

Schmucker, Susanne; Schöntube, Cornelia

Contribution to Periodical — Published Version

Fortschritte der Automatisierung und Kollaboration in der Sacherschließung. Bericht vom 6. Workshop „Computerunterstützte Inhaltserschließung“

O-Bib. Das Offene Bibliotheksjournal

Suggested Citation: Schmucker, Susanne; Schöntube, Cornelia (2023) : Fortschritte der Automatisierung und Kollaboration in der Sacherschließung. Bericht vom 6. Workshop „Computerunterstützte Inhaltserschließung“, O-Bib. Das Offene Bibliotheksjournal, ISSN 2363-9814, VDB, München, Vol. 10, Iss. 3, pp. 1–11,
<https://doi.org/10.5282/o-bib/5928>

This Version is available at:
<http://hdl.handle.net/11108/597>

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.



<https://creativecommons.org/licenses/by/4.0/>

Fortschritte der Automatisierung und Kollaboration in der Sacherschließung

Bericht vom 6. Workshop „Computerunterstützte Inhaltserschließung“

Bereits zum sechsten Mal fand am 16. und 17. November 2022 der Workshop „Computerunterstützte Inhaltserschließung“ statt. Vorträge rund um die Weiterentwicklung des Digitalen Assistenten (DA-3) sowie zu zukunftsweisenden Projekten der Sacherschließung wurden vorgestellt und diskutiert. Organisiert von der Universitätsbibliothek Stuttgart, der Staatsbibliothek zu Berlin – Preußischer Kulturbesitz und der Eurospider Information Technology AG in Zusammenarbeit mit dem Bibliotheksservice-Zentrum Baden-Württemberg und der Deutschen Nationalbibliothek fand der Workshop zum dritten Mal als Webkonferenz statt.¹

Frank Scholze, Generaldirektor der Deutschen Nationalbibliothek, begrüßte die Teilnehmer*innen. Er betonte die Bedeutung der digitalen Transformation und dass es Zeit brauche, bis diese in alle „Kapillaren“ eingezogen sei. Wichtig für die Umsetzung seien Vernetzung und Kooperation. Tools wie der Digitale Assistent (DA-3)² und Culturegraph³ würden dabei helfen, Metadaten bereitzustellen, Normdaten zu pflegen, intellektuelle und maschinelle Erschließung und somit Mensch und Maschine zu verflechten. In diesem Kontext gewinne der Ansatz des human in the loop (HITL) aus der Künstlichen Intelligenz an Bedeutung. Maschinelle Verfahren würden bei diesem Ansatz durch die Interaktion mit menschlichen Expert*innen kontinuierlich angepasst. Die Vorträge des Workshops griffen diese Themen auf und sollten zur Diskussion anregen.

Im Anschluss begrüßte auch Helge Steenweg, Leiter und Direktor der Universitätsbibliothek Stuttgart die Teilnehmenden. Er dankte Helga Karg von der Deutschen Nationalbibliothek und Volker Conradt vom Bibliotheksservice-Zentrum Baden-Württemberg für die Organisation der Veranstaltung über die Onlineplattform und betonte, dass das virtuelle Format sicher mit zur hohen Anzahl von 220 Anmeldungen beigetragen habe.

Qualitätsprüfung der maschinellen Erschließung mit Hilfe des DA-3

Den Start machte Anna Kasprzik, ZBW – Leibniz-Informationszentrum Wirtschaft, mit dem Vortrag „Auswertung der Ergebnisse aus den Bewertungen der maschinellen Erschließung (zbwase) im DA-3“. In der ZBW wird bereits nach dem HITL-Ansatz als wichtigem Baustein eines konsistenten Qualitätssicherungskonzeptes gearbeitet. Seit 2018 gibt es an der ZBW den produktiven Dienst der automatisierten Sacherschließung (AutoSE)⁴ für die Verschlagwortung der Dokumente im ZBW-Bestand mit

1 Das Programm des Workshops mit den Präsentationen zum Download kann im Wiki der DNB aufgerufen werden: <<https://wiki.dnb.de/pages/viewpage.action?pageId=252121510>>, Stand: 02.03.2023.

2 Derzeit gibt es drei verschiedene Profile des Digitalen Assistenten: für die K10plus-Bibliotheken des GBV und des SWB, den Österreichischen Bibliotheksverbund (OBV) und die Deutsche Nationalbibliothek (DNB).

3 Culturegraph, <https://www.dnb.de/DE/Professionell/Standardisierung/AGV/_content/culturegraph_akk.html>, Stand: 03.03.2023.

4 Für weitere Informationen siehe Kasprzik, Anna: Aufbau eines produktiven Dienstes für die automatisierte Inhaltserschließung an der ZBW. Ein Status- und Erfahrungsbericht, in: O-Bib. Das Offene Bibliotheksjournal 10 (1), 2023, S. 1-13, <<https://doi.org/10.5282/o-bib/5903>>.

dem Standard-Thesaurus Wirtschaft (STW).⁵ Methodisch wird Annif⁶ als Steckrahmen genutzt, ein von der Nationalbibliothek Finnland entwickeltes Werkzeug auf Python-Basis für die automatisierte Sacherschließung, ergänzt durch ein ZBW-eigenes Modell und ZBW-eigene Software-Komponenten. Stündlich wird die Datenbasis des Rechercheportals der ZBW – EconBiz⁷ – nach neuen Titeln durchsucht. Gefundene Publikationen werden automatisiert verschlagwortet, die Schlagwörter (im DA-3 gekennzeichnet mit „zbwase“) in die Datenbasis zurückgeschrieben und somit für den DA-3 abrufbar.⁸ 1,4 Millionen Metadatensätze wurden bereits automatisiert verschlagwortet. Ist eine Ressource intellektuell erschlossen, wird das Ergebnis über die EconBiz-Datenbasis abgegriffen und abgeglichen, ob die AutoSE-Vorschläge übernommen wurden. Dieser Abgleich dient dem AutoSE-Team als ein erstes einfaches Qualitätsmaß für die Richtigkeit und Vollständigkeit der maschinellen Erschließung. In der Vergangenheit führte die ZBW stichprobenartige jährliche Reviews der jeweils verbesserten Verfahren durch. Fachreferent*innen bewerteten die automatisiert generierten Schlagwörter mit Hilfe einer eigens dafür entwickelten Oberfläche. Seit 2022 gibt es ein Bewertungstool im DA-3, das eine kontinuierliche Bewertung der AutoSE-Vorschläge im Rahmen der täglichen Erschließungsarbeit ermöglicht. Soweit vorhanden, werden AutoSE-Vorschläge nach jeder intellektuellen Sacherschließung im Bewertungstool des DA-3 bewertet – zunächst jeder einzelne vorgeschlagene Deskriptor und abschließend die Qualität des Gesamtindexates. Diese abgestufte Bewertung liefert dem AutoSE-Team noch einmal detaillierte Informationen dazu, wie zutreffend ein vorgeschlagener Deskriptor bzw. das Gesamtindexat für die inhaltliche Beschreibung der Informationsressource ist. Neben der Einspielung als Vorschläge im DA-3 wird die AutoSE an Titeln in EconBiz ergänzt, die bisher noch nicht intellektuell verschlagwortet wurden. Die Verfahren der maschinellen Erschließung erzielen auf der Deskriptorebene zunehmend bessere Ergebnisse, auf Dokumentebene noch nicht. Anna Kasprzik betonte, dass der F1-Wert⁹ zur Qualitätsanalyse nicht ausreiche. Eine bessere Herangehensweise könne sein, den Recall höher zu gewichten (z.B. 1,5fach). Das Bewertungstool im DA-3 sei sehr gut für eine Standortbestimmung der maschinellen Methoden im laufenden Produktivbetrieb der Sacherschließung geeignet. Es gebe noch keine Informationen darüber, welche Titel bzw. Sachgebiete schwieriger zu erschließen seien, antwortete Anna Kasprzik auf eine Teilnehmerfrage. Auch werde die Qualität der Trainingsdaten nicht gemessen. Man nähme die Daten bewusst so, wie sie seien. Auf die Frage, ob die Methoden auf andere Einrichtungen übertragen werden könnten, antwortete Anna Kasprzik, dass es grundsätzlich möglich sei, da die Methoden als Open-Source-Software zur Verfügung stünden. Es sei aber ein großer Aufwand, diese auf die jeweiligen Informationsressourcen und Vokabulare der verschiedenen Fachgebiete anzupassen. Das Bewertungstool im DA-3 kann auch durch andere den DA-3 nutzende Bibliotheken für die Bewertung eigener Verschlagwortung angepasst und verwendet werden.

5 Standard-Thesaurus Wirtschaft (STW), <<https://zbw.eu/stw/version/latest/about.de.html>>, Stand: 02.03.2023.

6 Annif, <<https://annif.org>>, Stand: 02.03.2023.

7 EconBiz, <<https://www.econbiz.de>>, Stand: 02.03.2023.

8 Bisher nur im K10plus-Profil des DA-3 möglich.

9 Berechnung des F1-Wertes: das harmonische Mittel von Precision (welcher Anteil der vorgeschlagenen Deskriptoren wurde auch intellektuell vergeben?) und Recall (welcher Anteil der intellektuell vergebenen Deskriptoren wurde auch von der Maschine vorgeschlagen?).

Ähnlichkeitssuche als potenzielle Weiterentwicklung des DA-3

Der zweite Vortrag von Thomas Murphy, Eurospider, beleuchtete die „Ähnlichkeitssuche im DA-3“. In der Vorgängersoftware DA-2 gab es bereits eine einfache Umsetzung einer solchen Funktion, die nun auch für die Weiterentwicklung des DA-3 gewünscht ist. Es haben sich vier Anwendungsfälle für Ähnlichkeitssuchen herauskristallisiert. Fall 1 ist die Suche nach anderen Ausgaben eines Titels, deren Erschließung wegen unterschiedlicher ISBN durch den aktuellen DA-3 nicht angezeigt werden. Als Resultat werden andere Ausgaben im selben Katalog und andere sowie gleiche Ausgaben in anderen Katalogen angezeigt. Die dortigen Schlagwörter sollen per Knopfdruck in den aktuellen Titel übernommen werden können. Fall 2 ist der Anwendungsfall, den es bereits im DA-2 gab: zu einem Titel, der nicht mit GND, jedoch mit DDC-Sachgruppen, STW, Basisklassifikation (BK) oder Regensburger Verbundklassifikation (RVK) erschlossen ist, werden via Konkordanzen GND-Vorschläge berechnet. Mit der Ähnlichkeitssuche können Sacherschließende im Umfeld all dieser Erschließungsinformation zusätzliche oder geeignetere GND-Schlagwörter finden. Das System bietet eine Liste von GND-Schlagwörtern mit Rang an. Die Rangfolge der Schlagwörter kommt zustande, indem ein „Score“, basierend auf ähnlichen Dokumenten und der Häufigkeit der Schlagwörter in der Kollektion, gebildet wird. Es ist noch nicht sicher, welche Konkordanzen verwendet werden sollen. STW zu GND wäre z.B. eine Möglichkeit. Fall 3 deckt folgendes Anwendungsszenario ab: Nach der Erstellung einer provisorischen Erschließung durch die Sacherschließenden werden mit der Ähnlichkeitssuche Titel gesucht, die eine vergleichbare Erschließung haben. Das System bietet verschiedene Ausgaben des gleichen Titels und Titel in anderen Sprachen an (z.B. slowakisch). So kann überprüft werden, ob die eigene Erschließung vollständig ist oder ggf. noch Ideen für hilfreiche Ergänzungen gewonnen werden können. Im vierten Fall ist ein Informationsbedürfnis gegeben. Gesucht werden Titel, die für eine individuelle thematische Suche relevant sind und in diesem Zusammenhang auch Hinweise auf GND-Schlagwörter geben, die für die entsprechende Suche zielführend sind. Im Anschluss an eine erste Suche werden die Ergebnisse der Trefferliste durch die suchende Person im Hinblick auf ihre Relevanz für die thematische Suche bewertet. Auf dieser Basis wird eine neue Suche gestartet, um weitere relevante Titel im Katalog zu finden. In diesem Kontext kann auch nach weiteren relevanten Schlagwörtern gesucht werden. Eine sinnvolle und nutzungsfreundliche Umsetzung dieser vier Anwendungsfälle muss noch überlegt werden, da es viele Suchvarianten und Möglichkeiten zur Verschlagwortung im DA-3 gibt. Auf die Frage, ob eine Kopplung dieser vier Fälle auch in Discovery-Systemen vorstellbar wäre, entgegnete Thomas Murphy, dass dies unterschiedliche Dinge seien. Man könne technische Erkenntnisse sicher einfließen lassen, eine unmittelbare Kopplung wäre aber wohl nicht möglich. Peter Schäuble ergänzte, dass der vierte Anwendungsfall auch für die Erschließung von Forschungsdaten interessant sein könne. Forscher*innen würden sich mit Verschlagwortung oft nicht auskennen und eine Suche in Preprints ergäbe Schlagwörter, die man übernehmen könne.

Maschinelle Erschließung in der Biomedizin

Im anschließenden Vortrag von Anastasios Nentidis und Anastasia Krithara, beide National Centre for Scientific Research „Demokritos“, Griechenland, ging es um „Maschinelle Erschließung mit MeSH“.¹⁰ Beide Computerspezialist*innen mit Schwerpunkt in künstlicher Intelligenz arbeiten seit etwa zehn Jahren an semantischer Erschließung im Bereich Biomedizin. Der Wettbewerb BioASQ¹¹ organisiert sogenannte „Challenges“ im Bereich der biomedizinischen semantischen Indexierung und Fragenbeantwortung (englisch: Question Answering, QA) und misst z.B. die Fähigkeit der teilnehmenden Systeme, Dokumente in Echtzeit zu annotieren. Seit 10 Jahren gibt es jährlich eine Challenge, bestehend aus einem Bewertungszyklus und anschließendem Workshop. Das leistungsstärkste System eines Wettbewerbs wird mit einem Preis dotiert. 100 Teams aus 28 Ländern¹² haben im Laufe der Jahre teilgenommen. Die „Challenges“ umfassen Aufgaben zur hierarchischen Textklassifikation, zum maschinellen Lernen, zum Information Retrieval, zur Fragenbeantwortung aus Texten und strukturierten Daten, zur Zusammenfassung mehrerer Dokumente und vielen anderen Bereichen. Teilnehmende Institutionen können eine oder mehrere von verschiedenen Aufgaben (Tasks) übernehmen. In der Präsentation wurden sieben der BioASQ Tasks kurz genannt. Die teilnehmenden Teams kommen weltweit von verschiedenen Universitäten und Unternehmen, die an Methoden der Künstlichen Intelligenz arbeiten. Das BioASQ Ökosystem stellt Infrastruktur, Bewertungsservices und Datensätze zur Verfügung und ermöglicht den teilnehmenden Institutionen, ihre Systeme mit den Daten von BioASQ zu bedienen. Task A „Large-Scale Online Biomedical Semantic Indexing“ wurde in der Präsentation ausführlicher vorgestellt. Ziel dieser Aufgabe ist die automatisierte Annotation biomedizinischer Artikel mit MeSH-Klassen in Echtzeit. Die teilnehmenden Teams werden gebeten, innerhalb von 21 Stunden neue unklassifizierte MEDLINE-Artikel¹³, die ihnen von den Organisator*innen zugeteilt werden, zu klassifizieren. Sobald durch die MEDLINE-Kurator*innen manuell vergebene Annotationen zur Verfügung stehen, werden sie zur Bewertung der Klassifizierungsleistung der teilnehmenden Systeme herangezogen. Jedes Jahr wird eine neue Version der benötigten Trainingsdaten in Form neuer MeSH-annotierter Dokumente aus MEDLINE bereitgestellt. 2012 bestand das Trainingskorpus aus 10 Millionen Artikeln, 2022 waren es bereits 16,2 Millionen. Ein jährlicher Durchlauf umfasst 15 Wochen und etwa 150.000 Artikel. Genutzt werden englischsprachige Abstracts und Titel. Zur Bewertung werden neben flachen (z.B. Recall und Precision) auch hierarchische Metriken verwendet, um die Beziehung zwischen den Klassen zu berücksichtigen und die Schwere eines Fehlers bewerten zu können – je nachdem, wo in der Hierarchie eine falsche Annotation aufgehängt ist, kann sie inhaltlich völlig falsch oder z.B. thematisch nur zu allgemein sein.¹⁴ Die verwendeten Systeme nutzen zwei verschiedene Ansätze der Künstlichen Intelligenz: zum einen Verfahren der Merkmalsextraktion

10 MeSH: Medical Subject Headings Thesaurus – ein kontrolliertes und hierarchisch strukturiertes Vokabular, erstellt durch die National Library of Medicine (NLM), <<https://www.nlm.nih.gov/mesh/meshhome.html>>, Stand: 21.02.2023.

11 BioASQ challenge, <<http://bioasq.org/>>, Stand: 24.02.2023.

12 u.a. Google, Novartis, ETH Zürich, ZBMed, Universität Stuttgart.

13 MEDLINE: bibliografische Datenbank der National Library of Medicine (NLM), <https://www.nlm.nih.gov/medline/medline_overview.html>, Stand: 21.02.2023.

14 Siehe auch: A. Kosmopoulos, I. Partalas, E. Gaussier, G. Paliouras und I. Androutsopoulos: Evaluation Measures for Hierarchical Classification: A Unified View And Novel Approaches, <<https://arxiv.org/abs/1306.6802>>, Stand: 24.02.2023.

– hier geht der Trend zu neuronalen Wort- und Absatz-Embeddings – und zum anderen Verfahren der Konzeptvergabe, die zunehmend Methoden des Deep Learning mit tiefen neuronalen Netzen und Aufmerksamkeitsmechanismen¹⁵ nutzen. Die Verfahren haben sich deutlich verbessert. Der NLM Medical Text Indexer (MTI) – ein NLP¹⁶-basiertes System, das MeSH-Vorschläge generiert – hat sich durch die Umsetzung der Erkenntnisse aus den BioASQ Challenges im Laufe der Jahre um 10% verbessert. Die NLM ist 2022 dazu übergegangen, MEDLINE Zitate in PubMed ausschließlich automatisiert mit MeSH zu indexieren mit dem Ziel, eine umfassende und zeitnahe Indexierung aller Veröffentlichungen in MEDLINE zu erreichen. Peter Schäuble zeigte sich beeindruckt von den exzellenten Ergebnissen, betonte aber auch, dass der große Vorteil dieses Projektes in der großen Menge an Trainingsdaten sowie der Einschränkung auf das Fachgebiet der Biomedizin bestünde. Die DNB mit der Vielfalt an Sachgebieten und Sprachen habe es da weitaus schwerer, zu solchen Ergebnissen zu kommen.

Die Einbindung von coli-conc Mappings im DA-3

Im nächsten Beitrag stellten Uma Balakrishnan und Jakob Voß, beide Verbundzentrale des Gemeinsamen Bibliotheksverbundes (VZG), den Service „Einbindung von coli-conc Mappings in den DA-3“ vor. Coli-conc¹⁷ ist eine von der VZG angebotene Infrastruktur zur Bereitstellung, Verwaltung und Erstellung von Mappings und Konkordanzen zwischen Wissensorganisationssystemen (englisch Knowledge Organization Systems, kurz KOS).¹⁸ Die VZG garantiert den Weiterbetrieb des Projektes, dessen Förderung seit 2016 bestand und 2023 ausläuft. Cocoda ist ein Tool, das – ähnlich wie der DA-3 für die Sacherschließung – die computergestützte intellektuelle Erstellung von Mappings verschiedener Erschließungsvokabulare erleichtert. Beliebige KOS können eingebunden, vorhandene Mappings gesucht und bewertet sowie eigene Mappings angelegt werden.¹⁹ Durch die entsprechenden Mapping-Vorschläge ist es jetzt einfacher, eine Konkordanz z.B. zwischen RVK und DDC zu erstellen. Inzwischen beinhaltet Cocoda viele weitere Konkordanzen und Mappings.²⁰ Die Plattform dient der offenen Zusammenarbeit von Expert*innen und stellt Konkordanzen für eine freie gemeinsame Nutzung zur Verfügung. Coli-conc ist inzwischen direkt im DA-3 eingebunden und generiert über die vorhandenen Konkordanzen Vorschläge. Sollte man mit dem vorgeschlagenen Mapping nicht einverstanden sein, kann es direkt aus dem DA-3 heraus in Cocoda angepasst werden. Im DA-3 kann die Änderung dann unmittelbar übernommen werden. Dieser Vorgang wurde im Vortrag von David Rohrer live vorgeführt. Erforderlich für Änderungen in Cocoda ist ein Log-in über eine bestehende

15 Aufmerksamkeitsmechanismen sind eine Art von neuronaler Netzwerkschicht, die zu Deep-Learning-Modellen hinzugefügt werden kann. Sie ermöglichen es dem Modell, sich auf bestimmte Teile der eingegebenen Daten zu konzentrieren, indem sie den Daten unterschiedliche Gewichte zuweisen. Diese Gewichtung basiert in der Regel auf der Relevanz der einzelnen Teile der Daten für die jeweilige Aufgabe.

16 Die Verarbeitung natürlicher Sprache (Natural Language Processing, NLP) ist ein Teilbereich der künstlichen Intelligenz, der es Computern ermöglicht, menschliche Sprache zu verstehen, zu generieren und zu manipulieren.

17 Coli-conc, <<https://coli-conc.gbv.de/>>, Stand: 24.02.2023.

18 Der Service umfasst inzwischen mehrere webbasierte Dienste: das Mapping-Tool Cocoda, <<https://coli-conc.gbv.de/cocoda/>>, das Konkordanzregister BARTOC (Basic Register of Thesauri, Ontologies & Classifications), <<https://bartoc.org/registries>>, coli-rich, <<https://coli-conc.gbv.de/de/coli-rich/>> für die automatische Kataloganreicherung und coli-ana, <<https://coli-conc.gbv.de/coli-ana/>> für die Analyse und Zerlegung von DDC-Notationen, Stand: 24.02.2023.

19 Die Mappingtypen exact, close, narrower, broader und related können vergeben werden.

20 Z.B. 25 RVK-BK Konkordanzen mit 31.000 Mappings oder eine GND-LCSH Konkordanz mit ca. 43.400 Mappings.

GitHub-, Orcid- oder Wikidata-Kennung. Für einige Konkordanzen benötigt man gesonderte Rechte. Im DA-3 auf „Aktualisieren“ klicken, dann wird das neue Mapping übernommen.

Einbindung in den DA (Testsystem)

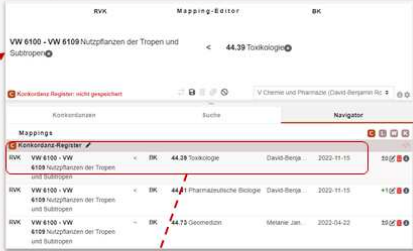
1. Titel suchen



2. Vorschläge sichten



3. Neues Mapping in Cocoda anlegen



4. In DA aktualisieren



Abb.: Aktive Bearbeitung eines Mappings. Vortragsfolie Nr. 11

Empfehlungen, die im DA-3 auftauchen, sind Mappings, die über eine SRU-Schnittstelle abgefragt werden. Es wird die übergeordnete Stelle angezeigt, falls das direkte Mapping nicht gefunden werden kann. Jakob Voß ergänzte Informationen zu geplanten Erweiterungen. Zukünftig sollen auch die Relationstypen „closeMatch“ und Mappings ohne Relation eingebunden werden, um mehr, leider aber auch ungenauere Vorschläge zu erhalten. Des Weiteren sollen zusätzliche Informationen zum Mapping angezeigt werden, wie z.B. Autor*in und Datum des Mappings. Testanwender*innen wünschten sich unter anderem eine bessere Sichtbarkeit des coli-conc Mappings in der Liste der möglichen Vorschlagsgeneratoren des DA-3. Das Teilprojekt coli-rich dient der Anreicherung der Bibliothekskataloge durch die mittels Konkordanzen generierte Sacherschließung aus den vorhandenen Erschließungssystemen.²¹ So werden z.B. Titeldaten im K10plus über Mappings der Relationen exactMatch und narrowMatch mit BK-Sacherschließung angereichert und in eigens dafür vorgesehene Felder eingespielt.²² Uma Balakrishnan wies noch einmal auf die Bedeutung der Community hin. Ein Beitrag der Community an der Pflege der Konkordanzen sei erwünscht und dringend erforderlich, um das Projekt am Leben zu halten.

21 Ausführlicher im Hintergrund-Artikel: Balakrishnan, Uma; Voß, Jakob: Anreicherung der Sacherschließung durch Konkordanzen (coli-rich), in: VZG Aktuell, Neues aus der Zentrale, (2), 2021, S.19. Online: <https://www.gbv.de/informationen/Verbundzentrale/Publikationen/broschueren/vzg-aktuell/VZG_Aktuell_2021_02.pdf>, Stand: 24.02.2023.

22 Die eingespielten Daten werden mit der Zeichenkette „coli-conc RVK->BK“ in \$A gekennzeichnet. Mit der Suchanfrage f seq „coli-conc RVK->BK“ können die entsprechenden Datensätze aufgerufen werden. Die Anführungszeichen müssen manuell eingegeben werden.

Demonstration des DA-3

Zum Abschluss des ersten Workshoptages führte Imma Hinrichs von der UB Stuttgart die Funktionen des DA-3 vor. Sie demonstrierte die Titelsuche mit einer Systemnummer oder alternativ über die erweiterte Suche (mit allen Suchschlüsseln, die die SRU-Schnittstelle zulässt) und im Anschluss daran die Listenerstellung auf Basis der gefundenen Treffer. Vermittelt wurde ein Überblick der Bearbeitungsoberfläche mit den sechs verschiedenen Boxen, deren Größe sich individuell einstellen lässt. Neu ist die Möglichkeit, zwischen zwei individuell eingerichteten Layouts zu wechseln – je nachdem, ob beispielsweise gerade der Laptopbildschirm oder ein anderes Bildschirmformat verwendet wird. Imma Hinrichs demonstrierte die Möglichkeiten der verschiedenen Tools für die Normdatensuche und die praktische Verwendung des sogenannten Scratchpad als Ablage für den schnellen Zugriff auf häufig genutzte Daten. Die Vorschlagsgeneratoren – das Herzstück des DA-3 – wurden in ihrer Funktion, möglichen Einstellungen und Hintergrundinformationen zur Herkunft der Daten erläutert. Der Einblick in die Anzeige bereits vorhandener Daten in der Kurzinfo-Box (Titeldaten) oder im Erschließungsfenster (vorhandene Sacherschließung) rundete die anschauliche Präsentation der Kernfunktionen des DA-3 ab.

Halbautomatische Vergabe der DDC-Sachgruppen²³

Am zweiten Workshoptag stellte Sebastian Gabler, Semantic Web Company, unter dem Titel „Thesaurus-unterstützte Sacherschließung“ ein Verfahren vor, mit dem DDC-Klassen für wissenschaftliche Publikationen (halb-)automatisiert vergeben werden können.²⁴ Dabei korrespondiert der Ansatz mit einer agilen Sacherschließung, die nicht nur auf wachsenden Bestand, aktuelle Retrievalmöglichkeiten sowie die heterogenen zugelieferten und maschinell erstellten Metadaten reagiert, sondern im besten Fall als polyzyklischer, kollaborativer Prozess sowohl Titeldaten als auch Metadaten immer wieder erzeugt, anreichert und auch kuratiert. Das Verfahren untersucht Volltexte mittels einer computerlinguistischen Methode und ermittelt anhand der GND-Normdaten passende Schlagwörter, die nach ihrer Relevanz im Text geordnet, gewichtet und schließlich nach mathematischen Modellen den entsprechenden DDC-Sachgruppen zugeordnet werden. Dazu kommt mit dem PoolParty Indexer, der einzelne Elemente und Konzepte aus dem Text extrahiert, ein Service der PoolParty Semantic Suite zum Einsatz.²⁵ Die DDC-Stellen werden entsprechend der Verortung eines Schlagwortes in der monohierarchischen Klassifikation vergeben: ist das gefundene Schlagwort Unterbegriff eines Elementes einer Sachgruppe, wird das Dokument dieser übergeordneten Klasse zugeordnet (transitiver Schluss). Dafür werden die bereits vorhandenen Anreicherungen der GND-Normsätze mit den passenden DDC-Notationen genutzt.²⁶ Da der auf Linked Data Prinzipien aufgebaute Ansatz nicht

23 DDC: Dewey Dezimal-Klassifikation - ein System zur Ordnung von Wissen, das in vielen Bibliotheken weltweit verwendet wird: <<https://www.dnb.de/wasistdieddc>>, Stand: 12.05.2023.

24 Sebastian Gabler entwickelte das Verfahren im Rahmen seiner Masterthesis an der Universität Wien. Siehe dazu <<https://theses.univie.ac.at/detail/60927/>>, Stand: 17.02.2023, sowie die Folien zum Vortrag.

25 Die Technologieplattform PoolParty Semantic Suite der Semantic Web Company ermöglicht das Management von Taxonomien, Ontologien und Wissensgraphen. Siehe dazu <<https://www.poolparty.biz/product-overview>>, Stand: 10.03.2023.

26 Diese Anreicherung wurde im Projekt CrissCross bereits begonnen und wird weitergeführt. Siehe dazu <<https://ixtrieve.fh-koeln.de/crisscross/index.html>>, Stand: 08.02.2023. Eine umfassende Anreicherung der GND mit den

auf maschinellem Lernen basiert, kommt er ohne Trainingsdaten aus. Zur Validierung des Verfahrens wurde ein Goldstandard mit 700 geeigneten Volltexten gebildet und die Indexierung dieses Referenzkorpus mit den extrahierten Metadaten verglichen und bewertet. Die Ergebnisse waren stets nachvollziehbar, stimmig mit den entsprechenden Regelwerken und für viele Bereiche auch vollständig und genau, was Sebastian Gabler anhand der Parameter Recall, Precision, F1-Score sowie des Mean Reciprocal Rank demonstrierte.²⁷ Das Verfahren eignet sich insgesamt sehr gut zur Ergänzung und Verbesserung automatisch und intellektuell erstellter Indexate und zur kollaborativen Erschließung und sollte durch die weitere Pflege der GND für alle Sachgebiete ermöglicht werden.

Direktlinks im DA-3

Helga Karg von der Deutschen Nationalbibliothek stellte in ihrem Beitrag die „Verbindung vom DA-3 zur WinIBW und zu WebDewey“ im DNB-Profil vor. Für beide Aspekte stellte der DA-3 bisher schon Möglichkeiten zur Verfügung, die nun jedoch entscheidend erweitert wurden. Zur Verbindung zur WinIBW: Bisher war es im DNB-Profil möglich, aus dem DA-3 heraus zum Titeldatensatz in der WinIBW zu wechseln. Auch eine detailliertere Ansicht der Normdaten war schon integriert: in der Box „Info“ führte ein Klick auf „GND“ zu einer Gesamtdarstellung des GND-Schlagwortes mit Definitionen, Verwendungshinweisen und sonstigen Informationen. Die ausführlichen Beziehungen zu hierarchisch und anderweitig verknüpften Schlagworten wurden durch einen weiteren Klick sichtbar. Per Klick auf die GND-Nummer war auch ein Wechsel zum entsprechenden Normsatz in der WebGND möglich.²⁸ Diese Ansichten sind für jede Sacherschließung äußerst hilfreich – für die Verknüpfung der bestpassenden Schlagworte. Sobald jedoch Änderungen oder Ergänzungen im Normdatensatz nötig wurden, musste eigens der GND-Zugang im ILTIS-Produktionssystem der WinIBW geöffnet werden, da nur dort eine umfassende Normdatenpflege möglich ist: einzelne Informationen ergänzen oder korrigieren, neue Beziehungen zu anderen Normsätzen herstellen, Mailboxen schreiben usw. Deshalb wurde ein Direktlink in den DA-3 eingefügt, der direkt in die PICA-Ansicht des Normdatensatzes führt.²⁹ So können die Normsätze bequem aus dem DA-3 heraus bearbeitet werden. Ein weiterer großer Vorteil dieses neuen Links besteht darin, dass nun auch eine Umfeldrecherche möglich ist: Welche Titel wurden bisher mit dem Schlagwort erschlossen? Mit welchen anderen Normdatensätzen wurden Schlagwortfolgen gebildet? Zur Verbindung zu WebDewey: Bislang wird in allen Profilen des DA-3 auf WebDeweySearch verlinkt, ein kostenfreies Portal.³⁰ Es ermöglicht die Suche nach Klassen, auch eine verbale Recherche und Browsing und zeigt mit ihnen verknüpfte Literatur verschiedener angebundener Kataloge. Ein Link auf WebDewey Deutsch, das Tool für das Klassifizieren mit der DDC, fehlte bislang und wurde nun im DNB-Profil eingerichtet. WebDewey

DDC-Klassen in Feld 083 sowie die fortlaufende Kuratierung des GND-Bestandes sind jedoch nötig, um eine breit gefächerte Nutzung des Verfahrens zu ermöglichen.

27 Der Mean Reciprocal Rank (MRR) dient der Bewertung eines Algorithmus. Er berücksichtigt, auf welchem Rang das erste relevante Element von möglichen Antworten auf eine Abfrage steht. Je näher der Einzelwert an der 1 liegt, desto genauer ist die Antwort und desto höher ihr Rang. Der mittlere reziproke Rang gibt einen Mittelwert für eine Reihe von Abfragen an.

28 WebGND von Eurospider, <<http://gnd.eurospider.com/s>>, Stand: 08.02.2023.

29 Aus dem DA-3 gelangt man in die WinIBW per Klick auf den Pfeil neben dem bevorzugten Namen in der Box „Info“ bei der Anzeige des GND-Schlagwortes.

30 WebDeweySearch, <<https://deweysearchde.pansoft.de/webdeweysearch/>>, Stand: 08.02.2023.

Deutsch ist ein Lizenzprodukt von OCLC und steht daher nur den lizenznehmenden Institutionen zur Verfügung.³¹ Der entscheidende Vorteil dieser Anwendung ist eine Kopierfunktion, mit der Notationen und ihre Bestandteile sowie synthetische Notationen einfach in die Katalogisierungsumgebung des K10plus, die WinIBW, integriert werden können.³² Helga Karg demonstrierte sowohl die Verlinkung aus dem DA-3 heraus, die aus den verschiedenen Boxen „Info – GND“, „Vorschläge“ und „Erschließung“ möglich ist, als auch die Kopierfunktion sowie wiederum die Anzeige im DA-3. In der regen Anschlussdiskussion kam zur Sprache, dass die Verknüpfung zwischen GND-Schlagwörtern und der DDC, wie sie im Projekt CrissCross³³ begonnen wurde, bereits über 150.000 Verbindungen enthält, jedoch längst noch nicht alle Fachgebiete umfasst. Außerdem wurde angeregt, auch die Ergebnisse von coli-ana in den DA-3 einzubinden.

Anwendung des DA-3 für kooperative Aufsatzerschließung

Martin Faßnacht und David Cloutier aus der UB Tübingen stellten in ihrem Vortrag „Artikellerschließung und Suche mit Schlagwortfolgen“ dar, wie sie den DA-3 im Kontext der Fachinformationsdienste Theologie und Religionswissenschaft für die Erschließung selbstständiger und unselbstständiger Literatur nutzen. Die Sacherschließung erfolgt kooperativ durch mehrere Institutionen, welche zwar nicht alle direkt im Katalogisierungstool WinIBW, aber bereits oder demnächst aktiv mit dem DA-3 arbeiten. Geplant ist vor allem eine multilinguale Erschließung, wofür das betreffende Vokabular bereits in neun Sprachen übersetzt wurde. Diese Übersetzungen sollten in den DA-3 integriert werden, damit sie dort allen DA-3-Nutzenden zur Verfügung stehen. Mit „MeisterTask“, in das David Cloutier Einblicke gewährte, wird die Sacherschließung in ihrem Ablauf organisiert. In diesem Projektmanagementtool gelingt es, die erfassten Zeitschriftenhefte außerhalb der WinIBW zeitnah für die Sacherschließung freizugeben und deren Bearbeitung durch die verschiedenen, auch externen Sacherschließenden transparent zu kennzeichnen sowie spätere Arbeiten in der GND kooperativ vorzubereiten oder fallbezogene Kommentare auszutauschen. In einem weiteren Teil zeigte Martin Faßnacht, wie das Discoverysystem des „Index Theologicus“ mit der neuen Funktionalität „Thema browsen“ ein benutzergesteuertes Browsing ermöglicht, das gut zwischen Suchen und Browsen vermittelt.³⁴ Nach einer Eingabe im Suchschlitz werden bereits vergebene Schlagwortfolgen aufgelistet. Einzelne Schlagwörter dieser Folgen können per Shift und Klick in eine weitere thematische Suche übernommen werden. Passt eine vorhandene Schlagwortfolge gut, kann man auch per Direktklick „Datensätze ansehen“.

Austausch der Community zum DA-3

Der letzte Abschnitt des Workshops war der „Diskussion und Erfahrungsaustausch der DA-3 Anwender“ gewidmet. Geleitet wurde er durch Regine Beckmann (Staatsbibliothek zu Berlin - Preußischer

31 WebDewey Deutsch, <<https://deweyde.pansoft.de/webdewey/login/login.html>>, Stand: 08.02.2023.

32 Berücksichtigt werden können alle Notationen, die durch Synthese in Abhängigkeit vom Titel gebildet und in WebDewey Deutsch abgelegt wurden. Das trifft auf besonders häufig genutzte Notationen zu.

33 CrissCross, <<https://ixtrieve.fh-koeln.de/crisscross/index.html>>, Stand: 08.02.2023.

34 Index Theologicus, Internationale Bibliographie für Theologie und Religionswissenschaft, <<https://ixtheo.de/>>, Stand: 17.02.2023.

Kulturbesitz, SBB) und Imma Hinrichs (UB Stuttgart) und ermöglichte den Austausch der zu diesem Zeitpunkt immer noch ca. 130 Teilnehmer*innen. Zunächst stellte Imma Hinrichs noch einmal gebündelt die neueren Entwicklungen des DA-3 vor, zu denen nicht nur die Einbindung von coli-conc und des Bewertungstools der ZBW zählen,³⁵ sondern auch die komfortable Nutzung alter Suchanfragen, die bequeme Navigation zwischen über- und untergeordneten Werken sowie die individuelle Gestaltungsmöglichkeit mittels zweier Layouts. In der anschließenden Feedbackrunde betonten die Nutzenden das bequeme Handling des Tools, wie gut sich der DA-3 in den Workflow einfüge und dass der DA-3 eine großartige Unterstützung im Erschließungsalltag darstelle. Außerdem biete er mehr Möglichkeiten, auch mit Partnern zu kooperieren, die nicht direkt im Katalogisierungssystem arbeiteten. Diskutiert wurde bspw. darüber, ob nachträgliche Anpassungen an Titeln vereinfacht werden könnten (bzw. ob das erwünscht ist) und ob bzw. wie eine Übertragung von Sacherschließungsdaten auf Werkcluster auch unabhängig vom eigenen Bestand möglich wäre. Der Wunsch nach Multilingualität, sowohl für die Oberfläche des DA-3 als auch für die Schlagwörter/Schlagwortsuche, wurde deutlich ausgesprochen. Christoph Steiner (Leiter der Sacherschließung der Österreichischen Nationalbibliothek, ÖNB, und Vertreter des OBV im DA-3 Advisory Board) gab durch dreijährige österreichische DA-3-Erfahrung gewonnene Praxistipps, z.B. wie man mit Literatur ohne ISBN-Barcode verfahren kann oder wie das Scratchpad beim Zeitsparen helfen kann. Obwohl es sich beim DA-3 um ein zusätzliches Tool und so um eine mögliche Verkomplizierung der Arbeitsabläufe handele, bliebe mehr Zeit für konzeptionelle Gedanken, mache die Sacherschließung mehr Spaß und es sei in der Nutzung mehr Sachverstand als technische Kompetenz gefragt.

Website zum DA-3

Regine Beckmann präsentierte schließlich die Entwürfe für eine zentrale Website zum Digitalen Assistenten und für ein Wiki, das als gemeinsame Arbeitsplattform dienen soll. Auf der Website werden Informationen zum Tool und zu Veranstaltungen sowie Formulare und nach Themenschwerpunkten sortierte FAQ angeboten.³⁶ Über einen Link gelangt man zum Wiki, in dem Unterlagen zur Bedienung, zu Updates, zu Besonderheiten des DA-3 im K10plus sowie zu Konferenzen und Publikationen ausgetauscht werden können.³⁷ Hier rief Regine Beckmann explizit zu einer regen Nutzung und Kollaboration, zum aktiven Teilen von Präsentationen und anderen Hilfsmitteln auf.

Fazit

Der DA-3 wächst mit seiner aktiven Nutzung und durch fortlaufende Anpassungen. Im Workshop wurde deutlich, wie Suchräume erweitert, wie Vokabularen noch sinnvoller vernetzt und genutzt werden können, dass Multilingualität wünschenswert wäre, wie sich Gewohnheiten und Workflows durch die Verfahren stetig ändern und wie die Inhaltsschließung in einer Interaktion zwischen Mensch und Maschine (HITL) immer besser werden kann. Rege Diskussionen und eine hohe Beteiligung bis zum Schluss zeigten, wie wichtig diese Themen sind. Mit großer Dankbarkeit aller Teilnehmenden

³⁵ Siehe dazu die Beiträge von Anna Kasprzik und Uma Balakrishnan.

³⁶ Website zum Digitalen Assistenten DA-3, <<https://www.da-3.de/>>, Stand: 10.03.2023.

³⁷ Wiki zum Digitalen Assistenten DA-3, <<https://wiki.k10plus.de/x/KwAul>>, Stand: 20.03.2023.

und in Vorfreude auf ein nächstes Mal ging der Workshop zu Ende. Bis zum nächsten Workshop gilt: Nutzen wir die schon vorhandenen Möglichkeiten, denken wir sie weiter, bringen wir unsere Ideen ein und die computerunterstützte Inhaltserschließung weiter voran!

Susanne Schmucker, ZBW - Leibniz-Informationszentrum Wirtschaft, Hamburg/Kiel,

<https://orcid.org/0009-0006-3756-1356>

Cornelia Schöntube, Staatliches Institut für Musikforschung Preußischer Kulturbesitz, Berlin,

<https://orcid.org/0009-0006-1443-5752>

Zitierfähiger Link (DOI): <https://doi.org/10.5282/o-bib/5928>

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