

Top-down versus Bottom-up: Wissensorganisation im Wandel. Von der traditionellen Wissenserschließung zu Folksonomies

Josef Herget und Sonja Hierl

Swiss Institute for Information Research
Hochschule für Technik und Wirtschaft (HTW) Chur,
Ringstrasse/Pulvermühlestrasse 57,
CH-7004 Chur, Schweiz
josef.herget@fh-htwchur.ch
sonja.hierl@fh-htwchur.ch

Abstract: Der vorliegende Beitrag gibt einen Überblick über den sich momentan vollziehenden Wandel der Wissenserschließung und –organisation. Nach einer Einführung in die klassischen Methoden und Instrumente werden neue Ansätze vorgestellt, die sich durch eine starke Partizipation der Nutzer im Erschließungsprozess auszeichnen. Auf Grundlage einer Abwägung der Herausforderungen und Chancen wird ein Ausblick auf vorhandene Potenziale gegeben, die in einer Kombination traditioneller und neuartiger Verfahren liegen.

1 Wissenserschließung im Wandel

Die Notwendigkeit eines intuitiven und unmittelbaren Zugriffs auf Information und Wissen ist in unserem durch Informationsfülle gekennzeichneten Zeitalter unumstrittener Konsens. Die sich für diesen Zweck derzeit herausbildende „Weltbibliothek“ bedarf jedoch zukünftig neuer Formen und Regeln der Wissensorganisation, die traditionellen Herangehensweisen und Methoden der Wissenserschließung und –organisation sind begrenzt und können den aktuellen Herausforderungen nicht mehr stand halten.

1.1 Traditionelle Verfahren der Wissenserschließung

Die Rahmenbedingungen in der Welt der Wissenserschließung haben sich in den letzten Jahren grundlegend verändert. Waren früher referenzierte Printquellen der Ausgangspunkt der Wissenserschließung und –organisation (z. B. in Form dokumentarischer Bezugseinheiten), ist nunmehr ein Wandel zu digitalen Primärdokumenten festzustellen.

Die textlichen Repräsentationen werden zudem vermehrt durch multimediale Elemente (Audio, Video und Bilder) ergänzt. Während bei nur in Druckformaten vorliegenden Quellen eine inhaltliche und intellektuelle präkoordinative Erschließung notwendig ist, um die anschließende Wiederauffindbarkeit durch Referenzierung zu gewährleisten, lassen sich elektronische Volltexte auch postkoordinativ während des Information Retrieval Prozesses erschließen. Hierbei werden zunehmend automatisierte auf statistisch und linguistisch basierten Verfahren eingesetzt (beispielsweise bei der Volltextsuche von Websuchmaschinen). Somit lassen sich folgende Tendenzen in der Wissenserschließung feststellen:

- Von Print- (bzw. analogen) Inhalten zu digitalen Inhalten
- Von manuellen Verfahren zu automatisierten Verfahren
- Von präkoordinativen Verfahren zu postkoordinativen Verfahren der Wissenserschließung

Traditionelle Ansätze zur Erschließung und Organisation von Wissen und Information folgen in der Regel einem Top-Down Ansatz, bei dem die erschließende Instanz unterschiedliche Methoden und Instrumentarien anwendet [La86], um anschließend das erschlossene Wissen in geeigneter Repräsentation den Informationssuchenden zur Verfügung zu stellen. Viele klassische Methoden der Wissenserschließung basieren auf einem solchen Top-Down Ansatz, beispielsweise [KR02], [KSS04]:

- Inhaltsverzeichnisse und Gliederungssysteme
- Register
- Klassifikationen
- Thesauri
- Indices (Verschlagwortung)
- Metadatenschemata
- Abstracts bzw. Kurzreferate

Diese Vorgehensweise hat sich im Laufe der vergangenen Jahrzehnte in Bibliotheken, Archiven und Dokumentationen sowie zur Erschließung größeren Datensammlungen, wie beispielsweise Repositorien und Datenbanken durchgesetzt und bewährt. Die Vorteile des Ansatzes sprechen für sich: Die Erschließung erfolgt anhand konsistenter Vokabularien und Beschreibungen, sie wird auf einem einheitlichen Abstraktionsniveau vorgenommen und die Einhaltung von festgelegten (Mindest-) Standards bei der intellektuellen Erschließung wird sichergestellt.

Als wesentliche Nachteile sind die hohen Kosten der zumeist vollständig intellektuellen Erschließung durch Fachleute und die teilweise fehlende Anpassungsfähigkeit der verwendeten Terminologien anzusehen. Voß konstatiert gar, dass Bestrebungen, die Ressourcen des Internet mit Hilfe der klassischen und traditionellen Methoden zu erschließen, bisher gescheitert sind und hierfür neuere Instrumente eingesetzt werden müssen, beispielsweise postkoordinative Verfahren wie Googles Ranking-Algorithmus [Vo07].

Als momentan allgegenwärtiges Beispiel der oben erläuterten Tendenzen lässt sich die Entwicklung vom so genannten Web1.0 zum Web2.0 betrachten. Dieser Wandel bringt hinsichtlich der Wissenserschließung und –organisation eine Reihe neuer Chancen und Möglichkeiten, im Gegenzug jedoch auch neue Voraussetzungen und Herausforderungen mit sich. Im Zuge der sich rapide verändernden Repräsentationsformen und Beschreibungen von Information und Wissen lassen sich diese Herausforderungen durch den Einsatz der oben erläuterten klassischen Ansätze alleine nicht mehr bewältigen [Bo02]. Vokabularien, umgangssprachliche Beschreibungen sowie auch fachliche Bezeichnungen unterliegen einem steten Wandel mit zunehmend kürzerer Halbwertszeit. Zum Zeitpunkt der Erschließung lässt sich für einige Wissensdomänen immer weniger absehen, wie die Nutzer digitaler Informationsquellen beispielsweise in nur in vier bis fünf Jahren nach den erschlossenen Quellen recherchieren und welche Suchtermini sie zum Auffinden verwenden werden. Hinzu kommt eine immer stärker werdende Diversifizierung von Nutzergruppen mit unterschiedlichen Anwendungsbereichen und sich herausbildendem Domänenwissen, was eine nachhaltige Auswirkung auf die verwendeten und üblicherweise eingesetzten Vokabularien hat [We85].

Eine weitere Herausforderung im Zeitalter des Web2.0 ist die verteilte und rasante Entwicklung von Daten, Informationen und digitalen Ressourcen sowie deren weltweite Verfügbarkeit über das Internet. Eine intellektuelle Erschließung nach dem Prinzip des Top-Down Ansatzes würde den Rahmen der sich herausbildenden Komplexität und des notwendigen Aufwandes schlichtweg übersteigen und zu immensen Kosten führen. Weiterhin wäre die Notwendigkeit einer Standardisierung durch regulierende Instanzen im Internet die Folge, was sowohl aus ethischen als auch rechtlichen Gesichtspunkten kaum realistisch erscheint und zudem einen immensen zeitlichen und logistischen Aufwand zur Folge hätte.

2 Neue Konzepte der Wissenserschließung

2.1 Tagging, Folksonomies und Ontologien

Aktuelle Entwicklungen im Web2.0 zeigen das Zusammentreffen sehr großer Nutzergruppen in kürzester Zeit, die eine Erschließung von digitalen Ressourcen anhand der freien Vergabe von Schlagworten – so genannten Tags – vornehmen [Vo07].

Der Erfolg einer derartigen Erschließung zeigt sich auf diversen sozialen Plattformen wie Delicious [De07], CiteUlike [Ci07] oder Flickr [Fl07]. Die zugrunde liegende Idee des Ansatzes ist ebenso simpel und einfach wie ihre Durchführung: Registrierte Nutzer können unterschiedlichsten Informationsobjekten und Daten Begriffe zuordnen, anhand derer sie sich bei einer anschließenden Suche wieder auffinden lassen. Beispielsweise wird die Organisation von Linksammlungen bzw. Bookmarks auf Delicious auf diese Weise gelöst. Sofern sich ein Nutzer nicht aktiv gegen die „Veröffentlichung“ seiner beschlagworteten Links auf der Plattform entscheidet, sind sowohl Links als auch die zugeordneten Begriffe für die gesamte Nutzergemeinschaft sichtbar und frei zugänglich [De07]. Bei der Vergabe der Schlagworte werden Nutzer unterstützt durch den Vorschlag von Begriffen, die mit den eingegebenen Zeichen übereinstimmen und bereits für den gleichen Link vergeben wurde. Die einer URL zugeordneten Begriffe werden anschließend in so genannten „Tag Clouds“ oder Wortwolken dargestellt, wobei häufiger verwendete Schlagworte (bei einer mehrfachen Erfassung der gleichen URL durch verschiedene Nutzer) beispielsweise durch einen größeren Schriftgrad prominenter dargestellt werden als Begriffe, die lediglich von ein oder zwei Anwendern zugeordnet wurden. Diese Art der Erschließung wird mit dem Begriff der „Folksonomies“ bezeichnet, abgeleitet von den englischen Begriffen Folk (Volk) und Taxonomy (Taxonomie). Die Verschlagwortung wird im Gegensatz zur klassischen Indexierung nicht von Fachexperten und auch nicht immer unter Verwendung kontrollierter oder einheitlicher Vokabularien vorgenommen, sondern durch die Nutzergemeinschaft der entsprechenden Plattform. Als weiteres Konzept, das zur Wissensordnung eingesetzt werden kann und das seit dem Aufkommen des Semantic Web verstärkt diskutiert wird, ist die Ontologie, die Gruber definiert als „a specification of a conceptualization“ [Gr93].

Beim direkten Vergleich einiger Kriterien von Folksonomien und Ontologien kommen Herzog et al. [HLH 07] zu untenstehender Einschätzung (Abb. 1), aus der sich die unterschiedliche Eignung der beiden Konzepte je nach Zielsetzung konkludieren lässt: Während Folksonomies breiter und weniger spezifisch sind, jedoch schneller und günstiger in der Erstellung als Ontologien, werden letztere durch Experten umgesetzt und sind hierdurch zielgerichteter, genauer aber auch teurer.

	Folksonomy	Ontology
Structure	flat	hierarchical structure
Creation	by users during the act of using the system	by ontology experts at a given time
Synonyms	no synonym control	synonym control possible
Precision	low precision	high precision
Flexibility	high flexibility	low flexibility
Creation Cost	low, rated by users	high, created by experts
Change	highly dynamic, changes constantly	rigid, often has to be recreated to accommodate change
Usability	no expertise required	requires proficiency in handling
Vocabulary	users vocabulary	experts vocabulary
Scalability	works better in a large scale	works better in a small scale

Abbildung 1: Vergleich von Folksonomies und Ontologien (entnommen aus [HLH 07])

2.2 Chancen und Herausforderungen neuer Erschließungsmethoden

Im Gegensatz zu den traditionellen Erschließungsformen ist das Vokabular in Social Software oder Web2.0 Anwendungen weder kontrolliert, reguliert, noch standardisiert, was den zugrunde liegenden Bottom-Up Ansatz verdeutlicht. Auch bei dieser Vorgehensweise sind die Vorteile offensichtlich: Eine sehr zeitnahe und schnelle intellektuelle Erschließung großer Datensammlungen wird ermöglicht, es erfolgt der Einbezug sowohl des Produzenten von digitalen Ressourcen als auch der Endanwender in den Indexierungsprozess. Nicht zuletzt entsteht ein dynamischer Indexierungs- und Erschließungsprozess, bei dem sich das Erschließungsvokabular im Laufe der Zeit mit der Veränderung von Terminologien gemeinsam an neue Vokabularien anpassen lässt und somit nicht obsolet wird. Diverse Forschungsansätze haben gezeigt, dass Folksonomies ein großes Potenzial als Grundlage für Metadaten zukommen kann [AD06]. Andererseits entstehen verschiedene Probleme, vor allem durch die fehlende Qualitätskontrolle und durch das Fehlen eines gemeinsamen Standards als Basis der Erschließung. Hinzu kommt die Problematik der erschließenden Nutzer, die nicht notwendigerweise Experten auf dem Gebiet sind, in dem sie Informationsquellen beschlagworten. Dennoch bekommt ihr Beitrag das gleiche Gewicht wie die Beiträge von Experten, die sich mit Inhalt und Kontext der erschlossenen Ressourcen tatsächlich gut auskennen.

3 Zukünftige Entwicklungen

Die Entwicklung der klassischen bzw. neuen Methoden und Instrumente der Wissenserschließung und -organisation findet derzeit in unterschiedlichen Communities statt, was eine Parallelentwicklung zur Folge hat, sofern nicht aktiv eine Zusammenführung und Kombination der Instrumentarien vorgenommen wird. Die Vor- und Nachteile der Verfahren zeigen ihre unterschiedliche Eignung je nach Zielsetzung und Rahmenbedingung deutlich auf. Als großes Potenzial lassen sich unterschiedliche Kombinationsmöglichkeiten von alten und neuen Methoden identifizieren, bei denen jeweils die Vorteile beider Ansätze ausgeschöpft werden und somit Synergieeffekte erzielt werden können. Beispielsweise lassen sich Folksonomies als Grundlage für die automatisierte Ontologiebildung nutzen, die eine qualitative Erschließung von digitalen Informationsressourcen auf Basis sozialen Taggings ermöglicht. Die methodische Vorgehensweise könnte sich hierbei bewährter Mittel wie beispielsweise des Latent Semantic Indexings (LSI) [AS06], [LFL98] bedienen. Bei diesem Verfahren werden auf Grundlage statistischer und linguistischer Verfahren durch Terminengewichtungen und -häufigkeiten innerhalb eines Dokumentenraums Begriffe indirekt in Verbindung gebracht mit zugrunde liegenden Konzepten.

Die Kombination traditioneller Erschließung mit neuartigen Ansätzen des gemeinsamen dynamischen Taggings einer Nutzergemeinschaft erlaubt die Bewältigung zahlreicher Herausforderungen und die Verbesserung der Qualität von Informationserschließung und Wissensrepräsentation:

So lassen sich hierdurch kostengünstig große Datenmengen bewältigen. Weiterhin entwickeln sich die verwendeten Vokabularien bei Veränderungen im Sprachgebrauch dynamisch weiter und die unterschiedlichen Perspektiven (Expertensicht und Nutzerperspektive) werden dargelegt. Die Kombination der Methoden kann auf unterschiedliche Weise vorgenommen werden beispielsweise durch die Verknüpfung traditioneller und neuer Ansätze oder die Anwendung neuer Methoden auf Grundlage von mit traditionellen Instrumenten erschlossenen Ressourcen.

Literaturverzeichnis

- [AS06] Alesso, H. P.; Smith, C. F.: Thinking on the Web: Berners-Lee, Gödel, and Turing, Hoboken, 2006.
- [AD06] Al-Khalifa, H. S. and Davis, H. C.: FolksAnnotation: A Semantic Metadata Tool for Annotating Learning Resources Using Folksonomies and Domain Ontologies. Proceedings of the Second International Conference on Innovations. In Information Technology. IEEE Computer Society, Dubai, UAE. 2006.
- [Bo02] Bohner, M.: Instrumente zur Metadatenerschließung im WWW am Beispiel der Online-Schlagwortnormdatei. Masterarbeit an der Universität Konstanz im Fachbereich Information Engineering. Konstanz, 2002.
- [Ci07] CiteULike: Social Filing System for academic papers, <http://www.citeulike.org>.
- [De07] Delicious: Social Bookmarking Website, <http://del.icio.us>.
- [Fl07] Flickr: Online Foto Management and Sharing Platform, <http://flickr.com/>.
- [Ga05] Gaus, W.: Dokumentations- und Ordnungslehre. 5. Auflage, Berlin/Heidelberg, 2005.
- [HLH07] Herzog, C.; Luger, M.; Herzog, M.: Combining Social and Semantic Metadata for Search in a Document Repository, In: Van Damme, C.; Hepp, M.; Siorpaes, K. (Hrsg.): Bridging the Gap between Semantic Web and Web 2.0 (SemNet 2007), Seiten 14-21, 2007.
- [KR02] Kiel E., Rost F.: Einführung in die Wissensorganisation: Grundlegende Probleme und Begriffe. Würzburg: ERGON Verlag, 2002.
- [KSS04] Kuhlen, R.; Seeger, T.; Strauch, D. (Hrsg.): Grundlagen der praktischen Information und Dokumentation. KG Saur, München, 2004.
- [La86] Lancaster, F.: Vocabulary Control for Information Retrieval. 2. Auflage, Information Resources Press, Arlington, 1986.
- [Le06] Lehnerer, S.: Wissensorganisation in der Online-Enzyklopädie Wikipedia. Magisterarbeit im Fach Informationswissenschaft an der Universität Regensburg, 2006.
- [LFL98] Landauer, T. K., Foltz, P. W., & Laham, D.: Introduction to Latent Semantic Analysis. Discourse Processes, 25, Seiten 259-284, 1998.
- [Vo07] Voß, J.: Tagging, Folksonomy & Co – Renaissance of Manual Indexing? In: Osswald, A.; Stempfhuber, M.; Wolff, C. (Hrsg.): Open Innovation. Proceedings des 10. Internationalen Symposiums für Informationswissenschaft in Köln, S. 243-254, 2007.
- [We85] Wersig, G.: Thesaurus-Leitfaden. Eine Einführung in das Thesaurus-Prinzip in Theorie und Praxis. 2. ergänzte Auflage, KG Saur, München, 1985.