchgeführt. Am Experiment nahmen insgesamt 174 Studierende teil. Die Studie wurde in zwei Versuchsreihen mit insgesamt 7 Übungsgruppen von je 20-28 Studierenden in einem PC-Labor durchgeführt. Die Anzahl der Teilnehmer in den einzelnen Durchgängen der zwei Versuchsreihen ist in Tabelle 1 abgebildet.

| | Versuchsreihe 1 | Versuchsreihe 2 |
|-------------|-----------------|-----------------|
| 1 Durchgang | 28 Teilnehmer | 27 Teilnehmer |
| 2 Durchgang | 20 Teilnehmer | 23 Teilnehmer |
| 3 Durchgang | 23 Teilnehmer | 27 Teilnehmer |
| 4 Durchgang | % | 26 Teilnehmer |

Tabelle 1: Anzahl der Teilnehmer in den Versuchsreihen

Im Rahmen der Studie versahen die Studierenden 10 vorgegebene LWR mit Schlagworten. Dabei wurden zwei kurze Lehr-Videos, drei Vorlesungsfolien, drei Screen-Shots und zwei Grafiken verwendet. Wie bei frei verfügbaren Tagging-Werkzeugen üblich, wurden dabei die häufigsten Schlagworte angezeigt und um die Möglichkeit der Eingabe zusätzlicher Schlagworte erweitert. Technisch wurde die Studie mit einem am Institut für Wirtschaftsinformatik entwickelten web-basierten elektronischen Fragebogen realisiert, bei dem die Studierenden die fünf häufigsten Schlagworte durch Anklicken wählen und eigene Schlagworte über ein Textfeld eingeben konnten. Dabei war es möglich eine beliebige Kombination aus eigenen und vorgegebenen Schlagworten zu wählen, wobei mindestens ein Schlagwort pro Ressource angegeben werden musste. Die Zusammensetzung der fünf häufigsten Schlagworte war innerhalb eines Durchgangs stabil.

Für 8 Ressourcen wurden je zwei verschiedene Startmengen an Schlagworten vorgegeben, um die Beeinflussbarkeit der Beschreibungsrichtung überprüfen zu können. Dabei wurden in der ersten Versuchsreihe die Ressourcen fachlich-inhaltlich beschrieben, während in der zweiten Versuchsreihe der Kontext der Ressourcen beschrieben wurde, z.B. der Lehrveranstaltungsleiter oder die verwendete Sprache. Zusätzlich wurden zwei Ressourcen in beiden Versuchsreihen mit der gleichen initialen Menge an Schlagworten versehen. Eine von diesen Ressourcen wurde mit- aus Sicht der Autoren - gänzlich unpassenden Schlagworten versehen. Die Menge der häufigsten Schlagworte wurde nach jedem Durchgang neu ermittelt und im nächsten Durchgang als Startmenge verwendet.

Im Rahmen des Experiments wurden von den Studierenden insgesamt 4.246 Zuordnungen vorgenommen, wobei den 10 Ressourcen insgesamt 404 verschiedene Schlagworte zugeordnet wurden. In der ersten Versuchsreihe wurden 10% und in der zweiten Versuchsreihe 20% der in den Top-5 vorgegebenen Schlagworte durch Schlagworte der Studierenden

den ersetzt. Insgesamt vergaben die Studierenden 481 zusätzliche Schlagworte (inklusive Doppelangaben).

Bei der Gesamtverteilung der zugeordneten Tags konnte eine klare Trennung zwischen Schlagworten mit hoher Zustimmung (gemeinschaftlich anerkannte Tags) und den Schlagworten mit einer geringen Zustimmung (spezifische Tags) festgestellt werden. Die Verteilungen der den 10 Ressourcen zugeordneten häufigsten sieben Schlagworte sind in Tabelle 2 abgebildet. Die Schlagworte sind nach ihrer Häufigkeit sortiert und zu der entsprechenden Platzierung sind die jeweiligen absoluten und relativen Häufigkeiten angegeben. Die ersten fünf bis sechs Elemente weisen durchgängig eine relativ hohe Häufigkeit auf und die nachfolgenden Elemente eine deutlich geringere Häufigkeit. Die restlichen Schlagworte sind nur als Summe angegeben, und wurden je maximal viermal, meist nur einmal zugeordnet. Die Separierung kann daher gut an dem deutlichen Abfall der Häufigkeiten vorgenommen werden. Die Verteilung ist in beiden Versuchsreihen sehr ähnlich, so dass in Tabelle 2 nur die Daten aus der zweiten Versuchsreihe dargestellt sind.

| | Ressource 1 | | Ressource 2 | | Ressource 3 | | Ressource 4 | | Ressource 5 | |
|-----------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|--------------|-------------|
| Platz 1 | 70 | 24% | 58 | 22% | 68 | 29% | 92 | 29% | 69 | 28% |
| Platz 2 | 60 | 20% | 55 | 21% | 56 | 24% | 52 | 17% | 47 | 19% |
| Platz 3 | 57 | 19% | 53 | 20% | 40 | 17% | 51 | 16% | 44 | 18% |
| Platz 4 | 37 | 13% | 34 | 13% | 25 | 11% | 46 | 15% | 37 | 15% |
| Platz 5 | 37 | 13% | 12 | 5% | 25 | 11% | 41 | 13% | 22 | 9% |
| Platz 6 | 5 | 2% | 8 | 3% | 3 | 1% | 7 | 2% | 3 | 1% |
| Platz 7 | 3 | 1% | 4 | 2% | 2 | 1% | 4 | 1% | 3 | 1% |
| Restliche | 25 | 9% | 38 | 15% | 15 | 6% | 19 | 6% | 21 | 9% |
| Summe | 294 | 100% | 262 | 100% | 234 | 100% | 312 | 100% | 246 | 100% |
| | Ressource 6 | | Ressource 7 | | Ressource 8 | | Ressource 9 | | Ressource 10 | |
| Platz 1 | 69 | 25% | 82 | 35% | 76 | 29% | 75 | 34% | 79 | 31% |
| Platz 2 | 57 | 21% | 51 | 22% | 57 | 21% | 54 | 25% | 62 | 24% |
| Platz 3 | 46 | 17% | 33 | 14% | 56 | 21% | 31 | 14% | 30 | 12% |
| Platz 4 | 40 | 15% | 30 | 13% | 25 | 9% | 30 | 14% | 30 | 12% |
| Platz 5 | 38 | 14% | 22 | 9% | 23 | 9% | 15 | 7% | 29 | 11% |
| Platz 6 | 13 | 5% | 3 | 1% | 4 | 2% | 1 | 0% | 4 | 2% |
| Platz 7 | 3 | 1% | 2 | 1% | 3 | 1% | 1 | 0% | 3 | 1% |
| Restliche | 8 | 3% | 13 | 6% | 22 | 8% | 12 | 5% | 20 | 8% |
| Summe | 274 | 100% | 236 | 100% | 266 | 100% | 219 | 100% | 257 | 100% |

Tabelle 2: Häufigkeiten der zugeordneten Schlagworte in Versuchsreihe 2

Als Separierungskriterium könnte in den vorliegenden Fällen beispielsweise ein 5% Kriterium eingesetzt werden. Erhielte ein Schlagwort mindestens 5% aller Schlagwortzuordnungen einer Ressource, so würde es als gemeinschaftlich anerkannter Tag betrachtet. Auf diese Weise könnten 5-6 Schlagworte pro Ressource ausgewählt werden, die in der ersten Versuchsreihe durchschnittlich 90,3% und in der zweiten Versuchsreihe 90,0% aller Zuordnungen einer Ressource repräsentieren würden. Alternativ könnte ein 0,9 Quantil angewendet werden. Dies hätte jedoch zur Folge, dass in wenigen Fällen weitere Werte mit einem Anteil von ca. 1% an den Gesamtnennungen in die Menge der gemeinschaftlich anerkannten Tags aufgenommen werden.

Über die sieben Durchgänge konnte in beiden Versuchsreihen eine Stabilisierung der Menge der fünf häufigsten Schlagworte beobachtet werden. Wie in Abbildung 1 veranschaulicht, nahm die Anzahl der Veränderungen in den Platzierungen der häufigsten Schlagworte in beiden Versuchsreihen mit jedem Durchgang monoton ab. Gab es im ersten Durchgang noch 40 bzw. 27 von 50^5 möglichen Platzierungsänderungen, so gab es im dritten Durchgang noch 15 bzw. 14 Änderungen. Darüber hinaus nahm die Zahl der neuen, also durch die Studierenden über die bisherige Menge hinaus eingegebenen Schlagworte ebenso kontinuierlich ab und war bereits nach dem zweiten Durchgang in beiden Versuchsreihen bei 0. So war die Menge der gemeinschaftlich anerkannten Schlagworte bereits nach dem zweiten Durchgang konstant und änderte sich nur noch in der Reihenfolge. Beide Beobachtungen unterstützen die Annahme, dass sich die Menge an Schlagworten stabilisiert und gegen eine stabile Zusammensetzung konvergiert.

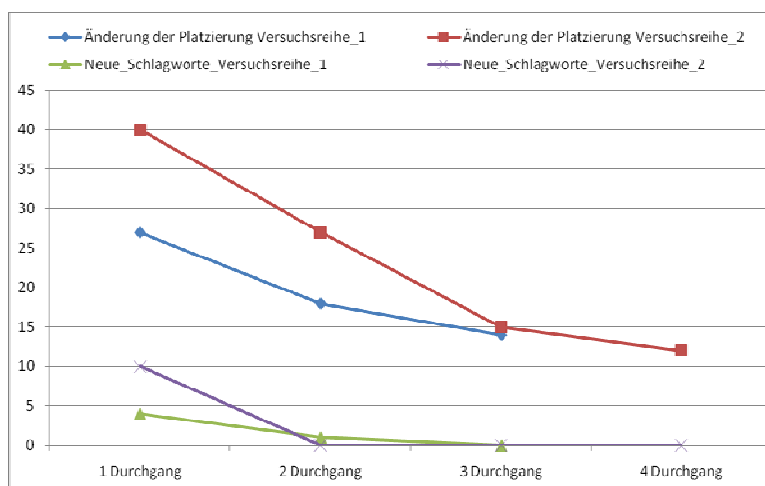


Abbildung 1: Veränderungen in der Menge der 5 häufigsten Schlagworte

Die Anzahl textueller Eingaben pro Teilnehmer unterschieden sich zwischen den Gruppen deutlich und lagen im Gruppendurchschnitt zwischen 1,6 und 4,0 Texteingaben je Ressource. Über alle sieben Gruppen wurden durchschnittlich 2,8 Texteingaben pro Teilnehmer je Ressource abgegeben. Die Gesamtanzahl der unterschiedlichen Tags stieg in beiden Versuchsreihen stetig an, verlor jedoch in beiden Fällen an Dynamik. Auf Grund des unterschiedlichen Eingabeverhaltens der Teilnehmer zwischen den Gruppen wurde nicht die Anzahl unterschiedlicher Tags zur Teilnehmeranzahl abgetragen, sondern das Verhältnis aus unbekanntem Schlagworten zu Texteingaben.

Dieses Verhältnis entwickelte sich wie in Abbildung 2 dargestellt. In der ersten Versuchsreihe sank die Quote zunächst stark und blieb dann stabil. In der zweiten Versuchsreihe sank die Quote von einem niedrigeren Ausgangsniveau über zwei Durchgänge, stieg aber im vierten Durchgang wieder an. Bezüglich der Entwicklung ergibt sich damit zwar eine leichte Tendenz, aber insgesamt kein eindeutiges Bild.

⁵ 10 Ressourcen mit je 5 Platzierungen = 50 mögliche Platzierungen

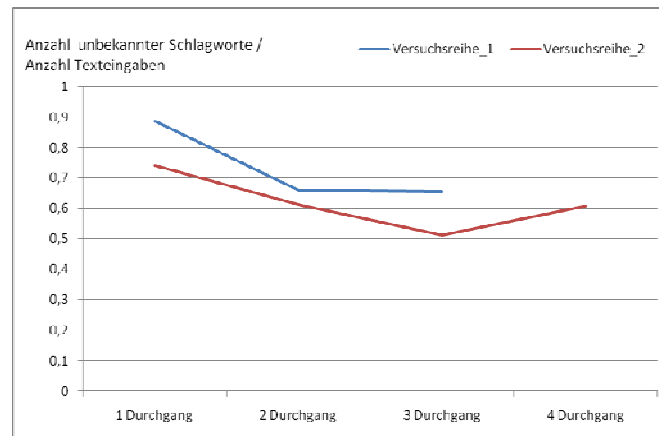


Abbildung 2: Quote von neuen Schlagworten zu Texteingaben

Durch verschiedene Kombinationsmöglichkeiten von zusammengesetzten Beschreibungen und Abkürzungen, wie beispielsweise Power_Point_Präsentation, PowerPoint, Power_Point oder PPT, wurden inhaltlich ähnliche Schlagworte abgegeben, jedoch als verschiedenartige Schlagworte gewertet. Des Weiteren wurden Schlagworte sowohl in Deutsch als auch in Englisch, wie beispielsweise Data und Daten und auch Synonyme, wie Rechner und Computer, eingegeben. Auf Grund der großen Anzahl von Kombinationsmöglichkeiten und verschiedenartigen Schlagworten erscheint der Stichprobenumfang jedoch zu gering, um hier eindeutige Aussagen treffen zu können.

Aus einer qualitativen Perspektive unterschieden sich beide Versuchsreihen inhaltlich bei den Texteingaben substantiell. So wurden in der Versuchsreihe mit der stark inhaltlich/fachlich orientierten Startmenge an Schlagworten überwiegend weitere inhaltlich/fachlich orientierte Schlagworte eingegeben. Bei der zweiten Versuchsreihe mit den Kontext der Ressource beschreibenden Schlagworten erfolgte primär die Eingabe weiterer den Kontext beschreibender Schlagworte. Insgesamt unterschieden sich die Texteingaben pro Ressource um 71% – 93% und durchschnittlich um 81% (d.h. 81% der Texteingaben waren unterschiedlich). Bei den gleichen Texteingaben handelte es sich meist um Begriffe, die gut platziert in der Ressource, wie z.B. eine Überschrift, standen. Die fünf häufigsten Schlagworte zu den Ressourcen, die in beiden Versuchsreihen mit den gleichen Startmengen begonnen wurden, waren bis auf geringe Unterschiede in der Platzierung identisch. Es setzte sich dabei sogar in beiden Versuchsreihen die gleiche Texteingabe der Studierenden durch. Diese Beobachtungen unterstützen die These, dass mit der Wahl der Startmenge die Richtung der vergebenen Schlagworte gezielt beeinflusst werden kann.

Zu den Ressourcen mit den, aus Sicht der Autoren, unpassenden Schlagworten gab es überdurchschnittlich viele textuelle Eingaben der Studierenden (durchschnittlich 12,8 Texteingaben zu 6,1 Texteingaben bei den anderen Ressourcen). Außerdem wurden hier überdurchschnittlich viele Schlagworte (durchschnittlich 2 Schlagworte zu 0,6 bei den restlichen Ressourcen) aus der Startmenge durch Texteingaben der Studierenden ersetzt. Einige Schlagworte konnten sich jedoch behaupten. Hier lässt sich vermuten, dass die

Studierenden versucht haben einen Bezug zu den vorgegebenen Ressourcen herzustellen und so nicht unbefangen über die Ressource nachgedacht haben.

Aus Sicht der Autoren konnten mit dem vorgestellten Experiment erste Erfahrungen mit dem Einsatz von Tagging für die Annotation von LWR gesammelt werden. In Bezug auf die Forschungsfragen lassen sich erste Tendenzaussagen treffen, die in weiteren Experimenten bzw. Feldstudien überprüft werden müssen. So lassen sich die gemeinschaftlich anerkannten Tags gut von den spezifischen Tags separieren. Die Konvergenz der Menge an Tags wird durch die Daten dagegen nicht eindeutig gestützt. Außerdem scheint die Startmenge von Tags großen Einfluss auf die Denkrichtung der Annotierenden zu haben. Während in dieser Studie commitment und coordination unterstützt werden, ist convergence eine weiter offene Frage.

5 Zusammenfassung und Ausblick

In der vorliegenden Arbeit wurde zunächst die Bedeutung von Metadaten für LWR beschrieben und anschließend auf die Probleme bei deren Erstellung eingegangen. Dabei wurde deutlich, dass es weder mit rein automatischen noch mit rein manuellen Ansätzen ökonomisch sinnvoll und praktisch möglich ist, adäquate Metadaten für die meist in Organisationen vorliegende große Zahl von LWR zu erstellen. Das Tagging wurde dabei als eine einfache Möglichkeit der gruppenbasierten Annotation vorgestellt, bei der einerseits Fachexperten und Themenverantwortliche entlastet werden und andererseits ein gruppenweiter Konsens über eine Beschreibung gefunden werden kann.

Im Rahmen eines Experiments ließen sich Hinweise darauf finden, dass Tagging in einem organisationalen Umfeld eingesetzt werden könnte. Gemeinschaftlich anerkannte Tags scheinen sich unter Anwendung einfacher Kriterien rasch von spezifischen Tags trennen zu lassen. Darüber hinaus kann durch die große Vielfalt an Schlagworten ein sehr breites Spektrum an Begriffen gewonnen und für den Aufbau von Begriffswelten, Taxonomien oder Ontologien verwendet werden.

Durch die Wahl der Startmenge an Schlagworten scheint es außerdem möglich zu sein, die Beschreibungsrichtung zu beeinflussen. Dies könnte zum einen in einem organisatorischen Prozess genutzt werden, um die gewünschte Form der Beschreibung durchzusetzen, Themenschwerpunkte zu setzen oder eine strategische Richtung vorzugeben. Andererseits könnten unachtsame oder böswillige Nutzer den Beschreibungsverlauf negativ beeinflussen. Nach den Ergebnissen dieser Studie scheint sich die Zusammensetzung der Schlagworte durch folgende Punkte maßgeblich beeinflussen zu lassen: die Startmenge an Schlagworten und deren Beschreibungsrichtung, die Anzahl der vorgegebenen und damit auswählbaren Tags, die Anzahl annotierender Anwender und die Motivation der Anwender.

Durch das rasche Commitment auf die TOP 5-6 Schlagworte fällt auf, dass auf Grund der hohen absoluten Werte, welche einzelne Schlagworte aufweisen, die Empfindlichkeit gegenüber neuen Strömungen sehr gering wird. So wurden einige Schlagworte in jedem Durchgang mehrfach eingegeben, schafften es jedoch nicht in die Top-5. Es lässt sich

dabei keine eindeutige Tendenz bezüglich der Dynamik bei der Generierung neuer Schlagworte erkennen. Um ein zu rasches Commitment und damit den Verlust potentiell interessanter Vorschläge zu vermeiden, bei Änderungen in der Umwelt bzw. nach einer gewissen Zeit sollte es möglich sein die stabile Menge aufzubrechen und empfindlicher für Änderungen zu gestalten. Dies könnte über eine zeitliche Diskontierung der Wortmeldungen, eine gezielte Auslese aus den zusätzlich eingebrachten Begriffen oder über einen völligen Neustart der Annotation erfolgen.

Limitationen der empirischen Studie sind der kleine Stichprobenumfang, die nicht fortlaufende Berechnung der Platzierung, die Festlegung auf 5 angezeigte Schlagworte und mögliche systematische Auswirkungen bedingt durch die Unterrichtssituation und das damit verbundene implizite Rollenverständnis der Studierenden. In einer aufbauenden Studie soll neben der Überprüfung der Unempfindlichkeit gegenüber neuen Strömungen und einer größeren Stichprobe auch eine fortlaufende Berechnung der Häufigkeiten implementiert werden. Des Weiteren soll die Anzahl der angezeigten Schlagworte an dem im Rahmen dieser Studie ermittelten Separierungskriterium orientiert werden.

Zusammenfassend lässt sich festhalten, dass kollaboratives Tagging einen viel versprechenden, partizipativen Ansatz für die semantische Beschreibung von LWR darstellt. Aus der Kombination der Essenz des derzeit stark propagierten Web 2.0-Ansatzes [OR05] mit den Vorzügen unternehmensweiter Wissensinfrastrukturen und des kollaborativen Taggings im Speziellen zusammen mit der gezielten Ausgestaltung dieses Ansatzes lassen sich hohe Potenziale für die Unterstützung von selbst gesteuerten bzw. gemeinschaftlichen Lern- und Wissensprozessen erhoffen.

Literaturverzeichnis

- [Bo03] Boyd, S.: Are You Ready for Social Software?, Darwin Online Magazine, May 2003, URL: <http://www.darwinmag.com/read/050103/social.html>, Zugriff am 08.09.2006.
- [Bu04] Bunschowski, M.; Röser, M.; Tavangarian, D.; Voigt, D.: Eine differenzierte Metadatennotation XML-basierter E-Learning-Ressourcen für ein zuverlässiges Metadatenmanagement. In: Seehusen, S. (Hrsg.): DeLFI 2004: Tagungsband der 2. e-Learning Fachtagung Informatik. Gesellschaft für Informatik, Bonn, 2004, 139-150.
- [CLP07] Cattuto, C.; Loreto, V.; Pietronero, L.: "Semiotic dynamics and collaborative tagging." PNAS 104(5), 2007: 1461-1464.
- [CMD05]Cardinaels, K.; Meire, M.; Duval, E.: Automatic Metadata Generation: the Simple Indexing Interface. The International World Wide Web Conference (WWW05), Chiba, Japan, ACM, 2005.
- [Co03] Collins, H.: Enterprise Knowledge Portals. New York 2003.
- [Du02] Duval, E.; Hodgins, W.; Sutton, S.; Weibel, S. L.: Metadata Principles and Practicalities, in: D-Lib Magazin, Vol. 8, Nr. 4, 2002.
- [Fi03] Firestone, J. M.: Enterprise Information Portals and Knowledge Management. Amsterdam 2003.
- [GH05] Golder, S. A.; Huberman, B. A.: The Structure of Collaborative Tagging Systems. Palo Alto, CA (USA), Information Dynamics Lab: HP Labs, 2005.
- [GS04] Götzer, K.; Schneiderath, U.; Maier, B.; Komke, T.: Dokumenten-Management. Informationen im Unternehmen effizient nutzen, 3. Auflage, Heidelberg 2004.

- [Gr03] Greenberg, J.: Metadata Generation: Processes, People and Tools, in: Bulletin of the American Society for Information Science and Technology, 2003, Vol. 29, Nr. 2, 2003, 16-19.
- [Gr04] Greenberg, J.: Metadata Extraction and Harvesting: A Comparison of Two Automatic Metadata Generation Applications. *Journal of Internet Cataloging*, 6(4), 2004, 59-82.
- [GT06] Guy, M.; Tonkin, E.: "Folksonomies - Tidying up Tags?" *D-Lib Magazine* 12(1), 2006.
- [HS00] Heuer, A.; Saake, G.: *Datenbanken Konzepte und Sprachen. 2. aktualisierte und erweiterte Auflage*, mipt-Verlag, Bonn, 2000.
- [Ho01] Hofmann, U.: *Netzwerk-Ökonomie*, Physica-Verlag, Heidelberg 2001.
- [KI06] Klamma, R.; Amine Chatti, M.; Duval, E.; et al.: *Social Software for Professional Learning: Examples and Research Issues*, in: Proceedings of the 6th IEEE International Conference on Advanced Learning Technologies ICALT. Kerkrade, The Netherlands, IEEE Computer Society, 2006.
- [Ma07] Maier, R.: *Knowledge Management Systems. Information and Communication Technologies for Knowledge Management*, Springer Verlag, Berlin 2007.
- [MHP05] Maier, R.; Hädrich, T.; Peinl, R.: *Enterprise Knowledge Infrastructures*. Springer-Verlag, Berlin 2005.
- [MS07] Maier, R.; Schmidt, A.: *Charcterizing Knowledge Maturing: A Conceptional Model Integrating E-Learning and Knowledge Management*. 4th Conference of Professional Knowledge Management (WM07), Potsdam, 2007.
- [MD05] Malaxa, V.; Douglas, I.: *A Framework for Metadata Creation Tools*, in: *Interdisciplinary Journal of Knowledge and Learning Objects*, Vol. 1, 2005.
- [MM06] McGregor, G.; McCulloch, E.: *Collaborative tagging as a knowledge organisation and resource discovery tool*. *Library Review* 55(5), 2006: 291 - 300.
- [MBP06] Motelet, O.; Baloian, N.; Pino, J. A.: *Learning Object Metadata and Automatic Processes: Issues and Perspectives*, Informing Science Press, 2006.
- [NPN02] Nilsson, M.; Palmér, M.; Naeve, A.: *Semantic Web Meta-data for e-Learning – Some Architectural Guidelines.*, Royal Institute of Technology Stockholm, 2002. URL: <http://kmr.nada.kth.se/papers/SemanticWeb/p744-nilsson.pdf>, Zugriff am 08.06.2007.
- [NT95] Nonaka, I.; Takeuchi, H.: *The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation*, Oxford University Press, New York, Oxford, 1995.
- [OR05] O'Reilly, T.: *What is Web 2.0? Design Patterns and Business Models for the Next Generation of Software*, URL: <http://www.oreilynet.com/>, Zugriff am 01.06.2007
- [Pi04] Pinkwart, N.; Jansen, M.; Oelinger, M.; et al.: *Partial Generation of Contextualized Metadata in a Collaborative Modeling Environment*. 2nd International Workshop on Applications of Semantic Web Technologies for E-Learning -International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems AH 2004, Eindhoven, The Netherlands, 2004.
- [Ro07] Rollett, H.; Lux, M.; Strohmaier, M.; et al.: "The Web 2.0 way of learning with technologies." *International Journal of Learning Technology* 3(1), 2007: 87-107.
- [Sa05] Sandkuhl, K.: *Wissensportale. Merkmale, Architekturen und Perspektiven*. In: *Informatik-Spektrum*. 28 (2005) 3, 193-201.
- [SM94] Shipman, F.; McCall, R.: *Supporting Knowledge-Base Evolution with Incremental Formalization*. Conference on Human Factors in Computing Systems, Boston, ACM Press, 1994.
- [Ta02] Tannenbaum, A.: *Metadata Solutions: using metamodels, repositories, XML, and enterprise portals to generate information on demand*. Addison-Wesley, Boston, 2002.