ZBW *Publikationsarchiv*

Publikationen von Beschäftigten der ZBW – Leibniz-Informationszentrum Wirtschaft
Publications by ZBW – Leibniz Information Centre for Economics staff members

Kasprzik, Anna

Article

Voraussetzungen und Anwendungspotentiale einer präzisen Sacherschließung aus Sicht der Wissenschaft

ABI Technik

Suggested Citation: Kasprzik, Anna (2018) : Voraussetzungen und Anwendungspotentiale einer präzisen Sacherschließung aus Sicht der Wissenschaft, ABI Technik, ISSN 2191-4664, De Gruyter, Berlin, Vol. 38, Iss. 4, pp. 332–335,,
https://doi.org/10.1515/abitech-2018-4006

This Version is available at:
http://hdl.handle.net/11108/383

Kontakt/Contact
ZBW – Leibniz-Informationszentrum Wirtschaft/Leibniz Information Centre for Economics
Düsternbrooker Weg 120
24105 Kiel (Germany)
E-Mail: info@zbw.eu
https://www.zbw.eu/de/ueber-uns/profil-der-zbw/veroeffentlichungen-zbw

Standard-Nutzungsbedingungen:
Dieses Dokument darf zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden. Sie dürfen dieses Dokument nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen. Sofern für das Dokument eine Open-Content-Lizenz verwendet wurde, so gelten abweichend von diesen Nutzungsbedingungen die in der Lizenz gewährten Nutzungsrechte.

Terms of use:
This document may be saved and copied for your personal and scholarly purposes. You are not to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public. If the document is made available under a Creative Commons Licence you may exercise further usage rights as specified in the licence.

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Mitglied der Leibniz-Gemeinschaft

Fachbeitrag

Anna Kasprzik

Voraussetzungen und Anwendungspotentiale einer präzisen Sacherschließung aus Sicht der Wissenschaft

http://doi.org/10.1515/abitech-2018-4006

Zusammenfassung: Große Aufmerksamkeit richtet sich im Moment auf das Potential von automatisierten Methoden in der Sacherschließung und deren Interaktionsmöglichkeiten mit intellektuellen Methoden. In diesem Kontext befasst sich der vorliegende Beitrag mit den folgenden Fragen: Was sind die Anforderungen an bibliothekarische Metadaten aus Sicht der Wissenschaft? Was wird gebraucht, um den Informationsbedarf der Fachcommunities zu bedienen? Und was bedeutet das entsprechend für die Automatisierung der Metadatenerstellung und -pflege?

Schlüsselwörter: Sacherschließung, Automatisierung, Semantic Web

Requirements and applications of high-quality subject indexing for scientific communities

Abstract: The potential use of automated methods in subject indexing and their interaction with intellectual methods receives a lot of attention at the moment. In this context, this report addresses the following questions: What are the requirements for library metadata from the viewpoint of science? What is required in order to meet the scientific communities' need for information? And what does that entail with respect to the automatization of metadata creation and maintenance?

Keywords: Subject indexing, Automatization, Semantic Web

1 Anforderungen an bibliothekarische Metadaten

Die Idee, aus Daten Metadaten zu machen, ist mindestens 500 Jahre alt: Schon um 1545 zerschnitt Conrad Gessner seine Notizen, Briefe, Bücher und weitere Quellen, um die Schnipsel zu seiner „Bibliotheca universalis" neu zusammenzusetzen, die den Anspruch erhob, jedes damals bekannte Buch zu verzeichnen – bei ihm waren das an die 10 000.[1] Mittlerweile ist seit der Einführung der elektronischen Datenverarbeitung mit dem weniger destruktiven „Copy & Paste" und diverser Software eine vergleichsweise weitreichende Grundlage für die Erstellung, Pflege und Vernetzung von Metadaten gegeben. Es gibt eine Reihe von Werkzeugen zur Erleichterung der Erschließung, so etwa den Digitalen Assistenten, der bei der Übernahme von Schlagwörtern aus Fremddaten und der Schlagwortrecherche unterstützt.[2] Solche pragmatischen Lösungen beantworten allerdings nicht die Frage, wie sich die Sacherschließung auf der konzeptionellen Ebene langfristig entwickeln sollte, gerade im Hinblick auf Machine-Learning- und semantische Technologien, die in den letzten Jahrzehnten aufgekommen sind.

Wenn man nach den Anforderungen an bibliothekarische Metadaten aus Sicht der Wissenschaft fragt, so bekommt man von Wissenschaftlern selbst selten eine Antwort, die über: „Dass ich finde, was ich suche", hinausgeht, und Ansätze zu explorativeren Suchszenarien werden noch kaum von der breiten Mehrheit einer Fachgemeinschaft angenommen. Nun sind an Bibliotheken im Fachreferat bereits Personen tätig, die ihrerseits Wissenschaftler oder Wissenschaftlerinnen sind oder waren

1 Wright, Alex. *Cataloging the world: Paul Otlet and the birth of the information age*. Oxford: Oxford University Press, 2014, 20.

2 Hinrichs, Imma et al. „Computerunterstützte Sacherschließung mit dem Digitalen Assistenten (DA-2)." *o-bib* 3,4 (2016): 156–185, https://www.o-bib.de/article/view/2016H4S156-185 (02.10.2018).

Bereitgestellt von | ZBW - Deutsche Zentralbibliothek fuer Wirtschaftswissenschaften Leibniz Institut
Angemeldet
Heruntergeladen am | 28.11.18 11:15

und die im Idealfall nicht nur Fachkenntnisse, sondern auch die nötigen Softskills besitzen, um sich mit Informationswissenschaftlerinnen oder -wissenschaftlern auszutauschen und neue Forschungsansätze auf ihre Alltagstauglichkeit für die Erschließung zu prüfen. Im besten Fall sitzen diese mit an der Bibliothek. Diese beiden Gruppen von Akteuren sind verantwortlich für die Formulierung der Anforderungen an bibliothekarische Metadaten, um die für einen optimalen Anwendungsnutzen notwendige Datenqualität auszuloten, diesen Nutzen der Wissenschaft zu vermitteln und deren Feedback in den Entwicklungsprozess wieder einzuspeisen. Das scheint – beim aktuellen Stand der Systeme und der Wissenschaftskultur – der direkteste gangbare Weg. Er erfordert allerdings einiges an Kommunikations- und Vermittlungsarbeit.

2 Wie kann hochqualitative Sacherschließung vorangetrieben werden?

Eine Grundvoraussetzung zur Verbesserung der Sacherschließung ist die Möglichkeit, die Struktur des verwendeten Normvokabulars intuitiv erfassen und verstehen zu können, das heißt, eine hochqualitative Visualisierung davon. Hier sollten angemessene Werkzeuge entwickelt werden, um der polyhierarchischen Graphstruktur der GND gerecht zu werden, so dass strukturelle Fehler von vornherein verhindert oder leicht korrigiert werden können.[3] Weiteres Potential für die Sacherschließung ergibt sich, wenn man die Struktur der GND auf Möglichkeiten zur Bereinigung und Aufwertung untersucht, das heißt, eine Analyse des Netzes von Konzepten und Relationen und deren semantischer Ausdruckskraft, und die Diagnose von zyklischen, unverbundenen oder anderweitig suboptimalen Unterstrukturen durchführt. Eine solche Diagnose könnte über Graphalgorithmen erfolgen oder über logische Regeln und Beschränkungen, die in formalisierte Begleitsysteme zur Beschreibung des GND-Datenmodells (z. B. die GND-Ontologie) eingearbeitet sind und automatisiert ausgewertet werden können.[4]

Und schließlich muss eine moderne Sacherschließung die Tatsache mitberücksichtigen, dass sich die Form, in der Forschungsergebnisse kommuniziert und zur Verfügung gestellt werden (Artikel, Blogbeiträge, Tweets, Videos, Rohdatensätze, Software, Linked-Data-Formate usw.), immer stärker verändert. Es wird ein differenzierterer Begriff des Forschungsbeitrags benötigt – auch die Ausformulierung einer einzelnen Hypothese, Beschreibung einer Methode oder Entwicklung eines Fachvokabulars können (zitierbare) Forschungsbeiträge sein. Basierend auf Linked Open Data (LOD) und den Prinzipien des Semantic Web könnten in Zukunft verschiedenste Artefakte aus dem wissenschaftlichen Schaffenszyklus in einem umfassenden Wissensgraphen (ähnlich dem Google Knowledge Graph, aber öffentlich zugänglich) miteinander vernetzt und besser auffindbar gemacht werden.[5] Die Wissenschaftskommunikation wird sich in den nächsten Jahrzehnten weg von dokumentenzentrierten und hin zu wissensbasierten Informationsflüssen bewegen. Für wissenschaftliche Bibliotheken bietet es sich an, diese Transformation aktiv zu begleiten – und sie erfordert entsprechend neue Formen der Sacherschließung im Sinn einer semantisch annotierten Verlinkung verschiedenster Entitäten. „Neu" bedeutet hier keinen kompletten Bruch, sondern eine kontinuierliche Anwendung der Grundprinzipien der Wissensorganisation, die ja auch der traditionellen Sacherschließung zugrunde liegen, jedoch jeweils im Kontext der neuesten Forschungsergebnisse aus der Informationswissenschaft und unter Verwendung der modernsten verfügbaren Technologien.

Eine Voraussetzung, die sowohl die strukturelle Aufwertung der GND als auch den Übergang der bibliothekarischen Erschließung in eine vernetzte Welt der Wissensgraphen signifikant erleichtern würde, ist die Übertragung aller bibliothekarischer Metadaten in Semantic-Web-Beschreibungssprachen (RDFS, OWL, SKOS).[6] Ein großer Vorteil dieser Sprachen ist es, Instanzdaten (z. B. Titel-

3 Kasprzik, Anna, Jürgen Kett. „Vorschläge für eine Weiterentwicklung der Sacherschließung und Schritte zur fortgesetzten strukturellen Aufwertung der GND." *o-bib* 5,3 (2018): Im Erscheinen. Kasprzik, Anna, Jürgen Kett. „Community-getriebene Entwicklung von Sacherschließungswerkzeugen rund um die GND; und: Von der Normdatei zum Thesaurus, und vom Thesaurus zur Ontologie – strukturelle Stolpersteine." *107. Deutscher Bibliothekartag in Berlin* (2018), https://opus4.kobv.de/opus4-bib-info/frontdoor/index/index/docId/15750 (02.10.2018).

4 Ebd., und: Haffner, Alexander. „Internationalisierung der GND durch das Semantic Web." *KIM*, Frankfurt a. M., 16.07.2012, http://www.kim-forum.org/Subsites/kim/SharedDocs/Downloads/DE/Berichte/internationalisierungDerGndDurchDasSemanticWeb.pdf?__blob=publicationFile (02.10.2018).

5 Auer, Sören et al. „Towards a Knowledge Graph for Science." *Proceedings of the 8th International Conference on Web Intelligence, Mining and Semantics (WIMS)*, 2018, https://doi.org/10.15488/3401 (02.10.2018).

6 Kasprzik, Anna. „Cleaning up a legacy thesaurus to make it fit for transformation into a Semantic Web KOS." *Proceedings of the NKOS Workshop at TPDL 2018*, Porto (2018): http://ceur-ws.org/Vol-2200/paper7.pdf (02.10.2018).

Bereitgestellt von | ZBW - Deutsche Zentralbibliothek fuer Wirtschaftswissenschaften Leibniz Institut
Angemeldet
Heruntergeladen am | 28.11.18 11:15

daten) und die Wissensorganisationssysteme, aus denen die zugehörigen semantischen Annotationen stammen (Normdateien, Klassifikationen, Fachvokabulare) auf der Basis ein und desselben Modells (Resource Description Framework) darstellen und austauschen zu können, und nicht mehr aus getrennten Silos holen zu müssen. Darüber hinaus kann z. B. die GND durch eine Bereitstellung in SKOS auch außerhalb der Bibliothekswelt zu einem interessanten Wissensorganisationssystem werden, das LOD-affine Fachgemeinschaften gerne für eigene Zwecke anpassen und nachnutzen, so dass anhand ihres Feedbacks aktuelle Forschungstrends schneller in die GND einfließen können.[7] Und nicht zuletzt ergeben sich dadurch, dass im Semantic Web semantische Informationen maschineninterpretierbar gemacht werden, ganz neue Möglichkeiten für die Automatisierung der Sacherschließung. Die LOD-Cloud könnte hier einen Paradigmenwechsel einläuten.

3 Die Zukunft der Sacherschließung

Wenn man den historischen Bogen schlägt – von Gessner über Bibliothekare wie Leibniz mit seinem Zettelkabinett und Dewey mit dem „Library Bureau", das später eine Kooperation mit der Tabulating Machine Company eingehen sollte, die letztendlich zu IBM wurde und die relationale Datenbank erfand – so steht fest, dass die digitale Verarbeitung von Wissen keine vom Bibliothekswesen isolierte Entwicklung ist.[8] Mit dem Aufkommen des World Wide Web, Web 2.0 und der Technologien, die das Semantic Web ermöglichen sollen, ergibt sich für Bibliotheken eine große Chance, diese Entwicklung auch in Zukunft zu begleiten – nicht nur, indem Grundlagenforschung auf ihre Anwendbarkeit für die Erschließung geprüft wird, sondern im Idealfall durch eine aktive Steuerung, jedenfalls eine möglichst präzise Formulierung von Anforderungen an moderne Methoden aus Bibliothekssicht.

Eine der Thesen, die nun wohl Konsens ist, besagt, dass intellektuelle und automatisierte Sacherschließung ineinandergreifen müssen. Dass letztendlich die breite Masse automatisiert erschlossen werden muss, lässt sich aufgrund der Publikationsflut kaum vermeiden. Automatisierte Methoden sind jedoch noch fehleranfällig und unscharf, und daher gilt es, strategisch günstige Punkte zu finden, an denen man intellektuelle Trennschärfe ins System geben kann, so dass diese tief ausgearbeiteten Kerne optimal hochskalieren und zu mehr Qualität im System führen, ohne dass der Aufwand verpufft. Zunächst einmal brauchen die aktuell verwendeten Machine-Learning-Methoden weiterhin hochqualitative – und das heißt zurzeit noch: intellektuell erstellte – Trainings(titel)daten, und gegebenenfalls eignen sich hier einige Publikationsarten mehr als andere. Aber auch Verfahren, die sich auf Wissensorganisationssysteme stützen, um semantische Annotationen abzuleiten, sind auf hochwertige Fachvokabulare und ein ausgewogenes Verhältnis von fachlicher Breite und Tiefe angewiesen (so gibt es für die universelle Abdeckung die GND und für Fachdetails z. B. den Standardthesaurus Wirtschaft der ZBW). Hier macht es beispielsweise Sinn zu prüfen, ob sich in einigen Themenbereichen besonders große positive Effekte ergeben, wenn man ein besonders dichtes Netz von Relationen spannt, und ob man in anderen so nur kleine Qualitätssteigerungen erzielt.

Diese beiden Methodenstränge zur Automatisierung schließen sich nicht aus, sondern sollten in einer nachhaltigen Mischung kombiniert werden, insbesondere, da statistisch-heuristische Machine-Learning-Methoden zwar aus Sicht der Informatik relativ gute Ergebnisse liefern können, jedoch undurchsichtige Blackboxen sind, während semantisch-logische Verfahren eine gewisse Kontrolle über die zugrundeliegenden Strukturen erlauben. Zusätzlich kann man die Qualität der Ergebnisse einer Methode streng genommen immer nur relativ zu ihrem Anwendungskontext beurteilen, es gibt keinen losgelösten Qualitätsbegriff. Das kann bedeuten, dass für verschiedene Anwendungen jeweils verschiedene Methoden die besten Ergebnisse liefern, und auch, dass von Wissensorganisationssystemen eventuell anwendungsspezifische Versionen entstehen müssen – so wie jetzt z. B. im Rahmen des GND-Entwicklungsprogramms Community-spezifische Erweiterungen der GND („GND PLUS").[9] Die Technologien rund um das Semantic Web mit ihren Möglichkeiten zur Nachnutzung von Vokabularen im eigenen Namensraum machen diese Aufgabe um Einiges einfacher.

7 Neubert, Joachim. „Turning the GND subject headings into a SKOS thesaurus: an experiment. " Blogbeitrag, 30.03.2016, http://zbw.eu/labs/de/blog/turning-the-gnd-subject-headings-into-a-skos-thesaurus-an-experiment (24.09.2018).

8 Wright 2014: 30–42.

9 DNB, „GND-Entwicklungsprogramm 2017–2021", https://wiki.dnb.de/display/GND/GND-Entwicklungsprogramm+2017-2021 (24.08.2018), 12.

Bereitgestellt von | ZBW - Deutsche Zentralbibliothek fuer Wirtschaftswissenschaften Leibniz Institut
Angemeldet
Heruntergeladen am | 28.11.18 11:15

Anmerkung: Dieser Beitrag fasst die von der Autorin eingenommene Position in einem Impulsvortrag und der Podiumsdiskussion beim Workshop der FAG „Erschließung und Informationsvermittlung“ des GBV zusammen. Der Workshop fand im Rahmen der 22. Verbundkonferenz des GBV statt.

Autoreninformationen

Anna Kasprzik
ZBW Hamburg
Neuer Jungfernstieg 21
20354 Hamburg
anna.kasprzik@googlemail.com
orcid.org/0000-0002-1019-3606

Bereitgestellt von | ZBW - Deutsche Zentralbibliothek fuer Wirtschaftswissenschaften Leibniz Institut
Angemeldet
Heruntergeladen am | 28.11.18 11:15