

Was vom Tagging übrig blieb – Erkenntnisse und Einsichten aus zwei Jahren Projektarbeit

Christian Hänger, Christine Krätzsch, Christof Niemann

1 Das DFG-Projekt „Collaborative Tagging als neue Form der Sacherschließung“

Im Oktober 2008 startete an der Universitätsbibliothek Mannheim das DFG-Projekt „Collaborative Tagging als neue Form der Sacherschließung“. Über zwei Jahre hinweg wurde untersucht, welchen Beitrag das Web 2.0-Phänomen Tagging für die inhaltliche Erschließung von bisher nicht erschlossenen und somit der Nutzung kaum zugänglichen Dokumenten leisten kann. Die freie Vergabe von Schlagwörtern in Datenbanken durch die Nutzer selbst hatte sich bereits auf vielen Plattformen als äußerst effizient herausgestellt, insbesondere bei Inhalten, die einer automatischen Erschließung nicht zugänglich sind. So wurden riesige Mengen von Bildern (Flickr), Filmen (YouTube) oder Musik (LastFM) durch das Tagging recherchierbar und identifizierbar gemacht. Die Fragestellung des Projektes war entsprechend, ob und in welcher Qualität sich durch das gleiche Verfahren beispielsweise Dokumente auf Volltextservern oder in elektronischen Zeitschriften erschließen lassen.

Für die Beantwortung dieser Frage, die ggf. weitreichende Konsequenzen für die Sacherschließung durch Fachreferenten haben konnte, wurde ein ganzer Komplex von Teilfragen und Teilschritten ermittelt bzw. konzipiert. Im Kern ging es aber in allen Untersuchungsschritten immer um zwei zentrale Dimensionen, nämlich um die „Akzeptanz“ und um die „Qualität“ des Taggings. Die Akzeptanz des Taggings wurde zunächst bei den Studierenden und Wissenschaftlern der Universität Mannheim evaluiert. Für bestimmte Zeiträume wurden Tagging-Systeme in unterschiedlichen Ausprägungen an die Recherchedienste der Universitätsbibliothek angebunden. Die Akzeptanz der einzelnen Systemausprägungen konnte dann durch die Analyse von Logfiles und durch Datenbankabfragen ausgewertet werden. Für die Qualität der Erschließung wurde auf einen Methodenmix zurückgegriffen, der im Verlauf des Projektes immer wieder an aktuelle Entwicklungen und an die Ergebnisse aus den vorangegangenen Analysen angepaßt wurde. Die Tags wurden hinsichtlich ihres Beitrags zum Information Retrieval mit Verfahren der automatischen Indexierung von Volltexten sowie mit der Erschließung durch Fachreferenten verglichen. Am Schluss sollte eine gut begründete Empfehlung stehen, wie bisher nicht erschlossene Dokumente am besten indexiert werden können: automatisch, mit Tags oder durch eine Kombination von beiden Verfahren.

2 Methodik und Arbeitsschritte

Bei der Sichtung der aktuellen Literatur zum Thema Tagging sowie bereits verfügbarer ähnlicher Tagging-Angebote in wissenschaftlichen Bibliotheken stellte sich heraus, dass insgesamt (national wie auch international) nur wenige Implementierungen existierten, die über Pilotprojekte hinausgingen. Dieses Rechercheergebnis machte es notwendig, auf der Basis der wenigen vorhandenen Lösungen und durch eigene Überlegungen zu einem Kriterienkatalog für einen bibliothekarischen Tagging-Service zu kommen. Als wesentliche Elemente wurden (a) eine intuitive Benutzeroberfläche, (b) die Anschlussfähigkeit des Systems an etablierte Standards (BibTeX), (c) seine Gebrauchswertorientierung (z.B. durch einen Literaturlistenexport oder erweiterte Recherchefunktionen), (d) präzise Hilfe-Seiten und Nutzer-Support sowie (e) eine allgemeine Qualitätskontrolle der Tags identifiziert. Mit diesen Kriterien ausgestattet wurde ein eigener Tagging-Service konzipiert, der zunächst eine wechselseitige und nahtlose Anbindung des Online-Katalogs an den frei zugänglichen, externen Tagging-Dienst "BibSonomy" vorsah. Während der folgenden Projektphasen präsentierte die Universitätsbibliothek dann ca. 20 Millionen Titeldaten von Büchern und Aufsätzen mittels der Suchmaschinensoftware Primo der Firma Ex Libris. Die Software beinhaltete unter anderem ein eigenes Tagging-System, mit dem die Nutzer Inhalte aus den Online Contents, den Nationallizenzen sowie sämtliche Katalogdaten für eine eigene Verschlagwortung heranziehen konnten. Gefundene Ressourcen konnten somit direkt gespeichert, mit Tags gekennzeichnet und anschließend über diese organisiert und recherchiert werden.

2.2 Erhebungsphase I: Der Vergleich von kontrollierten Vokabularen mit BibSonomy

Die Anbindung des Aleph Online-Katalogs an BibSonomy ermöglichte es, gefundene Titel mittels BibTeX nach BibSonomy zu übertragen. BibTeX ist ein Format zur Übertragung von Metadaten, welches sich im Internet (und nicht nur dort) als einer der wichtigsten internationalen Standards etabliert hat. Im Katalog wurden außerdem bei jedem Titel die in BibSonomy gefundenen Tags als Tag-Cloud angezeigt. Da Tagging-Systeme noch nicht als allen Nutzern bekannt vorausgesetzt werden konnten, wurde der Bekanntheitsgrad der neu implementierten Funktionen durch klassische Werbemedien, gezielte Veröffentlichungen und öffentlichkeitswirksame Aktionen gesteigert. Dies führte zu einer stetig wachsenden Nutzergemeinschaft. Die allgemeine Akzeptanz der Anbindung des Katalogs an das BibSonomy-System erwies sich als sehr hoch. Die Menge der für die Recherche verfügbaren Tags in BibSonomy wuchs in sechs Monaten um weit über 30% auf 97.777 Einträge. Diese Menge an Schlagwörtern wurde vollständig erhoben, um das Nutzerverhalten und die Qualität der Tags auf

Themen

empirisch breiter Basis analysieren zu können.

Für die Analysen der aus unterschiedlichen Zeichenfolgen bestehenden Tag-Daten (samt Nutzungsdaten und Informationen zur jeweils getaggt Quelle) wurde ein speziell zugeschnittenes Analyseverfahren entwickelt. Die Tags wurden bezüglich der verwendeten Schriftsätze bereinigt und von etlichen Sonderzeichen befreit. Durch ein sogenanntes Part-of-Speech Tagging wurde die Datenmenge in Subgruppen geteilt und somit ihre Handhabung stark vereinfacht. Innerhalb und durch dieses Verfahren konnten über einen Wörterbuchvergleich und verschiedene Heuristiken die Wortarten und Grundformen von etwa einem Drittel der Tags ermittelt werden. Ein weiteres Drittel bestand aus Zeichenfolgen, die durch bestimmte Trennzeichen aneinandergereiht wurden, während ein letztes Drittel überhaupt nicht zugeordnet werden konnte. Die überwiegende Mehrheit der Tags lagen in deutscher oder englischer Sprache vor. Für jede der drei Tag-Gruppen wurden individuelle Analysekonzepte entwickelt, um ihre jeweiligen Erkenntnismöglichkeiten bestmöglich auszuschöpfen. Diese werden zusammen mit den ersten Analyseergebnissen im folgenden skizziert.

2.2.1 Lexikalisch existente Begriffe

Die als lexikalisch bereits existent erkannten Begriffe, die zu 80% aus Nomen bestanden, wurden zur Bestimmung ihrer Qualität mit ausgewählten kontrollierten Vokabularen verglichen (RVK, SWD, STW). Die Vokabulare wiesen bezüglich ihrer Struktur erhebliche Unterschiede auf: die SWD war durch die geringste, die RVK hingegen durch die stärkste Hierarchisierung ihrer Konzepte gekennzeichnet. Im Ergebnis zeigte sich, dass sich in absoluten Zahlen deutlich weniger Treffer in stark hierarchisierten Vokabularen erzielen ließen, d.h. nur wenige der verfügbaren Tags fanden sich etwa in der RVK. Dafür repräsentierten diese wenigen Tags prozentual gesehen einen viel größeren Teil dieses Vokabulars, d.h. relativ wenige Tags reichten aus, um bezogen auf die gesamte Konzeptmenge eine hohe Abdeckung zu erreichen. Dieses Ergebnis wurde nicht zuletzt darauf zurückgeführt, dass sich in stark hierarchisierten Vokabularen Strukturwörter (z. B. „Allgemeines“, „Geschichte“, „Deutschland“ usw.) häufig wiederholen. In flach hierarchisierten Vokabularen wie der SWD verhält es sich dagegen genau umgekehrt. Nach verschiedenen Vergleichsschritten und -methoden zeigte sich, dass insgesamt bestenfalls ein Fünftel der Tags dem „Goldstandard“ der von Fachreferenten regelhaft gebildeten kontrollierten Vokabulare entsprach. Ein Rückschluss auf die Qualität der Tags war anhand der Daten allerdings nur bedingt möglich, da für einen seriösen Vergleich und insbesondere die Qualifizierung der inhaltlichen Struktur des Tagging-Systems eine weit höhere Abdeckung als die je nach Vokabular erreichten 3 – 18 % notwendig gewesen wäre.

Vokabular	SWD	STW	RVK
Struktur	flach, wenige Relationen	preferred terms	Hierarchie (Aufstellungssystematik)
N/Deskriptoren	540166	22800	795000

Matching			
N/Treffer	17520	2790	2210
Anteil genutzter Tags (%)	17,92	2,85	2,26
erreichte Abdeckung des Vokabulars, Mehrfachzählungen möglich (%)	7,47	12,23	18,14

Abbildung 1: Ergebnisse des Abgleichs von Tags mit ausgewählten, kontrollierten Vokabularen

2.2.2 Wortketten

Wortketten bestehen aus mehreren Zeichenfolgen, die durch Trennzeichen wie etwa Binde- oder Unterstriche verknüpft sind. Bei diesen Verkettungen stellte sich zusätzlich zur Strukturdimension (siehe Abschnitt „Lexikalisch existente Begriffe“) die Frage, ob es Unterschiede bezüglich der Qualität der Begriffe an verschiedenen Positionen innerhalb der Kette geben und/oder ob sich hinter der Anordnung der Tags eine eigene Form der Hierarchisierung verbergen würde. Deshalb wurden die einzelnen Begriffe bis zur fünften Position abgeglichen und die individuellen Treffermengen ermittelt. Aus den Ergebnissen konnte geschlossen werden, dass die „Qualität“ der Tags (gemessen an dem Kriterium äquivalenten Vorkommens in einem kontrollierten Vokabular) mit der Position innerhalb der Wortkette signifikant steigt. Die intellektuelle Betrachtung der Ketten bezüglich einer inneren Relation (etwa derart, dass der zweite Begriff immer ein Oberbegriff zum ersten wäre) ließ sich hingegen in der Mehrheit der Fälle nicht bestätigen. Zumeist wurde durch die Verkettung das „Problem“ nicht zulässiger Leerzeichen, die im Tagging-System als Trenner fungieren, gelöst.

2.2.3 Lexikalisch nicht erkannte Begriffe

38.337 Tags konnten durch das Part-of-Speech Tagging nicht automatisch zugeordnet werden und mussten intellektuell erschlossen werden. Abkürzungen und Akronyme (Kurzwörter aus Anfangsbuchstaben, z. B. „EDV“), aber auch Neologismen oder neue Wortkombinationen lieferten

Themen

einen äußerst heterogenen Datenbestand, für dessen Analyse ein umfangreiches Kategoriensystem entwickelt wurde. Durch dieses System konnten z. B. eventuelle Hierarchien, die Komplexität und die Funktionen der Tags, der Fachbereich oder ihr semantischer Gehalt eingeschätzt werden. Um die teils aufwändigen Analysen effizienter zu gestalten, wurde das Vorgehen durch den Rückgriff auf die Metadaten der durch ein Tag bezeichneten Ressourcen unterstützt. Beispielsweise handelte es sich oftmals um Artikel aus Fachzeitschriften, dessen Titel evident darüber Aufschluss gaben, um welchen Fach- bzw. Themenbereich es sich handelte. Die Betrachtung der häufigsten, in den Titeln der Zeitschriften auftretenden Begriffe ermöglichte die Zuordnung der Tags zu maximal drei relevanten Fachbereichen. Ohne dieses Verfahren wäre in etlichen Fällen eine inhaltliche Zuordnung der Tags überhaupt nicht möglich gewesen, da die Tagger bei ihrer Bildung keinerlei Konventionen befolgt hatten.

Insgesamt bestätigten bereits die ersten Analysen die Annahme, dass sich die bibliothekarische Sacherschließung von der Erfassung und Organisierung von Quellen mittels Tagging erheblich unterscheidet. Es zeigte sich, dass angesichts der hohen Qualitätsstandards bei einer Verschlagwortung durch Fachreferenten das Potenzial des Taggings nicht in einer ebenso qualifizierten Inhaltserschließung, sondern in seiner ausgeprägten Dynamik und Kreativität gesehen werden musste. Dieser Erkenntnis sollte in den folgenden Projektphasen durch entsprechende Analysen Rechnung getragen werden.

2.3 Erhebungsphase II: Der Vergleich von kontrollierten Vokabularen mit der freien Vergabe im Online-Katalog

In der zweiten Erhebungsphase wurde ein unabhängiges Tagging-System, das Teil der Software Primo ist, allein für die Nutzer der Universitätsbibliothek Mannheim zur Verfügung gestellt. Bei der Beobachtung des Nutzungsverhaltens zeigte sich allerdings schnell, dass der Service nur wenig genutzt wurde. Dies war auffällig, da zum Start des Systems (analog zur ersten Erhebungsphase) erneut umfassende Marketingmaßnahmen durchgeführt wurden. Die Anzahl der vergebenen Tags war dennoch zu gering, um sie mit dem Ziel einer Qualitätsbestimmung sinnvoll untersuchen zu können. Im Unterschied dazu ermöglichte das weiterhin schnell wachsende BibSonomy-System mittlerweile den Zugriff auf 204.674 verschiedene Tags, die insgesamt 2,5 Millionen Mal vergeben wurden. Für die Zielsetzung des Projektes war es deshalb der beste Weg, mit den Analysen der BibSonomy-Tags fortzufahren.

2.3.1 Analyse der inhaltlichen Strukturen im Tagging-System

Unter Einsatz der Software „Semtinel“¹ wurden die Ergebnisse des Matchings von BibSonomy-Tags auf kontrollierte Vokabulare (s. o.) genutzt, um Visualisierungen der Datenstrukturen herzustellen (vgl. Abbildung 2). Hierdurch konnten die Themenbereiche „Informationswissenschaften“, „Bildungswesen“ und „Mathematik“ klar als inhaltliche Schwerpunkte ermittelt werden. Dagegen waren Themen wie Rechtswissenschaft, Volkswirtschaftslehre oder Agrarwissenschaften so gut wie gar nicht vertreten. Außerdem wurde die vertiefende Analyse der lexikalisch nicht erkannten Zeichenfolgen ins Auge gefasst. Die Wortarten der 500 häufigsten Tags aus dieser Gruppe wurden intellektuell zugewiesen und um tagspezifische Wortklassen (etwa Mischformen aus Zahlen und Abkürzungen usw.) ergänzt. Die Längen aller Zeichenketten dieser Gruppe wurden mit der Verteilung in einem natürlichsprachigen Korpus verglichen und die Nutzer- und Nutzungszahlen jedes Tags ausgewertet. Die Ergebnisse zeigten, dass nicht erkannte Begriffe häufig aus Zusammensetzungen lexikalisch ursprünglich getrennter Wörter bestanden und dass lediglich 22 % der Tags von mindestens 6 Nutzern verwendet wurden. Dieses Fünftel der nicht erkannten Begriffe konnte im eigentlichen Sinne als „kollaborativ“ gelten und enthielt besonders aussichtsreiche Kandidaten für allgemeinverständliche Wortneuschöpfungen.

¹ <http://www.semtinel.org/>

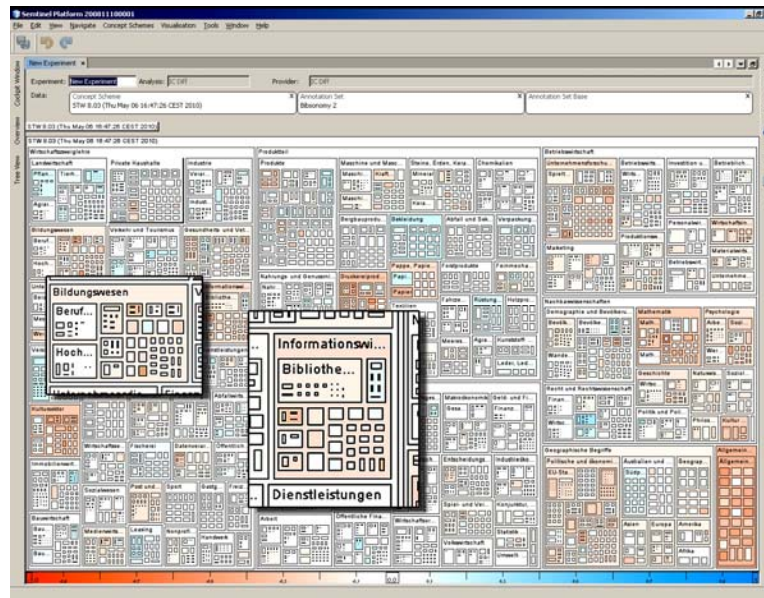


Abbildung 2: Graphische Treemap des Matchings der BibSonomy-Tags auf den Standard Thesaurus Wirtschaft (die roten Bereiche sind häufig vertreten)

2.3.2 Die Nutzbarmachung des Chaos

Die große semiotische Heterogenität der lexikalisch unbekanntem Tags trat zunächst als Beeinträchtigung der Erschließungsqualität in Erscheinung. Es setzte sich aber bald der Eindruck durch, dass diese Tags als neue Themenbezeichnungen oder innovative Wortkombinationen durchaus auch einen bibliothekarisch nutzbaren Wert haben könnten. Die Annahme, dass nach bestimmten Kriterien ausgewählte Tags z. B. zur Erweiterung oder Aktualisierung kontrollierter Vokabulare eingesetzt werden könnten, sollte einer validen Prüfung unterzogen werden. Hierfür wurde ein mehrstufiges Analysemodell entwickelt und eingesetzt (vgl. Abbildung 3). Nach der Anwendung des Part-of-Speech Taggings wurden methodisch relevante Ausschnitte aus der Menge der Tags zunächst auf ihr Vorkommen in weiteren Korpora bzw. in dynamischen Quellen (z. B. Google Search) überprüft, um ihre lexikalische Existenz endgültig auszuschließen. Abkürzungen und Akronyme wurden recherchiert und bei erfolgreicher Identifikation ausgesondert. Tags, die noch nicht als Wortform in einem Standardwörterbuch aufgenommen wurden und auch sonst nirgendwo

auftauchen, konnten dann grundsätzlich als neue Lösung des Erschließungsproblems betrachtet werden. Dies machte sie für eine Analyse besonders interessant. Diese besonderen Tags wurden nun auch im Zeitverlauf bezüglich ihrer Nutzungshäufigkeit sowie der Anzahl ihrer unterschiedlichen Nutzer betrachtet, um Indikatoren für die Allgemeinverständlichkeit bzw. Akzeptanz eines Tags zu erhalten. Das sich aus diesen verschiedenen Teilaspekten ergebende Gesamtbild wurde für jedes Tag zu einer Rangposition aggregiert, die als starkes Indiz für eine „sinnvolle“ und „gute“ Verschlagwortung galt. Schließlich wurden die im Ranking als beste Kandidaten identifizierten Tags unter der Bezeichnung „Neotags“ zusammengefasst und intellektuell unter Einsatz eines nach der ersten Erhebungsphase zusätzlich erweiterten Kategoriensystems evaluiert.

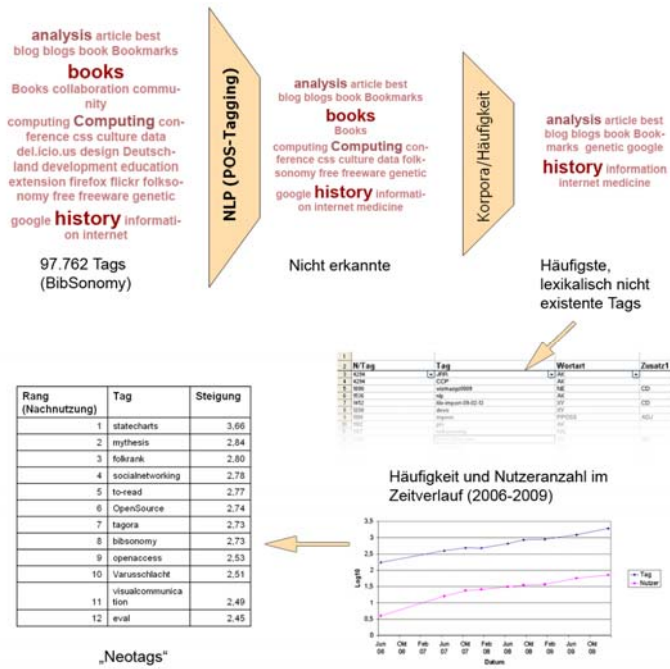


Abbildung 3: Filterprozesse zur Identifikation von Schlagwortkandidaten

2.3.3 Die Messung der Kreativität

Das Kategoriensystem zur Untersuchung der Neotags beinhaltete auch die Evaluation der kreativen Leistung in Tagging-Systemen, d. h. inwiefern die Zeichenfolgen an sich das Potenzial eines neuen und fortgeschrittenen Konzepts in sich tragen. Auf der Basis aller Möglichkeiten, die ein Sprachsystem zur Formung neuer Begriffe bereithält, wurden die Neotags bezüglich ihres kreativen Gehaltes eingeschätzt und in eine zehnstufige Rangliste eingeordnet. Die Analyse der 500 häufigsten Tags ergab, dass die ermittelten Kandidaten häufig aus der Komposition zweier Wörter entstanden waren. Auch die Übernahme neuer Wörter aus dem Titel, dem Abstract oder dem Text der getaggtten Ressource sowie Verkürzungen jeglicher Art traten häufig auf. Ausgesprochene kreative Wortbildungsprozesse wie Analogiebildungen oder Mehrfachkonversionen waren – wenn auch selten – ebenfalls zu beobachten. Der Kompressionsgrad, d. h. die Mehrfachsynthese semantischer Komponenten zu einem einzelnen Ausdruck, trat hingegen fast gar nicht auf. Insgesamt zeigte sich ein nennenswerter Anteil neuer Begriffe und Zeichenfolgen, deren Übernahme in kontrollierte Vokabulare zu erwägen wäre.

2.4 Die Qualität der Verschlagwortung durch Tags: Untersuchungsmethodik

Die mangelhafte Akzeptanz von "Insellösungen", also von in ihrer Nutzung beschränkten Tagging-Systemen, führte zu Datenmengen, die keine seriösen Aussagen über die Qualität der Verschlagwortung durch Tags ermöglichten. Um aber die Aussagekraft der aus den BibSonomy-Daten gewonnenen Erkenntnisse besser einschätzen zu können, wurde ein Vergleich der Datenstrukturen mit einem zweiten großen Tagging-System, nämlich "Citeulike", durchgeführt. Auf dieser Grundlage konnte die Übertragbarkeit der Ergebnisse aus den BibSonomy-Analysen beurteilt werden. Anschließend konnte der Vergleich der Erschließungsmethoden anhand eines Daten-Samples intellektuell evaluiert werden, um zu einer gültigen Empfehlung im Rahmen der bibliothekarischen Sacherschließung zu gelangen.

2.4.1 Vergleich von BibSonomy und Citeulike

Der Vergleich der Datenstrukturen von BibSonomy und Citeulike jeweils für deutsche und englische Begriffe diente der Identifikation von allgemeinen Ähnlichkeiten und Unterschieden zwischen Tagging-Systemen. Als zentrale Vergleichsdimension wurde die grammatikalische Struktur sowie der Anteil lexikalisch noch unbekannter Einträge bestimmt. Dabei zeigte sich, dass die

Verteilung der Wortarten in beiden Systemen weitgehend homogen verläuft. Zudem waren in beiden Systemen eine relativ geringe Anzahl, nämlich 245 identische Einträge, für einen signifikanten Anteil der Nutzung (BibSonomy: 18,5 %, Citeulike: 24,9 %) verantwortlich. Es bestätigte sich die Vermutung, dass es sich bei diesen Tags um in hohem Maße allgemein gültige und leicht verständliche Zeichenfolgen handelte. Die individuelle Bedeutung einzelner Tags unterschied sich hingegen in beiden Systemen: es ließ sich etwa kein Ranking aufstellen, indem z. B. bei BibSonomy und Citeulike ein Tag „semantic“ in gleicher Weise häufiger benutzt wurde, als ein Tag „linguistic“. Bezogen auf die gesamte Datenmenge ließ sich außerdem feststellen, dass sich mit 32.014 (etwa 1/3) der Tags aus BibSonomy in exakt der selben Schreibweise auch in Citeulike wiederfanden.

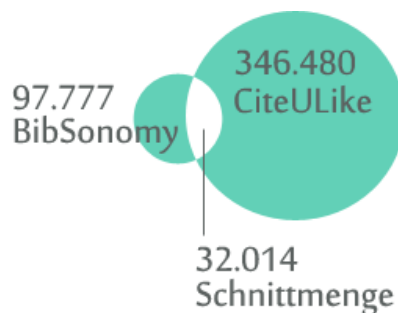


Abbildung 4: Übereinstimmungsgrad von BibSonomy und Citeulike

2.4.2 Vergleich der Struktur der Verschlagwortung von Tagging-Systemen und Fachreferenten

Für den Vergleich der Struktur der Verschlagwortung von Tagging-Systemen und Fachreferenten wurden 100 Tags aus dem BibSonomy-System als zufällige Stichprobe gezogen. Für alle mit diesen Tags verschlagworteten Bücher (im BibTeX als "@book" gekennzeichnet) wurden die relevanten Metadaten (Autor, Titel, weitere Tags usw.) erhoben. Auf diese Weise konnten 1.387 Titel identifiziert werden, die mit hoher Wahrscheinlichkeit bereits bibliothekarisch erfasst wurden, sodass gleichzeitig Nutzertags und Schlagwörter von Fachreferenten verfügbar waren. Die SWD-Schlagwortketten wurden dem Datensatz hinzugefügt, um anschließend beide Begriffsmengen (Tags und Schlagwörter) bezüglich ihrer Wortarten in deutscher und englischer Sprache zu analysieren. Die Auswertung zeigte, dass für die 1.387 Titel in der Stichprobe insgesamt etwa ein Viertel mehr

Themen

Tags als Schlagwörter vergeben wurden. Erwartungsgemäß fanden sich deutlich mehr Schlagwörter von Fachreferenten als Tags im deutschen Wörterbuch bzw. konnten als deutsche Begriffe identifiziert werden. Legte man jedoch englische Wörterbücher zugrunde, drehte sich dieses Verhältnis überraschenderweise nicht etwa um, sondern schwächte sich lediglich ab. Demnach würde selbst ein ausschließlich englischsprachiger Nutzer eher in den Schlagwörtern der Fachreferenten als im BibSonomy-System auf lexikalisch existente, englische Ausdrücke treffen. Weiterhin spiegelten die Präferenzen bei der Nutzung bestimmter Wortarten die häufige Nutzung klassifizierender Adjektive (z. B. „europäisch“, „multinational“ oder „politisch“) durch Fachreferenten bzw. die häufige Nutzung von Verben durch Tagger. Als Fazit wurde die inhaltliche Erschließung durch Fachreferenten unter strukturellen Gesichtspunkten als höherwertig eingestuft.

2.4.3 Vergleich der inhaltlichen Qualität der Verschlagwortung von Tagging-Systemen und Fachreferenten

Nach der Strukturanalyse wurde die Stichprobe einem zweistufigen inhaltlichen Vergleich unterzogen: für jeden Titel wurde im ersten Schritt der Übereinstimmungsgrad beider Verschlagwortungsverfahren berechnet. Knapp 4 % der Titel wurden identisch verschlagwortet, etwa 52 % zeigten Ähnlichkeiten und weitere 42 % entsprachen sich überhaupt nicht. Die Ähnlichkeit der Verschlagwortung durch Tagging und Fachreferenten konnte bezüglich der Identität der Zeichenketten folglich als „gering“ bezeichnet werden.

Im zweiten Schritt wurden alle Titel, die nicht zu 100 % identisch durch Tags und SWD-Schlagwörter erfasst worden waren, auf eine eventuelle semantische Identität bzw. Analogie hin untersucht. Hierzu wurden alle Tags bzw. SWD-Schlagwörter eines Titels in einer Mindmap visualisiert, intellektuell verglichen und den Kategorien „identisch“, „analog“ oder „alleinstellend“ zugeordnet. Auch auf inhaltlicher Ebene bestätigte sich daraufhin die Einschätzung, dass sich die Verschlagwortung mittels Tagging bzw. durch Fachreferenten grundsätzlich unterscheidet: nur etwa 14 % der Konzepte waren semantisch identisch oder wenigstens analog. Im letzten qualitativen Analyseschritt wurde schließlich die Güte der Verschlagwortung nach dem Kriterium „liefert zusätzliche Aspekte bzw. Informationen zu einem Titel“ auf einer Skala zwischen 1 und 5 eingeschätzt. Im Ergebnis war eine annähernde Gleichverteilung zu konstatieren: Zu jeweils einem Drittel war die Verschlagwortung durch Fachreferenten oder durch die Tagger inhaltlich ergiebiger. In einem weiteren Drittel der Fälle war die inhaltliche Erschließung in etwa gleich gut. In etwa 8 % der Fälle war die Erschließung entweder nicht vergleichbar oder in beiden Systemen unzureichend.

2.5 Zusammenfassung der Ergebnisse

Die abschließende Auswertung der Ergebnisse der einzelnen Projektphasen wurde unter den Gesichtspunkten (a) der allgemeinen Akzeptanz des Taggings bei den Nutzerinnen und Nutzern sowie (b) der Qualität der Erschließung im Vergleich zu automatischen Verfahren und zur Sacherschließung durch Fachreferenten durchgeführt. Grundsätzlich ist die Effektivität und Effizienz eines Tagging-Systems ganz wesentlich von der Anzahl und Aktivität der Nutzer abhängig. Erst in zweiter Hinsicht ist die konkrete technische Ausgestaltung eines Systems relevant, das immer das gleiche Set an Basisfunktionen bereitstellen muss: Quellensuche- bzw. Metadatenerstellung, Verschlagwortung, Recherche und Organisation. Entsprechend hängt der Erfolg des Taggings wesentlich davon ab, ob direkte und intuitiv nutzbare Anbindungen der Online-Kataloge an die großen und erfolgreichen Tagging-Systeme angeboten werden. Umgekehrt sollten die dort vorhandenen Daten (soweit möglich) zur Anreicherung der Metadaten im Online-Katalog dienen. Dabei ist auf einen klaren Herkunftsnachweis zu achten, damit den Nutzern eine Einschätzung der Qualität einer Verschlagwortung ermöglicht wird.

Bezüglich der Qualität der einzelnen Erschließungsmethoden haben sich je spezifische Vorzüge und Mängel ergeben: Kontrollierte Vokabulare (1) bieten Sicherheit und Struktur, sind aber bezüglich neuer Entwicklungen und aufkommender Themen träge und statisch. Ihr hoher Formalisierungsgrad sichert eine ebenso hohe Präzision, die allerdings an entsprechende Vorüberlegungen und/oder Kenntnisse bezüglich der Anfrageformulierung geknüpft ist. Dies muss als Barriere für viele Nutzerinnen und Nutzer betrachtet werden, die derartige Kenntnisse oft nicht besitzen. Automatische Verfahren (2) bestechen durch ihre hohe Effizienz und Verfügbarkeit, können aber ebenso wenig selbst Innovationen produzieren, wie sie Ambiguitäten und inhaltliche Besonderheiten blind ignorieren. Tagging-Systeme (3) sind äußerst dynamisch und bezüglich der von ihnen abgedeckten Themen effizient, hängen bezüglich ihres Erschließungsgrades aber stark von den Interessen ihrer Nutzer ab. Ihre Mängel im Hinblick auf die Strukturiertheit und die begriffliche Schärfe des Vokabulars mindert offenbar nicht die inhaltliche Abdeckung einer Ressource, insofern eine gewisse Mindestanzahl von Nutzern diese Ressource getaggt hat.

3 Was vom Tagging übrig blieb: Empfehlungen und Fazit

3.1 Akzeptanz des Taggings

Es kann von einer grundsätzlich hohen Bereitschaft der Nutzer ausgegangen werden, wissenschaftliche Quellen durch Tags zu organisieren und zu erschließen. Diese Bereitschaft hängt allerdings wesentlich davon ab, ob ein System durch entsprechende Datenbestände genügend Ergebnisse liefert, um für eine Recherche reizvoll zu erscheinen. Tagging-Systeme, die als „Insellösung“ auf die Nutzer einer einzelnen Institution beschränkt sind, werden deshalb nicht ausreichend angenommen. Anbindungen an externe Dienste, deren Datenbestand sich aus vielen verschiedenen Quellen und Verknüpfungen speist, erfahren dagegen eine sehr gute Resonanz. Wissenschaftlichen Bibliotheken wird deshalb empfohlen, möglichst schnelle und einfache Verlinkungen zu erfolgreichen Tagging-Plattformen wie BibSonomy oder Citeulike anzubieten. Die Anzeige der dort verfügbaren Daten im eigenen Katalog ist ebenfalls wünschenswert und wird von den Nutzern befürwortet.

3.2 Verfahren zur Analyse von Tagging-Daten

Für die Analyse der äußerst heterogenen Textdaten, die in Tagging-Systemen entstehen, wurden spezifische Verfahren entwickelt und angewendet, die je nach Datenausschnitt und Erkenntniszweck optimiert wurden. Nach erfolgreichen Testläufen wurde der Methodenmix jeweils für größere Datenmengen eingesetzt, um die aus den explorativen Studien gewonnen Hypothesen zu überprüfen. Dieses Vorgehen hat sich als äußerst fruchtbar herausgestellt. Alle durchgeführten Schritte und die daraus gewonnenen Erkenntnisse wurden in diversen Artikeln und Beiträgen veröffentlicht sowie auf zahlreichen nationalen und internationalen Konferenzen vorgestellt, um sie der Wissenschaftsgemeinschaft zur Verfügung zu stellen.²

3.3 Struktur der Tags

Der Vergleich von zwei großen Tagging-Systemen hat große Ähnlichkeiten in der grammatikalischen Struktur der Tagging-Daten ergeben. Es werden mehrheitlich Substantive bzw. Eigennamen zur Erschließung sowie auch Verben zur Organisation der Quellen eingesetzt. Systembedingt kann

² Weitere Referenzen und Publikationen: <http://www.bib.uni-mannheim.de/485.html>

außerdem eine große Menge von Wortkombinationen und Wortneuschöpfungen konstatiert werden, die aus den unterschiedlichsten Beweggründen und für sehr unterschiedliche Zwecke gebildet werden. Nur ein geringer Teil der Tags entspricht den formalen Kriterien kontrollierter Vokabulare. Eine besondere Hierarchisierung der Tags innerhalb eines Tagging-Systems über den Indikator der Häufigkeit der Nutzung hinaus hat sich nicht ergeben. In inhaltlicher Hinsicht hat sich eine klare Dominanz informatiknaher bzw. naturwissenschaftlicher Disziplinen gezeigt, wobei es sich hierbei um systemspezifische Präferenzen handelt. Insgesamt ist eine klare Tendenz zu zunehmender inhaltlicher Diversifikation in den Tagging-Systemen zu erkennen, was mit hoher Wahrscheinlichkeit der wachsenden Akzeptanz durch breitere Nutzergruppen zuzuschreiben ist.

3.4 Qualität der Tags

Bei der Evaluation der Qualität der Tags bestätigte sich die Einschätzung, dass sich die Verschlagwortung mittels Tagging von jener durch Fachreferenten grundsätzlich unterscheidet. Nur ein kleiner Teil der Konzepte wurde in den beiden Systemen semantisch identisch oder wenigstens analog vergeben. Grundsätzlich liegen für eine Ressource fast immer mehr Tags als Schlagwörter vor, die zudem wesentlich häufiger exklusiv im Tagging-System zu finden sind. Diese Tatsache berührt jedoch nicht den inhaltlichen Erschließungsgrad einer Quelle, der sich trotz einer geringeren Anzahl an SWD-Schlagwörtern pro Ressource in beiden Systemen als gleichwertig gezeigt hat. Dennoch ist das Ausmaß der semantischen Abdeckung des Taggings überraschend, da sie der allgemeinen Erwartungshaltung von einer deutlich höheren Qualität der Verschlagwortung durch die professionelle Inhaltserschließung teilweise widerspricht. Diese Erwartung ist zumindest bezüglich der inhaltlichen Dimension zu relativieren.

3.5 Fazit

Der Beitrag des Taggings im Rahmen des bibliothekarischen Kontextes ist vor allem in der ergänzenden Erweiterung der Recherche- und Literaturverwaltungsfunktionalitäten der Online-Kataloge zu sehen. Durch Tagging können diese um eine nutzerorientierte Komponente ergänzt werden und signifikant an Attraktivität gewinnen. Systeme mit einem begrenzten Nutzerkreis sind allerdings zugunsten der Anbindung an etablierte Systeme zu vernachlässigen. Diese können einen parallel existierenden Zugang zu den vorhandenen Ressourcen liefern, der seine Stärken in einer explorativen, eher „unscharfen“ Recherche entfaltet. Somit wird einem speziellen Bedürfnis der Nutzerinnen und Nutzer Rechnung

Themen

getragen, dem durch die voraussetzungsreiche Verwendung von präzisen bibliothekarischen Schlagwörtern nicht immer entsprochen werden kann. Bezüglich der inhaltlichen Abdeckung einer Ressource erfüllt das Tagging jedenfalls die Anforderungen eines Recherchesystems, insofern eine ausreichende Mindestanzahl von Tags vorliegt. Natürlich ist es sehr wichtig, die Nutzerinnen und Nutzer ausreichend darüber zu informieren, dass Tagging – wie alle anderen Erschließungsmethoden auch – keine vollständige Abbildung der verfügbaren Ressourcen leistet. Es stellt lediglich *einen* von verschiedenen Zugangswegen mit spezifischen Besonderheiten und Ergebnissen zur Verfügung. Eine Kombination der Erschließungsverfahren „Fachreferenten“, „Tagging“ und „automatisch“ ist hingegen nur für sehr spezielle Zielsetzungen und als Abfolge von Ergänzungs- und Aktualisierungsschritten sinnvoll. Eine gleichzeitige Integration der Verfahren würde aufgrund ihrer erheblichen Unterschiede eine deutliche Verschlechterung der Erschließungsqualität zur Folge haben. Sinnvoll ist daher eine gleichberechtigte Bereitstellung dieser Zugangswege bei sichtbarer Trennung für die Nutzer. Auf diese Weise können die Vorteile aller Verfahren genutzt werden, ohne sich ihre jeweiligen Nachteile zu eigen zu machen.