

Der Weg zur Gründung der AG Erschließung ÖB-DNB und die Entwicklung eines maschinellen Verfahrens zur Verschlagwortung der Kinder- und Jugendliteratur mit GND-Vokabular

Der Weg zur Gründung der AG Erschließung ÖB-DNB

Öffentliche Bibliotheken und die Deutsche Nationalbibliothek haben viele Gemeinsamkeiten. Öffentliche Bibliotheken und die Deutsche Nationalbibliothek haben aber auch viele Unterschiede. Zu den Gemeinsamkeiten zählt zweifelsohne die inhaltliche Vielfalt des Medienangebots. Anders als institutionell gebundene Bibliotheken wie Hochschulbibliotheken, seien es Universitäts- oder Fachhochschulbibliotheken, offerieren Öffentliche Bibliotheken wie auch die DNB ein über institutionelle Belange hinausreichendes universelles Angebot. Sie haben Kinderbücher und philosophische Abhandlungen, Ratgeber und Romane, Spiele und Noten im Bestand. Die Vielfalt der inhaltlichen und formalen Medien korrespondiert mit der Vielfalt der Nutzerinnen und Nutzer. Die Nutzerinnen und Nutzer der Öffentlichen Bibliotheken und der DNB müssen nicht Angehörige einer Institution sein; es reicht, dass sie ein wie auch immer geartetes je persönliches Informationsbedürfnis haben. Zu den Unterschieden zählen neben den gesetzlichen Aufgaben, für die DNB festgelegt als Bundesgesetz, für Öffentliche Bibliotheken in einigen Bundesländern durch entsprechende Landesgesetze, der ganz unterschiedliche Umgang mit Medien. Haben Öffentliche Bibliotheken den Anspruch, Gebrauchsbibliotheken zu sein, in denen Medien, intensiv genutzt, ein zeitlich begrenztes Aufenthaltsrecht haben, so fungiert die DNB als Gedächtnisinstitution, um die Medien auch für nachfolgende Generationen zu erhalten. Die DNB hat dabei die Aufgabe, die Medien „zu erschließen und bibliografisch zu verzeichnen“ und hierbei „zentrale bibliothekarische und nationalbiografische Dienste zu leisten“ sowie die Medien „für die Allgemeinheit nutzbar zu machen“ (DNBG §2, Satz 1)¹. Die Gebrauchsorientierung der Öffentlichen Bibliotheken impliziert, dass mit der hohen Kundenorientierung die gute Auffindbarkeit von Medien im Zentrum der Erschließung steht. Was liegt daher näher, als hierfür die zentralen Metadatendienste der DNB zu nutzen? Die Versorgung mit zentral erfassten Metadaten in hoher Qualität für die Erschließung lokaler Medienbestände ist wirtschaftlich und ermöglicht, die knappen personellen Ressourcen auf dringend erforderliche Vermittlungstätigkeiten zu konzentrieren. Soweit die Theorie, soweit auch die Praxis, bis vor etwa zehn Jahren Veränderungen eintraten.

Als Folge der Erweiterung ihres Aufgabenspektrums auf digitale Medien entschied sich die DNB, die intellektuelle Erschließung vermehrt durch maschinelle Verfahren zu ergänzen oder zu ersetzen. Dies betraf auch Medienwerke, die für öffentliche Bibliotheken relevant sind. Es begann 2013 mit der Ankündigung im DNB-Rundschreiben 20130206, einzelne Segmente der Reihe A der Deutschen Nationalbibliografie, wie Ratgeber, Erlebnisberichte, Bildbände, Wörterbücher, Anleitungen zu den Themen Basteln, Handarbeiten und Hobby, Kochbücher, Reiseführer, Schul- und Berufsschulbücher

1 Gesetz über die Deutsche Nationalbibliothek, DNBG. <<http://www.gesetze-im-internet.de/dnbg/index.html>>, Stand: 07.09.2020.

und Zeitschriften von der intellektuellen Verschlagwortung auszunehmen. Der Dresdener Vorstoß zum 5. Kongress Bibliothek & Information im Jahr 2013, die Verschlagwortung kooperativ innerhalb der Lektoratskooperation zu leisten, fand keine Unterstützer.

Die nächste Stufe der Einschränkungen, die Ankündigung der Einstellung der Verschlagwortung der DDC-Gruppen B - Belletristik und K - Kinder- und Jugendliteratur ab September 2017 führte zu breiten Diskussionen in der bibliothekarischen Fachöffentlichkeit und einem kollektiven Ruf nach Korrektur und erreichte sogar die Feuilletons der deutschen Presse (K. Ceynowa, FAZ vom 31.7.17.²). Unter Federführung des Deutschen Bibliotheksverbandes wurde der Dialog der DNB mit den Öffentlichen Bibliotheken angestoßen. In einem grundlegenden Gespräch am Rande des dbv-Sektion1-Treffens im Herbst 2017 äußerten die Öffentlichen Bibliotheken den Wunsch nach einem gemeinsamen Workshop von Vertretern der DNB und der Öffentlichen Bibliotheken.

Dieser fand am 14. März 2018 mit Vertretungen der Deutschen Nationalbibliothek, der Stadtbibliothek Darmstadt, den Stadtbüchereien Düsseldorf, der Zentral- und Landesbibliothek Berlin, den Städtischen Bibliotheken Dresden, den Bücherhallen Hamburg und der Münchner Stadtbibliothek statt. Seitens der DNB wurden die Beweggründe, Rahmenbedingungen und grundlegenden Bestandteile des Konzepts zur Inhaltserschließung und die Unausweichlichkeit ihres Vorgehens wegen grundlegender Veränderungen im Publikationsverhalten und fehlender Personalressourcen transparent gemacht. Die Verfahren der maschinellen Inhaltserschließung mit dem Schwerpunkt maschinelle Verschlagwortung wurden erläutert und das internationale System zur inhaltlichen Erschließung für die Anwendung durch Verlage und Buchhandel – die THEMA-Klassifikation³ – vorgestellt. Die Vertretungen der öffentlichen Bibliotheken machten noch einmal ausführlich die Notwendigkeit der verbalen Erschließung ihrer Bestände deutlich. Konsens bestand über die GND als Basis der sachlichen verbalen Erschließung und die Notwendigkeit ihrer Aktualität. Die DNB sicherte zu, aktuelle Begriffe in die GND einzupflegen, auch wenn Medien zu diesen Themen nicht intellektuell von der DNB erschlossen werden. Unter der Mailadresse inhaltsererschliessung@dnb.de können Öffentliche Bibliotheken Vorschläge für neue GND-Begriffe an die DNB richten. Es erfolgte der Hinweis auf die hohe Dynamik der Metadaten der DNB und die Möglichkeit Metadatenänderungen sekunden genau über die OAI-Schnittstelle an die lokalen Bibliotheksmanagementsysteme auszuliefern, eine Möglichkeit, zu deren Umsetzung die Bibliotheksmanagementsystem-Anbieter der ÖBs angehalten werden müssen.

Man einigte sich auf das Vorhaben, die THEMA-angereicherten Daten des Buchhandels bibliotheksspezifisch aufzuarbeiten und maschinell in verbale Erschließungskategorien zu transferieren. Es wurde darüber informiert, dass die Marketing- und Verlagsservice für den Buchhandel GmbH – MVB sich intensiv bemüht, die Verwendung von THEMA durch die Verlage im Rahmen einer Offensive zur Verbesserung der Qualität der Daten im Verzeichnis Lieferbarer Bücher zu forcieren. Die

2 Ceynowa, Klaus: Deutsche Nationalbibliothek. In Frankfurt lesen jetzt zuerst Maschinen, in: Frankfurter Allgemeine, 31.07.2017. Online: <<https://www.faz.net/aktuell/feuilleton/buecher/maschinen-lesen-buecher-deutsche-nationalbibliothek-setzt-auf-technik-15128954.html>>, Stand: 07.09.2020.

3 Thema-Klassifikation 1.4. <<https://ns.editeur.org/thema/de>>, Stand: 07.09.2020.

Workshop-Teilnehmenden waren sich einig in der Auffassung, an der Normierung des Schlagwortvokabulars, wie sie über die GND stattfindet, festzuhalten. Denkbar sei, freie Schlagwörter aus den MVB-Daten mit der GND abzugleichen und/oder eine Liste von Begriffen zu erstellen, die für ein Mapping der THEMA-Notationen auf die GND geeignet wären. Als geeignetstes Segment empfahl sich die Kinderliteratur, zum einen wegen ihrer großen Menge und hohen Bedeutung in den ÖBs und der mangelnden Erschließung durch jedwede Klassifikation, zum anderen wegen der Menge und Qualität der freien Schlagwörter in den Daten des Buchhandels. Verabredet wurde, dass die DNB eine Skizze für ein Projekt zur Nutzbarmachung von THEMA und der freien MVB-Schlagwörter erarbeiten und zur Verfügung stellen sollte, während die ÖB-Vertretungen eine Liste von Schlagwörtern im Bereich der Kinderliteratur, die von besonderer Relevanz z.B. bei der Auskunftserteilung sind, an die DNB übermitteln wollten.

In der Folgezeit konnten ein regelmäßiger Austausch und eine kontinuierliche Zusammenarbeit der Teilnehmenden des Workshops mittels Telefonkonferenzen etabliert werden. Man begann mit der Erstellung einer Liste von ca. 8.000 für Kinderliteratur relevanten Schlagwörtern, auf die die freien Schlagwörter der MVB-Datensätze gemappt werden sollten. Die DNB extrahierte Beispielcluster von MVB-Datensätzen für Kinder- und Jugendliteratur zur Prüfung durch die ÖBs.

Am 30. Januar 2019 fand ein weiteres Treffen in Frankfurt statt, bei dem sich eine dauerhaft agierende Arbeitsgemeinschaft mit dem Titel „AG Erschließung ÖB-DNB“ konstituierte. Aktive Mitglieder mit Stand Juli 2020 sind Anne Dreger (Zentral- und Landesbibliothek Berlin), Klaus Peter Hommes (Stadtbüchereien Düsseldorf), Elisabeth Mödden (Deutsche Nationalbibliothek), Letitia Mölck (Deutsche Nationalbibliothek), Loredana Pinna (Bücherhallen Hamburg), Daniela Sitte-Zöllner (Städtische Bibliotheken Dresden). Man beschloss, in monatlichen Telefonkonferenzen zu arbeiten und sich zusätzlich einmal jährlich persönlich zu treffen.

Für ein Verfahren für maschinell generierte Schlagwörter wurden folgende Grundsätze vereinbart:

1. Die maschinell generierten Schlagwörter reichern die Titeldaten an, sie bieten keinen vollständigen Ersatz für die intellektuelle Erschließung.
2. Die maschinell generierten Schlagwörter geben den wesentlichen Inhalt der Publikation wieder, so dass sie bei einer Recherche gefunden werden kann.
3. Die maschinell generierten Schlagwörter werden als solche für die Dateneinspielung gekennzeichnet.
4. Es gilt der Qualitätsmaßstab 80:20, d.h. die maschinell generierten Schlagwörter treffen zu mindestens 80 Prozent zu.

Diese Grundsätze ermöglichen es, einerseits mit hinreichender Präzision themenbezogene Medien bei Recherchen zu finden, andererseits pragmatisch mit den vorhandenen Daten umzugehen.

Beschreibung des Verfahrens und der Entwicklung des Wörterbuchs für die Kinder- und Jugendliteratur

Grundlage für das Verfahren sollten Erschließungsdaten sein, die die DNB durch den MVB Marketing- und Verlagsservice des Buchhandels GmbH regelmäßig geliefert bekommt (siehe Abbildung 1).

Der MVB bietet Verlagsprodukte und Dienstleistungen für die Vermarktung von Büchern an. Über das Verzeichnis Lieferbarer Bücher (VLB), seine zentrale Plattform, können Verlage ihre Publikationen mit Hilfe gut gepflegter Metadaten besser such- und sichtbar machen. Das aktuelle Preismodell des MVB berechnet die Gebühren der Titelmeldung in Abhängigkeit von der Qualität der Metadaten, die der Verlag an das VLB liefert: Je besser die an das VLB gelieferten Daten sind, desto niedriger der Preis.⁴ Laut Angaben des MVB sind dort mehr als 22.000 Verlage und Verleger registriert und rund 2,5 Millionen Titel verzeichnet. Damit ein Verlag den günstigen Goldstatus erlangen kann, muss er neben bestimmten anderen Angaben folgende inhaltsbeschreibende Kategorien ausfüllen: Verschlagwortung, Hauptbeschreibung, THEMA-Klassifikation.⁵

Die Basis für die maschinelle GND-Schlagwortvergabe in der Kinder- und Jugendliteratur sind die MVB-Schlagwörter, die mit dem Vokabular der GND abgeglichen werden. Dafür wird die Software der Firma Averbis aus Freiburg, die Averbis Extraction Platform, eingesetzt, mit der auch die maschinelle Verschlagwortung und Klassifizierung der Online-Publikationen der Reihe O erfolgt.

Titel: Eulenzauber. Im Kreis der Goldflügel
Urheber: Ina Brandt; mit Illustrationen von Irene Mohr
...
Freie Schlagwörter:
Abenteuer, Eule, Familie, Fantasie, Flora, Freunde, Freundschaft, Für Leser der Zauberschneiderei, Für Leser von Linda Chapman, Für Leser von Sternenschweif, Geheimnis, Goldene Flügel, Goldwing, Helfen, Irene Mohr, Kinderbuch, Mädchen, Magie, Natur & Tiere, Reihe, Spiegel-Bestseller-Autorin, Zaubereule
Thema-Klassifikation:
YFH → Kinder/Jugendliche: Fantasy
YFM → Kinder/Jugendliche: Liebesromane, Geschichten über Liebe und Freundschaft
5AJ → empfohlenes Alter: ab 8 Jahre
YFP → Kinder/Jugendliche: Natur- und Tiergeschichten

Abbildung 1: Beispiel für mitgelieferte MVB-Schlagwörter und Thema-Klassifikation

4 Verzeichnis Lieferbarer Bücher (VLB), Informationen – Preise. <<https://vlb.de/leistungen/titelmeldung>>, Stand: 13.10.2020.

5 Verzeichnis Lieferbarer Bücher (VLB), Informationen – Titelmeldung. <<https://vlb.de/leistungen/titelmeldung>>, Stand: 04.09.2020.

Der Ablauf des maschinellen Erschließungsprozesses für die Printpublikationen der Kinder- und Jugendliteratur lässt sich folgendermaßen beschreiben: Regelmäßig übermittelt werden aktuelle MVB-Daten über neue Printpublikationen an die DNB. Im Katalogisierungssystem der DNB wird ein entsprechender Datensatz für jede Publikation angelegt. Wenn dann eine neue Printpublikation aus der Kinder- und Jugendliteratur in der DNB abgeliefert wird, wird diese zunächst im Medieneingang ausgepackt, der entsprechende Datensatz aufgerufen, die Publikation inventarisiert und formal erschlossen. Anschließend folgt der automatische Erschließungsprozess (siehe Abbildung 2). Dabei werden nur der Titel und die MVB-Schlagwörter aus dem Titeldatensatz der Publikation über eine Verarbeitungssteuerung an die Averbis-Software übermittelt. Die Software führt eine linguistische Vorverarbeitung durch und gleicht auf Segmentebene sowohl die Titel als auch die MVB-Schlagwörter mit den GND-Schlagwörtern des speziellen Wörterbuchs für Kinder- und Jugendliteratur (im Folgenden KiJuLit-Wörterbuch genannt) ab. Die Verarbeitungssteuerung schreibt die fünf ermittelten GND-Schlagwörter mit den höchsten Konfidenzwerten⁶ in den Titeldatensatz der Publikation zurück. Die Inhaltserschließung der DNB ergänzt dann noch die Angaben zu Zielgruppe(n) und Art des Inhalts und schaltet den Datensatz für die Anzeige in der Nationalbibliografie frei.

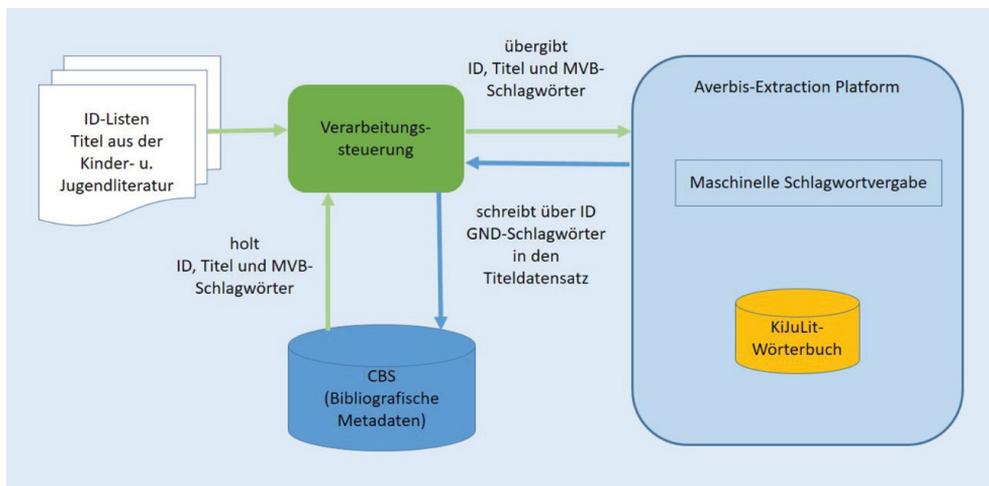


Abbildung 2: Erschließungsprozess der Kinder- und Jugendliteratur

Die maschinelle Erschließung von Online-Publikationen der Kinder- und Jugendliteratur unterscheidet sich etwas davon. Hier werden die Titeldaten direkt vom Verlag geliefert. Sobald eine Online-Publikation eingetroffen ist, werden in der folgenden Nacht die Titeldaten und die Verlagsschlagwörter über die Verarbeitungssteuerung an die Averbis-Software übergeben, dort erfolgt dann wie oben beschrieben die maschinelle Verschlagwortung. Alle ermittelten GND-Schlagwörter werden mit entsprechenden Kennzeichnungen in den Titeldatensatz geschrieben (siehe Abbildung 3).

6 Schätzwert der Averbis-Software auf einer Skala von 0 bis 1 bzgl. der inhaltlichen Genauigkeit der einzelnen vorhergesagten Schlagwörter.

```

MARC21-Format (gekürzt)
001 1155611047 --> IDN des Datensatzes
082 74$aK$qDE-101$223sdbn --> DDC-Sachgruppe K = Kinderliteratur
385 $0(DE-588)4028859-6$0https://d-nb.info/gnd/4028859-6$0(DE-101)040288595$aJugend$2gnd --> Zielgruppe

490 1 $aElena - ein Leben für Pferde / Nele Neuhaus$vBd. 7

650 7$81\p$0(DE-588)4045503-8$0https://d-nb.info/gnd/4045503-8$0(DE-101)040455033 $aPferd$2gnd
650 7$82\p$0(DE-588)4035646-2$0https://d-nb.info/gnd/4035646-2$0(DE-101)040356469 $aLiebe$2gnd
650 7$83\p$0(DE-588)4036871-3 $0https://d-nb.info/gnd/4036871-3$0(DE-101) 040368718 $aMädchen$2gnd
--> GND-Schlagwörter

653 $aJugendbücher
653 $aLiebe
653 $aPferd
653 $aNele Neuhaus
653 $aOstwind --> freie Deskriptoren aus Fremddaten
653 $aElena
653 $aCharlottes Traumferd
653 $aBibi und Tina
653 $aPferdebuch für Mädchen

883 0 $81\p$amaschinell gebildet$c0,09274$d20200430$qDE-101
883 0 $82\p$amaschinell gebildet$c0,02584$d20200430$qDE-101
883 0 $83\p$amaschinell gebildet$c0,02579$d20200430$qDE-101 --> Angabe zur Erstellung von
GND-SW, Konfidenzwert und Datum der Erstellung
    
```

Abbildung 3: Kennzeichnung der maschinell erstellten GND-Schlagwörter im Marc21-Feld 883

Das Wörterbuch der für Kinder- und Jugendliteratur relevanten GND-Begriffe bildet den grundlegenden Wortschatz für das Mapping der MVB-Schlagwörter. Für die Erstellung des Wörterbuchs wurden alle bisher in der DNB für die Erschließung der Kinder- und Jugendliteratur intellektuell vergebenen Schlagwörter extrahiert (verwendetes Tool: Metafacture) und daraus ein Auszug für die Averbis-Software erstellt. Das KiJuLit-Wörterbuch enthält neben Sachschlagwörtern auch Personennamen, Geografika und Werktitel – insgesamt 7.880 Begriffe in der ersten Version.

Es folgten drei aufeinander aufbauende Tests des Verfahrens.

Im Oktober 2019 wurden im Testsystem der DNB 1.000 Titeldatensätze des MVB, die über eine Sachgruppe K und über MVB-Schlagwörter in den Pica-Feldern 5560 verfügten, mit diesem KiJuLit-Wörterbuch maschinell verschlagwortet. Das Testkorpus enthielt Titeldatensätze von 2017 bis 2019. Es erfolgte ein Abgleich zwischen den abgelieferten MVB-Schlagwörtern mit der bevorzugten Benennung sowie den abweichenden Benennungen der GND-Schlagwörter aus dem KiJuLit-Wörterbuch (siehe Abbildung 4). Der Titel der Publikation fand bei diesem ersten Test noch keine Verwendung für die Verschlagwortung.

Aus den 1.000 Titeldatensätzen wurde eine Stichprobe von 200 Titeln ausgewählt und von den Mitgliedern der AG nach einem einfachen Ampelsystem ausgewertet (siehe Abbildung 5). Grün stand dabei für „sehr gut“, Gelb bedeutete „noch ok“ und bei Rot war die maschinelle Erschließung des Titels „ungenügend“. Parallel dazu erfolgte auch eine Bewertung durch Kolleg*innen der Inhaltserschließung der DNB.

Die Ergebnisse der Stichproben konnten die Mitglieder der AG nicht überzeugen (siehe Abbildung 6)

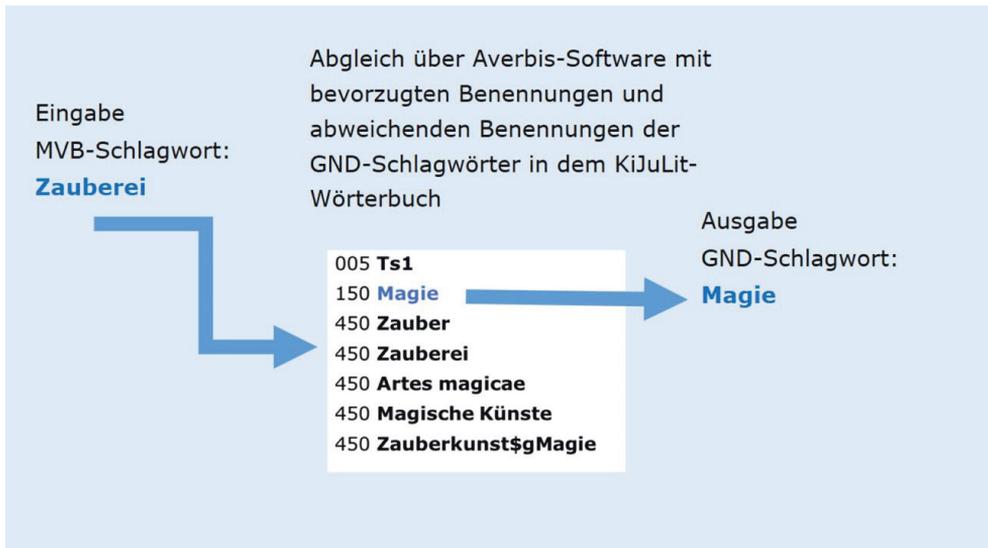


Abbildung 4: Abgleich MVB-Schlagwörter und GND-Schlagwörter

Titel	MVB-Schlagwörter (frei vergeben)	GND-Schlagwörter (automatisch ermittelt)				Kommentare
Kein Platz für uns 9783219117936	Gemeinschaft; Flucht; Ausgrenzung; Empathie; Fremdsein; Flüchtlinge; Mitgefühl; Felsen; Meer; Ozean; Seelöwe; Saebär; Ohrenrobbe; Robbe; Seehund; Palindrom; fliehen; Migration; ausschließen; willkommen heißen; Krieg	s Seehund; s Meer; s Fremdheit; s Flüchtling; s Ausgrenzung				
Zu viel Plastik 9789463414296	Chemikalie; Einwegplastik; Granulat; Mikroplastik; Plastikflaschen; Strand säubern; Strömung; Tiere; recyceln	s Strand; s Plastik; s Tiere				
Wundertier Schwamm 9783715207490	Sachbilderbuch; Meer; Natur; Schwämme; Dinosaurier	s Dinosaurier; s Pilze; s Meer; s Natur				Vorhandene SW geben den Inhalt nicht wieder. Wesentliches SW "Schwämme" fehlt. Warum hat der Deskriptor nicht gematcht?
Spiele mit Python supereasy programmieren 9783831036752	Python Buch; programmieren; Programmieren lernen; Python lernen; Python programmieren; Programmieren für Kinder; Jugendbuch; Spiele programmieren; Programmieren Buch; Computerspiele programmieren; Computer; Informatik; EDV; Informatik Unterricht; Programme schreiben; Geschenk Jungen; Bücher Jungen; Mädchen; Anwendung; Programmiersprache; Computerspiel; Kinder; Jugendliche; IDLE; IDE; Turtle; Shell; Code; Pygame; Tkinter; Spiel entwickeln; Python; Digitalpakt; Schule; Freizeit; PC Spiel	s Computerspiel; s Programmierung; s Spiel; s Junge				Python <Programmiersprache>; Programmierung
Schneckenleben 9783737356398	Gehäuse; Baumschnecke; Schnecken; Garten; Gehäuseschnecken; Banderschnecke; Weinbergsschnecke; Kriechtiere; Wiese; Kinder; Natur; Schneckenhaus; Lernen; Balkon; Weichtiere; Insekten; Wald	s Schneckenhaus; s Reptilien; s Schnecken; s Wiese; s Insekten				Schnecken

Abbildung 5: Auszug aus Bewertungstabelle Test 2

Für den zweiten Test fand nun auch die Vergabehäufigkeit der Schlagwörter Berücksichtigung. Alle Schlagwörter, die bisher nur ein einziges Mal bei der intellektuellen Erschließung der Kinder- und Jugendliteratur vergeben worden waren, wurden eliminiert, mit dem Ziel, unter anderem die Zahl der falsch vergebenen mehrdeutigen Terme zu verringern. Auch Angaben zum Inhaltstyp wie Atlas, Kinderbuch, Enzyklopädie, Biografie etc. oder zur Zielgruppe wie Kind, Vorschulkind etc. wurden entfernt. Alle ausgeschlossenen GND-Schlagwörter wurden in einer Negativliste zusammengestellt. Ergänzend wurden Listen der in den Bibliotheken der AG-Mitglieder verwendeten Schlagwörter mit dem KiJuLit-Wörterbuch abgeglichen und die fehlenden GND-Schlagwörter aufgenommen. Im Ergebnis enthielt das KiJuLit-Wörterbuch 2.158 GND-Schlagwörter und die Negativliste 3.822 GND-Schlagwörter.

Im Januar 2020 erfolgte der zweite Test mit der Averbis-Software, dem neuen KiJuLit-Wörterbuch und neu selektierten 1.107 Titeldatensätzen, die sich aus allen im Jahr 2019 von den Bücherhallen Hamburg neu erworbenen Publikationen der Kinder- und Jugendliteratur speisten. Eine Stichprobe von 200 Titeldatensätzen bewerteten die Beteiligten der ÖBs erneut nach dem Ampelsystem. Die Modifikationen führten zu einer deutlichen Verbesserung der Bewertung, waren allerdings nach einstimmiger Meinung der Beteiligten immer noch nicht für eine Freigabe des Verfahrens geeignet (Siehe Abbildung 6).

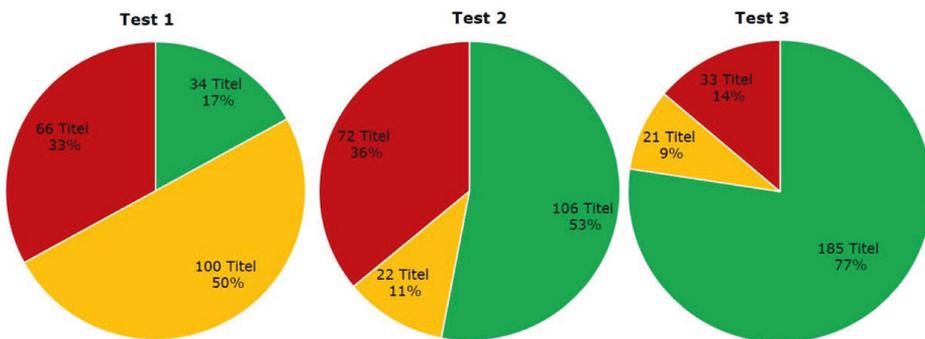


Abbildung 6: Ergebnisse der Bewertungen von Test 1 / 200 Beispiele; Test 2 / 200 Beispiele; Test 3 / 239 Beispiele

Die AG Mitglieder beschlossen einen dritten Test mit weiter optimiertem Verfahren und KiJuLit-Wörterbuch. Mittels Selektion der MVB-Daten anhand der Warengruppennummer [VLB-WN]1212 konnten Pappbilderbücher ausgeschlossen werden. Diese hatten einen zu großen Anteil an der schlechten Bewertung der Ergebnisse, so dass die AG sich darauf einigte, auf die Anreicherung dieser Datensätze zu verzichten. Anhand konkreter Einzelbeispiele wanderten einige Schlagwörter von der Negativliste wieder zur Positivliste und einige von den Verlagen mit komplett anderer Intention vergebene Begriffe, wie Gutenachtgeschichte, Kindergeburtstag, Geburtstagsgeschenk in die Negativliste.

Die Kolleg*innen der DNB modifizierten die linguistischen Prozesse des Mappingverfahren, damit auch Wortbestandteile erkannt werden konnten. Zusätzlich wurden von der Averbis-Software nicht nur die abgelieferten MVB-Schlagwörter für den Abgleich herangezogen, sondern auch die jeweiligen Titel der Publikationen. Der dritte Test lief im April 2020.

Nach der Bewertung der Ergebnisse haben sich die AG Mitglieder dafür ausgesprochen, ab September 2020 mit der maschinellen Erschließung der Kinder- und Jugendliteratur zu beginnen. Die Ergebnisse des produktiven Verfahrens sollen laufend über Stichproben und Fehleranalysen ausgewertet und durch weitere Anpassungen kontinuierlich verbessert werden.

Ausblick

Die maschinelle Erschließung der Kinder- und Jugendliteratur mit Hilfe der freien Schlagwörter der Verlage ist ein erster Schritt in der Verbesserung der verbalen Sacherschließung der für Öffentliche Bibliotheken relevanten Literatur. Im weiteren Prozess wird dieses Verfahren auf andere Bereiche ausgedehnt werden können, wie z.B. Belletristik, Ratgeber und Reiseliteratur. Die Erfahrungen mit der Kinder- und Jugendliteratur lassen jedoch erwarten, dass das Mappingverfahren nicht ohne Modifikationen auf andere Segmente übertragen werden kann, sondern dass es einer genauen Analyse bedarf, welche Stoppwortlisten segmentspezifisch erstellt und eingesetzt werden müssen, um zufriedenstellende Resultate zu erhalten.

Die Analyse und Umsetzung der verlagsseitig vergebenen Schlagwörter schöpft das in den Daten vorhandene Potenzial jedoch bei weitem noch nicht aus. Bücher werden von den Verlagen mit der THEMA-Klassifikation erschlossen. Hier ist zu prüfen, wie diese klassifikatorische Erschließung für verbale Sucheinstiege nutzbar gemacht werden kann. Mit dem THEMA-Code werden z.B. Inhalte von Romanen gekennzeichnet. Mit der Notation FBA-1DFG-DE-VRB wird z.B. angegeben, dass Düsseldorf der Schauplatz eines zeitgenössischen Romans ist. Durch Hinzunahme weiterer Qualifier lassen sich sprachliche, zeitliche und pädagogische Aspekte sowie Zielgruppen angeben.

Die Umsetzung dieser Daten für die inhaltliche Erschließung ÖB-relevanter Medienwerke wird ein Prozess sein, dessen Ende derzeit noch nicht absehbar ist. Zugleich gilt es daran zu arbeiten, dass die in den ÖBs eingesetzten Bibliotheksmanagementsysteme mit einer regelmäßigen Aktualisierung von Daten umgehen können. Nachweisdaten sind zunehmend weniger unveränderlich, sondern sollten angereichert und verbessert werden können. Neben dem technischen Prozess der Nutzbarmachung von im Hintergrund vorhandenen Daten ist der Erkenntnisprozess, dass Metadaten nicht mehr sofort abgeschlossen und vollendet sind, entscheidend, um Veränderungsprozesse im Kollegium und gegebenenfalls bei den Anbietern von Bibliotheksmanagementsystemen zu initiieren und umzusetzen.

Anne Dreger, Zentral- und Landesbibliothek Berlin
Klaus Peter Hommes, Stadtbüchereien Düsseldorf
Elisabeth Mödden, Deutsche Nationalbibliothek Frankfurt
Letitia Mölck, Deutsche Nationalbibliothek Frankfurt
Loredana Pinna, Bücherhallen Hamburg
Daniela Sitte-Zöllner, Städtische Bibliotheken Dresden

Zitierfähiger Link (DOI): <https://doi.org/10.5282/o-bib/5637>

Dieses Werk steht unter der Lizenz [Creative Commons Namensnennung 4.0 International](#)