„WLEFormativeEAssessment ? “ Zur Qualität der inhaltlichen Erschließung in Tagging-Systemen

Christof Niemann

This document appeared in

Bernhard Mittermaier (Eds.):
eLibrary - den Wandel gestalten
5. Konferenz der Zentralbibliothek
Proceedings of the WissKom 2010: 5. Konferenz der Zentralbibliothek, 08.-11. November 2010, Jülich
Schriften des Forschungszentrums Jülich / Reihe Bibliothek/Library, Vol. 20
Zentralbibliothek (ZB)
Forschungszentrum Jülich GmbH, Zentralbibliothek, Verlag, 2010
ISBN: 978-3-89336-668-2

„WLEFormativeEAssessment ? “ Zur Qualität der inhaltlichen Erschließung in Tagging-Systemen

Christof Niemann

Zusammenfassung

Der Beitrag behandelt die Frage, ob sich die kollaborative Intelligenz der Nutzer von Tagging-Systemen für die inhaltliche Erschließung wissenschaftlicher Ressourcen eignet. Aufbauend auf frühere Forschungsergebnisse kommt ein Modell zur Anwendung, das computerlinguistische bzw. qualitative Methoden zur Analyse der schnell wachsenden Tagging-Systeme „BibSonomy“ und „CiteULike“ kombiniert. Im ersten Schritt zeigt ein Strukturvergleich große Ähnlichkeiten in der grammatikalischen Struktur beider Systeme bei nur geringer Überschneidung der verwendeten Terme. Aussagen zum inhaltlichen Gehalt eines Systems müssen demzufolge systemspezifisch erfolgen. Im zweiten Schritt erfolgt ein explorativer Tiefenvergleich der Erschließungsqualität zwischen Taggern und professionellen Sacherschließern. Im Ergebnis ist ein erhebliches Gefälle bezüglich formaler Qualitätskriterien zugunsten der SWD-Schlagwörter der Fachreferenten zu konstatieren. Gleichzeitig wirkt sich dies jedoch *nicht* auf die inhaltliche Qualität der Erschließung aus, insofern diese am Kriterium semantischer Abdeckung bzw. Güte gemessen wird. Die *inhaltliche* Erschließung durch Tagger und Informationsexperten kann deshalb als gleichwertig beurteilt werden.

Abstract

The article deals with the question of whether the collaborative intelligence of users of tagging systems is suited for the subject indexing of scientific resources. Building on earlier research a model is applied, which combines computational linguistics with qualitative methods to analyze the rapidly growing tagging systems „BibSonomy“ and „CiteULike“. In the first step, a structural comparison shows great similarities in the grammatical structure of both systems with only a slight overlap of the used terms. Statements with regard to the content of a system must therefore be system-specific. In the second step, an explorative in-depth comparison between taggers and librarians in terms of the quality of subject indexing is conducted. As a result, a significant gap regarding formal quality criteria in favor of the SWD-keywords of subject librarians has to be stated. At the same time, however, this has no effect on the quality of subject indexing, inasmuch as it is measured by the criterion of semantic

coverage and quality. In this regard, the subject indexing by taggers and information management professionals can be judged as equivalent.

1 Einleitung

Die inhaltliche Erschließung digitaler Medien durch Tagging ist zur festen Größe für das Wissensmanagement im Internet avanciert. Der *information overload*, der diese Entwicklung hervorgerufen hat, konfrontiert professionelle Informationsversorger mit einer nie da gewesenen Flut von digital publizierten Artikeln und Werken. Um dieser Herausforderung zu begegnen haben verschiedene Dienstleister Tagging in ihren Systemen integriert. In diesem Beitrag werden abschließende Erkenntnisse darüber präsentiert, ob die kollaborative Intelligenz der Tagger für die inhaltliche Erschließung wissenschaftlicher Ressourcen eingesetzt werden kann, während diese von einer intuitiven und individuellen Wissensorganisation profitieren. Die große Freiheit bei der Vergabe von Deskriptoren durch die Nutzer eines Tagging-Systems ist nämlich ein ambivalentes Phänomen. Kundennähe und kreatives Potenzial stehen der großen Menge völlig unkontrollierter Meta-Informationen gegenüber, deren inhaltliche Qualität und Aussagekraft noch unklar ist. Bisherige Forschungsbemühungen konzentrieren sich hauptsächlich auf die automatische Hierarchisierung bzw. Relationierung der Tag-Daten (etwa mittels Ähnlichkeitsalgorithmen) oder auf die Analyse des (Miss-)Erfolgs, den ein Nutzer bei einer Suchanfrage subjektiv erfährt. Aus der Sicht stark strukturierter Wissensorganisation, wie sie Experten z.B. durch die Anwendung von Klassifikationen realisieren, handelt es sich bei den zunächst unvermittelt nebeneinander stehenden Tags also kurz gesagt um Chaos. Das in diesem Chaos aber auch Struktur und wertvolles Wissen als Gemeinschaftsprodukt erzeugt werden kann, ist eine der zentralen Thesen, die im Forschungsprojekt „Collaborative Tagging als neuer Service von Hochschulbibliotheken“[1] untersucht werden.

Zielsetzung des Beitrages

Nachdem zunächst die Analyse der linguistischen Struktur des gesamten Tagging-Systems in Vordergrund stand (vgl. Niemann, 2009), geht es in diesem Beitrag neben dem Vergleich der Strukturen zweier Tagging-Systeme um eine inhaltliche Evaluierung individueller Tags, die vorrangig intellektuell geleistet wurde. Das einzelne Tag wurden nicht nur als leere Zeichenfolge im Kontext anderer

[1] http://www.ub.uni-mannheim.de/485.html

Zeichenfolgen betrachtet, sondern als Träger mit tatsächlichem, semantischen Gehalt. Beide Analyseschritte, der Strukturvergleich und die semantische Analyse der Tags, sind die letzten Bausteine auf dem Weg zu einer abschließenden Beurteilung der Leistungsfähigkeit von Tagging-Systemen im Kontext professioneller Inhaltserschließung.

Forschungsstand zum Thema Tagging

Die Forschung zum Thema Tagging folgt einigen klar erkennbaren Trendlinien. Eine Evaluierung von 73 wissenschaftlichen Studien und Beiträgen der letzten zwei Jahren zeigt einige immer wiederkehrende Themen: *a)* allgemeinen Betrachtungen jeglicher Art, etwa zum Wissensmanagement oder zum generellen Potenzial von Tagging-Systemen, *b)* Gruppen, typische Strukturen und Verteilungen von Tags, Nutzern und/oder Ressourcen, *c)* Interaktionsbeziehungen zwischen Nutzern, Tagging-Systemen und/oder Umweltbedingungen, *d)* Analysen von Netzwerkstrukturen und/oder systemimmanenten Dynamiken sowie *e)* (eher selten ausschließlich, aber in fast allen Beiträgen als methodische Zielsetzung) die Verbesserung und Bereitstellung von Algorithmen und automatischen Verfahren. Die Ausdifferenzierung der Forschung zeigt sich außerdem in diversene „Orchideenstudien“ (wie etwa jene zur Charakterisierung von Angriffen auf Tagging-Systeme oder zu Reaktionen der Nutzer auf Veränderungen einer bestimmten Systemarchitektur). Erwartungsgemäß finden sich auch linguistisch orientierte Annäherungen, die sich etwa mit dem benutzten Vokabular, dem automatischen Auffinden und der Verknüpfung ähnlicher Begriffe oder der Anreicherung derselben mit Metainformationen beschäftigen.

Trotz aller Komplexität und Vielschichtigkeit der Beiträge werden fast immer sprachstatistische Methoden und/oder maschinelles Lernen (Clustering Verfahren, Ähnlichkeitsalgorithmen usw.) zur Analyse eingesetzt. Nur sehr vereinzelt und nachgeordnet finden sich intellektuelle Evaluierungen, noch seltener Befragungen zur Validierung der automatisch generierten Ergebnisse. Dies überrascht zunächst, da Tagging-Plattformen ja ganz wesentlich auf sprachlichen Ausdrücken basieren, die ausschließlich durch den Menschen valide interpretiert und hinsichtlich ihrer semantischen Qualität bestimmt werden können. Trotzdem ist eine Erfassung der Denotation, d.h. der Kernbedeutung und der substantielle Sinnhaftigkeit individueller Tags, bisher nur sehr sporadisch geschehen. Die Schwerpunkte der untersuchten

Studien zeigen insofern deutlich die methodischen Präferenzen der interessierten Fachöffentlichkeit, deren Vertreter im Wesentlichen aus informatiknahen Disziplinen bzw. solchen mit entsprechenden Schnittstellen (etwa digitale Bilbiotheksdienste, eLearning in der Pädagogik usw.) stammen. Vor dem Hintergrund dieser Fehlbestände und aus dem Blickwinkel professionellen Wissensmanagements soll in diesem Beitrag deshalb eine erste Grundlage zur intellektuellen Bestimmung des Potenzials von Tagging-Systemen geschaffen werden. Aufgrund der erheblichen Datenmengen wird dabei ein Methodenmix konzipiert, der automatische Verfahren zur ersten Strukturierung sowie intellektuelle Analysen zur Tiefenerfassung der Erschließungsqualität einsetzt.

2 Datenerhebung

2.1 Datenstruktur

Da es zum Prinzip des Taggings gehört, dass die Nutzer bei der Vergabe von Deskriptoren in keiner Weise eingeschränkt werden, sind die zugrunde gelegten Daten höchst heterogen. Entsprechend werden etliche äußerst eigenwillige Wortkombinationen gebildet, die für einen Außenstehenden kaum zu dechiffrieren sind, z. B. ein Tag wie „WLEFormativeEAssessment“. Um derartige Zeichenfolgen überhaupt einordnen zu können, bedarf es der systematischen Berücksichtigung bestimmter Kontextinformationen (vgl. Niemann, Im Druck). Es hat sich allgemein als sinnvoll erwiesen, jeder auf bestimmte Details abzielenden Untersuchung eine globale Analyse der linguistischen Struktur des gesamten Tagging-Systems vorzuschalten (vgl. hierzu Niemann, 2009). Die Separierung grammatikalischer Einheiten kann z. B. mittels Part-of-Speech-Tagging („POS-Tagging“) realisiert werden. Bei dieser computerlinguistischen Methode wird ein zu analysierender Begriff mit einem Wörterbuch verglichen, um seine Wortart und seine Grundform zu bestimmen. Zusätzlich werden über verschiedene Heuristiken, etwa über die Großschreibung deutscher Substantive u. Ä., Wahrscheinlichkeiten für eine Zuordnung ermittelt, sollte ein direkter lexikalischer Vergleich erfolglos bleiben. Das geschilderte Verfahren wurde auf Datenabzüge der Tagging-Systeme „BibSonomy“[2] (vgl. Abschnitt 2.2) und „CiteULike“[3] angewandt, um ein Vergleich der beiden Portale zu ermöglichen.

[2] http://www.bibsonomy.org
[3] http://www.citeulike.org

2.2 Stichprobenziehung zur intellektuellen Analyse

Als Datenbasis zur Tiefenanalyse der Qualität von Tags in Tagging-Systemen diente ein Abzug aller Nutzereingaben in das erfolgreiche, deutsche (aber nur etwa zu einem Drittel deutsch*sprachige*) Tagging-System „BibSonomy". Verfügbar waren somit alle 97.762 Tags, die bis zum ersten Juli 2009 für Artikel und Publikationen ins BibSonomy-System eingetragen wurden. Für den Vergleich der Verschlagwortungsqualität von Tagging-Systemen und Fachreferenten wurden 100 Tags aus BibSonomy als zufällige Stichprobe gezogen. Alle mit diesen Tags verschlagworteten Titel wurden in Form von BibTeX-Snippets exportiert, um anschließend relevante Metadaten (etwa Autor, Titel, weitere Tags usw.) aus ihnen zu gewinnen und in eine relationale Datenbank zu überführen. Aus dem auf diese Weise erstellten Datensatz wurden nur jene Fälle berücksichtigt, die als Typenbezeichnung „@book" enthielten, weil diese als „Bücher" gekennzeichneten Titel im Gegensatz zu Artikeln („@article") mit hoher Wahrscheinlichkeit bereits bibliothekarisch erfasst wurden. Nach der Filterung befanden sich 1.387 Titel im Datensatz, für die gleichzeitig Nutzertags sowie potenziell Schlagwörter von Fachreferenten verfügbar waren. Um diese zu ermitteln, wurde jeder Titel einzeln in der bibliographischen Datenbank „WorldCat"[4] recherchiert, von der aus eine Weiterleitung auf die Online-Kataloge deutschsprachiger Bibliotheken erfolgte, die den Titel im Bestand führten. Die in den Katalogen gefundenen Schlagwortketten der Fachreferenten wurden ermittelt und dem Datensatz hinzugefügt. Die Datenbank enthielt schließlich zu jedem einzelnen Titel alle Tags aus BibSonomy sowie alle vom Fachreferenten vergebenen Schlagwörter. Die Begriffsmengen wurden durch POS-Tagging jeweils getrennt bezüglich ihrer Wortarten in deutscher und englischer Sprache analysiert.[5]

3 Strukturvergleich von BibSonomy und CiteULike

Der Vergleich der Datenstrukturen von BibSonomy und CiteULike dient der Identifikation von allgemeinen Ähnlichkeiten und Unterschieden zwischen Tagging-Systemen, die als zentrale Kontextbedingungen einer Tiefenanalyse unabdingbar sind. Die Schätzung der Wortarten zeigte, dass die Verteilung in beiden Systemen

[4] http://www.worldcat.org

[5] Die Ergebnisse dieser Wortartenidentifikation sind mit 80% Trefferwahrscheinlichkeit allerdings nicht valide genug, um präzise Aussagen über die grammatikalische Struktur der Daten zu treffen. Allenfalls erlauben sie Schätzungen, aus denen sich Hypothesen ableiten lassen, die anschließend genau geprüft werden müssen.

weitgehend homogen verläuft. Tabelle 1 zeigt außerdem die wichtigsten Kennzahlen des Strukturvergleichs.

	CiteULike	BibSonomy
Gesamtzahl Tags (Types)	346.480	97.777
Verwendungshäufigkeit (Tokens)	5.875.178	873.773
Durschnittliche Verwendungshäufigkeit eines Tags	17	9
Nutzungsanteil der 500 häufigsten Types des eigenen Systems am eigenen System	35,83%	44,37%
Nutzungsanteil der 500 häufigsten Types des anderen Systems am eigenen System	20,6%	19,73%
Anteil der häufigsten, in beiden Systemen vorhandenen Types an der Nutzung	24,91%	18,5%
Vom POS-Tagger (dt.) erkannte Wörter	6,75%	16%
Übereinstimmende Tags	32.014 (9,24%)	32.014 (32,74%)

Tabelle 1: Vergleichstabelle BibSonomy und CiteULike[6]

Im wesentlich größeren CiteULike-System bilden die am häufigsten benutzten 500 Tags 35 % der gesamten Nutzung ab (bei BibSonomy sind es gut 44 %). Betrachtet man diese Abdeckung über Kreuz, so genügen dieselben 500 Tags, um jeweils etwa 20 % des anderen Systems abzubilden. Demnach ist in beiden Systemen eine relativ geringe Anzahl, nämlich 245 identischer Types, für einen signifikanten Anteil der Nutzung (BibSonomy: 18,5 %, CiteULike: 24,9 %) verantwortlich. Es liegt die Vermutung nahe, dass es sich bei diesen Tags um in hohem Maße allgemein gültige und leicht verständliche Zeichenfolgen handelt. Wie wichtig die in beiden Systemen vorhandenen Tags für die Systeme jeweils individuell sind, ist bezüglich ihres Ranges hingegen nicht identisch. So lässt sich etwa kein Ranking aufstellen, indem z. B. bei BibSonomy und CiteULike ein Tag „semantic“ in gleicher Weise häufiger benutzt wird, als ein Tag „linguistic“. Bezogen auf die gesamte Datenmenge lässt sich außerdem feststellen, dass sich mit 32.014 (etwa 1/3) der Tags aus BibSonomy eine erhebliche Menge der Zeichenfolgen in exakt der selben Schreibweise auch in CiteULike wiederfinden (vgl. auch Abbildung 1).[7] Ob dieser Anteil zeitlich und systembedingt konstant bleibt oder dynamisch ist, muss allerdings in weiteren Erhebungen geklärt werden.

[6] Unter „Type“ wird eine individuelle Zeichenfolge, unter „Token“ hingegen die Häufigkeit des Vorkommens eines Types verstanden.

[7] Die semantische Übereinstimmung der Tags beider Systeme liegt aufgrund der unterschiedlichen Schreibweisen der selben Konzepte sogar noch etwas darüber.

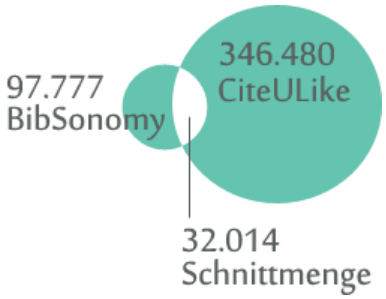

Abb. 1: Schnittmengendarstellung der übereinstimmenden Tags aus BibSonomy und CiteULike

4 Vergleich der Erschließungsqualität von Fachreferenten und Tagging- Systemen

Im Folgenden werden die Ergebnisse der Tiefenanalyse der zufälligen Stichprobe mit 1.387 Titeln vorgestellt, für die jeweils SWD-Schlagwörter von Fachreferenten sowie Tags von Laien (sprich: Taggern) ermittelt wurden.

4.1 Strukturvergleich Tag-Stichprobe/SWD-Schlagwörter

Tabelle 2 gibt einen ersten Überblick über die Mengenverhältnisse der vom POS-Tagger (nicht) als korrekte Zeichenfolge identifizierten Schlagwörter bzw. Tags.

	SWD-Schlagwörter	BibSonomy-Tags
Gesamtanzahl	7678	10136
Deutscher POS-Tagger gefunden	5752 (74,52 %)	3587 (35,39 %)
nicht gefunden	1958 (24,48 %)	6549 (64,61 %)
Englischer POS-Tagger gefunden	3413 (44,45 %)	3330 (32,85 %)
nicht gefunden	4265 (55,55 %)	6806 (67,15 %)

Tabelle 2: Mengenverhältnisse der vom POS-Tagger analysierten Zeichenfolgen

Für die Stichprobe von 1.387 Büchern gilt, dass mit 10.136 Zeichenfolgen die Menge der Tags insgesamt um etwa 25% höher liegt, als die der verfügbaren Schlagwörter (7.678 Begriffe). Außerdem ist auffällig, dass die Trefferquoten für die Fachreferenteschlagwörter sowohl für die deutsche als auch für die englische POS-Tagging Analyse um etwa 40% respektive 11,6% signifikant höher liegen, als jene der BibSonomy-Tags. Für die deutschen Begriffe war dies zu erwarten, da die überwiegende Mehrheit der Schlagwörter nach SWD deutschsprachig ist. Zudem ist selbst das deutsche Tagging-System BibSonomy von englischen Begriffen dominiert, sodass hohe Trefferqoten überraschend gewesen wären. Unerwartet ist jedoch, dass

sich dieses Verhältnis im Falle des englischen POS-Taggings nicht etwa umdreht, sondern lediglich abschwächt. Demnach würde selbst ein ausschließlich englischsprachiger Rezipient eher in den Schlagworten der Fachreferenten als im BibSonomy-System auf lexikalisch existente Ausdrücke treffen. Dass für beide Sprachen höhere Trefferquoten in den Schlagwörtern der Fachreferenten erzielt wurden, ist ein klarer Hinweis auf eine bessere orthographische Qualität dieser Schlagwörter.

4.1.1 Wortartenvergleich für deutsche Begriffe

Für den Wortartenvergleich wurden zunächst ausschließlich Zeichenfolgen berücksichtigt, die als korrekt erkannte Begriffe lexikalisch existent sind. Für deutschsprachige Zeichenfolgen wurden folglich ca. 3/4 der SWD-Schlagwörter und 1/3 der Bibsonomy-Tags untersucht (vgl. Tabelle 2). Für die Ergebnisse des deutschen POS-Taggers zeigt sich, dass bei beiden Verschlagwortungsmethoden in etwa gleich viele Nomen und Eigennamen zum Einsatz kommen. Bei den weniger häufig verwendeten Wortarten wird dagegen ersichtlich, dass die Fachreferentenverschlagwortung im Vergleich zu der BibSonomy-Verschlagwortung deutlich mehr Adjektive (11 %, BibSonomy: 2 %) und Zahlen (5 %, BibSonomy: 1 %) aufweist (vgl. Abbildung 2). Dieser Unterschied ist auf die systematische Verschlagwortung durch die Fachreferenten zurückzuführen, die wesentlich öfter Begriffe mit übereordnender, klassifizierender Funktion (wie z. B. „europäisch“, „multinational“, „künstlich“, „natürlich“, „politisch“, „strategisch“, „sozial„ usw.) einsetzen. Die häufiger auftretenden Zahlen in den Verschlagwortungssystemen der Bibliotheken lassen sich hauptsächlich auf zwei Ursachen zurückführen: zum einen werden Schlagwörter von Zahlen begleitet, wenn sie in einer bibliotheksinternen Zahlensystematik organisiert sind. Zum anderen wird auch das Schlagwort „Geschichte“ oftmals mit einer Angabe des entsprechenden Jahres oder Jahrhunderts versehen. Die relativ häufige Verwendung von Zahlen zur chronologischen Einordnung der Inhalte ist folglich für die Struktur der Wortartenverteilung an dieser Stelle verantwortlich. Umgekehrt gilt dies für die Verben: Ihr fehlender Anteil innerhalb der Schlagwörter der Fachreferenten erklärt sich dadurch, dass Verben in der professionellen Sacherschließung nicht eingesetzt werden.

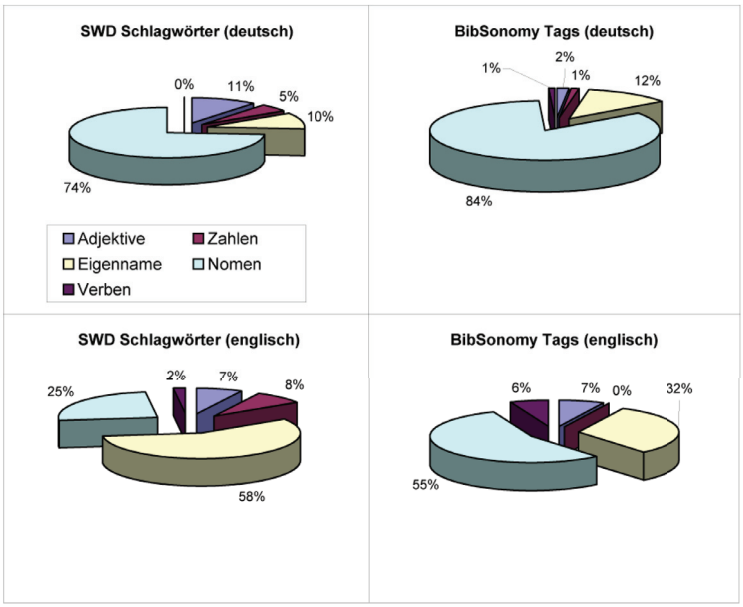

Abb. 2: Wortartenvergleich zwischen Fachreferentenschlagwörtern (SWD) und der Zufallsauswahl aus BibSonomy (TAG) aufgrund der deutschen und englischen POS-Tagger-Analyse

4.1.2 Wortartenvergleich für englische Begriffe

Im Falle der englischsprachigen Zeichenfolgen wurden die Hälfte der SWD-Schlagwörter und 1/3 der Bibsonomy-Tags untersucht (vgl. Tabelle 2). Die Schlagwörter der Fachreferenten wurden gegenüber den Tags aus Bibsonomy demnach fast doppelt so häufig zu den Eigennamen gezählt (vgl. Abbildung 2). In der Summe ergeben die jeweiligen Mengen der für den POS-Tagger schwer zu differenzierenden Wortarten „Nomen“ und „Eigennamen“ hingegen für beide Begriffsmengen einen fast identischen Wert (81 % bzw. 83 %). Der Unterschied ist somit lediglich der Methode des Analysewerkzeuges geschuldet. Auffällig ist, dass in beiden Systemen jeweils 7 % der Zeichenfolgen als englische Adjektive identifiziert werden konnten. Der etwas geringere Anteil war im Falle der Verschlagwortung durch deutsche Fachreferenten zu erwarten, nimmt sich bei den Tags jedoch als Besonderheit aus. Englischsprachige Nutzer von Tagging-Systemen scheinen demnach häufiger als deuschsprachige bei der inhaltlichen Erschließung auf Adjektive zurückzugreifen.

Führt man die Ergebnisse der Adjektivhäufigkeiten für beide Verschlagwortungs-systeme sowie für deutsch und englisch zusammen, so lässt sich im Falle der

SWD-Schlagwörter eine hohe Strukturiertheit erkennen, während im Taggingsystem ein ausgeprägter Wortpluralismus herrscht. Tabelle 3 zeigt die Treffer jeweils für Types und Tokens an. Während im Schlagwortsystem der Fachreferenten ein Adjektiv durchschnittlich 5 mal verwendet wird, liegt die Verwendungshäufigkeit für Adjektive in BibSonomy etwas unter 2.

	SWD DE	SWD EN	Bibsonomy EN	Bibsonomy DE
Type	125	56	2370	682
Token	551	244	4070	1067

Tabelle 3: Vergleich der Type-Token-Verhältnisse der Adjektive

5 Inhaltliche (qualitative) Analyse

5.1 Identität der Zeichenfolgen

Die zugrunde liegenden Daten wurden für die Betrachtung auf inhaltlicher Ebene um Dopplungen und um fehlerhafte Datensätze reduziert, wodurch die Datenbasis auf 1288 Bücher sank.[8] Bei diesen wurde mittels automatischer Abfrage geprüft, welche identischen Zeichenfolgen sowohl als SWD-Schlagworte sowie auch als Tag vorhanden waren. Hierbei zeigte sich, dass 48 Bücher (3,72 %) exakt dieselben Schlagwörter bzw. Tags in beiden Systemen erhalten hatten. 17 Bücher (1,32 %) hatten eine hohe Übereinstimmung, in 296 (22,96 %) Fällen gab es eine mittlere, in 377 (29,25 %) Fällen eine geringe Übereinstimmung. In 550 (42,67%) Fällen gab es zwischen den Tags aus BibSonomy und den SWD-Schlagwörtern überhaupt keine Übereinstimmungen.

5.2 Semantische Analyse der Stichprobe

Semantische Identität und semantische Analogie

Der erste inhaltliche Analyseschritt bestand darin zu prüfen, welche der Tags bzw. Schlagwörter in semantischer Hinsicht identisch sind bzw. analog gebraucht werden können. Hierzu wurden zunächst die 550 Tags bzw. Schlagwörter, die eine Übereinstimmung von 0 aufwiesen, mit dem Mindmapping-Programm „Freemind“

[8] Die Mehrzahl der nicht berücksichtigten Titel bestand hierbei aus Büchern, die in keiner deutschen Bibliothek mit Schlagworten versehen wurden.

den Kategorien „semantisch identisch, „thematisch analog“ oder „alleinstellend“ zugeordnet (vgl. Abbildung 3). Als „semantisch identisch“ (7 % der Zeichenfolgen) wurden alle Ausdrücke betrachtet, die im Sinne des allgemeinen Sprachverständnisses den gleichen Inhalt repräsentieren (z. B. „Kongress“ und „Congress“, „Wirtschaftsentwicklung“ und „Economic development“ oder „künstliche Intelligenz“ und „ai“). „Thematisch analog“ (7 % der Zeichenfolgen) waren Ausdrücke, die thematisch in etwa den gleichen Inhalt repräsentieren oder zumindest ihren Hauptaspekt teilten. Diese Kategorie wurde tendenziell „wohlwollend“ zugeordnet, d. h. es wurde eher nach Gemeinsamkeiten als nach trennenden Eigenschaften gesucht (z. B. „Soziale Ungleichheit“ und „Social Equality“, „Geschichte Anfänge bis 800“ und „history“ oder „Wirtschaftslage“ und „Wirtschaftsstruktur“). Als „alleinstellend“ (BibSonomy: 58 %, Schlagworte: 28 %) wurden schließlich Tags bzw. Schlagwörter verstanden, die weder als identisch noch als thematisch analog zu einem Tag bzw. Schlagwort betrachtet werden konnten (z. B. „Kind, Denken, Kognitive Entwicklung„ und „devo, object, permanence“, Schriftguss, Technik und „mediengeschichte, www01, wwwbook, wwwkap2“ oder „International economic relations, International trade“ und „Wohlfahrtstheorie, Theorie_der_Wirtschaftspolitik, Methoden_und_Techniken_der_Volkswirtschaft“).

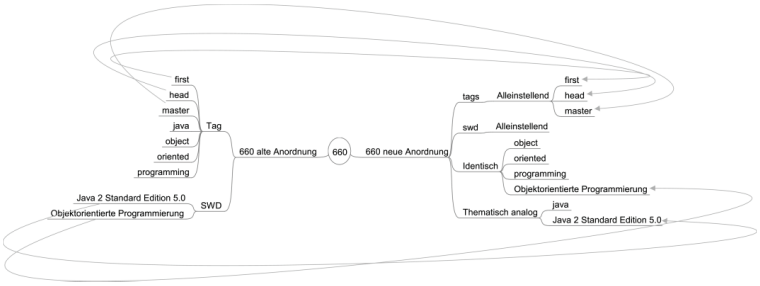

Abb. 3: Freemind Beispiel

Auf semantisch-inhaltlicher Ebene bestätigte sich somit die Einschätzung, dass sich die Verschlagwortung mittels Tagging bzw. durch Fachreferenten grundsätzlich unterscheidet: nur etwa 14 % der Konzepte waren semantisch identisch oder wenigstens analog. Mit knapp 60 % steuerten die Tags nicht nur absolut eine höhere Anzahl, sondern auch fast doppelt so viele alleinstellende Konzepte zur Datenmenge bei. Zu beachten ist, dass der Terminus „alleinstellend“ keinerlei Aussage über den inhaltlichen Mehrwert einer Zeichenfolge beinhaltet, mithin hier rein quantitativ gemeint ist.

Güte der inhaltlichen Erschließung

Im zweiten 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.

6 Resümee

In den vorgestellen Analysen sollte der Frage nachgegangen werden, ob sich die kollaborative Intelligenz der Nutzer von Tagging-Systemen für die inhaltliche Erschließung wissenschaftlicher Ressourcen eignet. Die präsentierten Ergebnisse, die auf einen Strukturvergleich der Tagging-Systeme BibSonomy und CiteULike sowie auf einer Tiefenanalyse eines Datensamples aus BibSonomy und entsprechenden SWD-Schlagworten deutscher Universitätsbibliotheken basieren, legen den Schluss nahe, dass bezüglich der semantischen Abdeckung keine Qualitätsunterschiede zwischen beiden Verschlagwortungssystemen zu ermitteln sind. Dies ist insofern überraschend, als die allgemeine Erwartungshaltung von einer deutlich höheren Qualität der Verschlagwortung durch die professionelle Inhaltserschließung ausgeht. Auch wenn im Endergebnis eine vergleichbare Abdeckung erreicht wird, muss dieser Schluss allerdings an verschiedenen Stellen relativiert werden. Insbesondere bezüglich der Präzision einer Recherche ist davon auszugehen, dass die nach strengen Regeln gebildeten SWD-Schlagwörter wesentlich bessere Ergebnisse erzielen. Allerdings wird dies mit der Notwendigkeit erkauft, dass ein Nutzer das passende Schlagwort kennen muss, um eine Recherche überhaupt korrekt ausführen zu können. Demgegenüber ist ein Tagging-System diesbezüglich erheblich „fehlertoleranter“ und ermöglicht ein ausgesprochen exploratives und dynamisches Suchen. Verschiedene Schreibweisen, Ausdrucksformen usw., die in Recherchesystemen zumeist als störender „Noise“ empfunden werden, können aus diesem Blickwinkel die notwendige Unschärfe liefern, die ein Recherchierender ohne exakte Vorstellung bzw. Explikation seines Suchbedürfnisses benötigt, um überhaupt zu Ergebnissen zu kommen.

Die erläuterten Befunde können bezüglich der grammatikalischen Struktur der Wortmengen zunächst über die beiden untersuchten Tagging-Systeme hinweg als gültig bezeichnet werden. Für die inhaltliche Struktur müssen hingegen individuelle Analysen erfolgen, da sich die einzelnen Systeme bezüglich ihrer Zielpublika und ihrer generellen Ausrichtung unterscheiden. Auch der Umfang der semantischen Analysen bedarf einer Ausweitung, um die hier vorgestellten Schlussfolgerungen auf breiterer Ebene zu evaluieren. Als fruchtbar hat sich dabei der methodische Ansatz erwiesen, die großen und schnell wachsenden Datenmengen in den Tagging-Systemen durch verschiedene Verfahren zu bearbeiten, die nach der Strukturierung die gezielte Selektion relevanter Datensegmente für eine (aufwändigere) inhaltlich-semantische Analyse ermöglichen.

Literatur

Niemann, C. (2009). Intelligenz im Chaos: Collaborative Tagging als neue Form der Sacherschließung. In: Barth, R., (Hrsg.), *Die Lernende Bibliothek 2009 : Wissensklau, Unvermögen oder Paradigmenwechsel?* , Jahrgang 33 von *Churer Schriften zur Informationswissenschaft*, Seite 59–66. Chur.

Niemann, C. (Im Druck). Von Tags und Neotags: Die intellektuelle Evaluierung von Tagging-Systemen. In: Danowski, P. und Bergmann, J., (Hrsg.), *Handbuch Bibliothek 2.0*, Bibliothekspraxis. Saur.