

Aufbau eines produktiven Dienstes für die automatisierte Inhaltserschließung an der ZBW

Ein Status- und Erfahrungsbericht

Anna Kasprzik, ZBW – Leibniz-Informationszentrum Wirtschaft, Hamburg/Kiel

Zusammenfassung

Die ZBW – Leibniz-Informationszentrum Wirtschaft betreibt seit 2016 eigene angewandte Forschung im Bereich Machine Learning mit dem Zweck, praktikable Lösungen für eine automatisierte oder maschinell unterstützte Inhaltserschließung zu entwickeln. 2020 begann ein Team an der ZBW die Konzeption und Implementierung einer Softwarearchitektur, die es ermöglichte, diese prototypischen Lösungen in einen produktiven Dienst zu überführen und mit den bestehenden Nachweis- und Informationssystemen zu verzahnen. Sowohl die angewandte Forschung als auch die für dieses Vorhaben („AutoSE“) notwendige Softwareentwicklung sind direkt im Bibliotheksbereich der ZBW angesiedelt, werden kontinuierlich anhand des *State of the Art* vorangetrieben und profitieren von einem engen Austausch mit den Verantwortlichen für die intellektuelle Inhaltserschließung. Dieser Beitrag zeigt die Meilensteine auf, die das AutoSE-Team in zwei Jahren in Bezug auf den Aufbau und die Integration der Software erreicht hat, und skizziert, welche bis zum Ende der Pilotphase (2024) noch ausstehen. Die Architektur basiert auf Open-Source-Software und die eingesetzten Machine-Learning-Komponenten werden im Rahmen einer internationalen Zusammenarbeit im engen Austausch mit der Finnischen Nationalbibliothek (NLF) weiterentwickelt und zur Nachnutzung in dem von der NLF entwickelten Open-Source-Werkzeugkasten Annif aufbereitet. Das Betriebsmodell des AutoSE-Dienstes sieht regelmäßige Überprüfungen sowohl einzelner Komponenten als auch des Produktionsworkflows als Ganzes vor und erlaubt eine fortlaufende Weiterentwicklung der Architektur. Eines der Ergebnisse, das bis zum Ende der Pilotphase vorliegen soll, ist die Dokumentation der Anforderungen an einen dauerhaften produktiven Betrieb des Dienstes, damit die Ressourcen dafür im Rahmen eines tragfähigen Modells langfristig gesichert werden können. Aus diesem Praxisbeispiel lässt sich ableiten, welche Bedingungen gegeben sein müssen, um Machine-Learning-Lösungen wie die in Annif enthaltenen erfolgreich an einer Institution für die Inhaltserschließung einsetzen zu können.

Summary

Since 2016, ZBW – Leibniz Information Centre for Economics has been conducting their own research in the area of machine learning with the goal to develop viable solutions for automated or machine-assisted subject indexing in-house. In 2020, a team at ZBW started designing and implementing a suitable software architecture in order to transfer these prototypical solutions into a productive service and to integrate it into the existing metadata systems and workflows. Both the applied research and the software development necessary for this endeavour (dubbed “AutoSE”) are executed by an organizational unit of the library department of ZBW, are continually pushed forward following the state of the art and benefit from a close communication with the staff responsible for intellectual subject indexing. This article reports on the milestones that the AutoSE team has reached over the last two years with respect to the implementation and the integration of the software and outlines those that are yet to be delivered until the end of the pilot phase (2024). The architecture is based on open source software and its machine-learning-based components are developed in close communication

with the National Library of Finland (NLF) and, where possible, adapted to be integrated into NLF's open source toolkit Annif. The operating model of the AutoSE service includes periodical reviews of individual components and of the productive workflow in its entirety and allows continuous improvements of the architecture. One of the results to be delivered by the end of the pilot phase is a documentation of the requirements for running the productive service on a permanent basis so that the necessary resources can be secured. This practical example shows which conditions have to be met by an institution in order to successfully use machine learning solutions such as the ones offered in Annif for subject indexing.

Zitierfähiger Link (DOI): <https://doi.org/10.5282/o-bib/5903>

Autorenidentifikation:

Kasprzik, Anna: ORCID: <https://orcid.org/0000-0002-1019-3606>; GND: 1022595687

Schlagwörter: Inhaltserschließung, Automatisierung, Machine Learning, Metadaten, IT-Infrastruktur, Personalressourcen, human in the loop

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

1. Weichenstellung für den Transfer von Ergebnissen aus der angewandten Forschung in den Bibliotheksbetrieb

Die ZBW – Leibniz-Informationzentrum Wirtschaft ist die weltweit größte Informationsinfrastruktur für wirtschaftswissenschaftliche Literatur, pflegt als solche ihr eigenes normiertes Vokabular – den Standard-Thesaurus Wirtschaft (STW)¹ – und ist oft die erste Einrichtung, die eine wirtschaftliche Informationsressource erschließt, kann also in vielen Fällen keine Fremderschließung nachnutzen. Entsprechend wurde die Automatisierung der Erschließung über die letzten beiden Jahrzehnte hinweg zu einem immer dringlicheren Anliegen. Seit ca. 2016 betreibt die ZBW eigene angewandte Forschung im Bereich Machine Learning mit dem Zweck, praktikable Lösungen für eine automatisierte oder zumindest maschinell unterstützte Inhaltserschließung zu entwickeln.² Die in dem Projekt AutoIndex (bis Ende 2018) entstandenen Prototypen waren jedoch ihrem Forschungscharakter gemäß in der Erschließungspraxis nicht direkt als Dienst einsetzbar: Der Code erfüllte nicht die für einen Produktivbetrieb erforderlichen Standards³ und es fehlte eine umgebende IT-Infrastruktur, um den Dienst mit den anderen Metadatenverarbeitungsprozessen im Haus sinnvoll zu verzahnen.

1 Standard-Thesaurus Wirtschaft (STW), <<https://zbw.eu/stw/version/latest/about.de.html>>, Stand: 30.09.2022.

2 Siehe auch Tochtermann, Klaus; Kasprzik, Anna: Auf Augenhöhe mit Forschungspartnern aus der Wissenschaft – Anwendung von Künstlicher Intelligenz in der ZBW, in: BuB – Forum Bibliothek und Information 74 (6), 2022, S. 306–311. Online: <<https://pub.zbw.eu/dspace/bitstream/11108/526/2/2022-Kasprzik-Tochtermann-Augenh%C3%B6he.pdf>>.

3 Etwa automatisierte Tests des Codes, Code Reviews (Überprüfung des Codes durch mehr als eine Person), eine automatisierte Überwachung (Monitoring) des Betriebs und Maßnahmen zu dessen Ausfallsicherheit.

2019 wurden nach zwei Personalwechseln erneut die konkreten Anforderungen für einen erfolgreichen Transfer der bisher erarbeiteten Lösungen in die Praxis evaluiert. Ein zentraler strategischer Meilenstein hierbei war es, den Projektstatus hinter sich zu lassen und die Automatisierung der Inhaltsererschließung (fortan „AutoSE⁴“) durch die Leitung der ZBW zu einer Daueraufgabe erklären zu lassen. In der Folge wurde für die Konzeption und Implementierung einer geeigneten Softwarearchitektur eine Pilotphase bis 2024 ausgerufen und eine weitere Softwareentwicklungsstelle für diese Aufgabe eingerichtet. Die wissenschaftliche Methodenentwicklung sollte parallel dazu weitergeführt werden und deren Ergebnisse sollten ständig über Experimentier- und Testphasen in das produktive System einfließen. Das Team besteht seither aus drei Personen für die drei Rollen „Leitung/Koordination“, „angewandte Forschung“ und „Softwarearchitekturentwicklung“. Mit den beiden Abteilungen, die für die IT-Entwicklung und für den Alltagsbetrieb der IT-Infrastruktur zuständig sind, und auch mit der an der ZBW als Leibniz-Einrichtung ebenfalls existierenden Forschungsabteilung wird anlassbezogen eng zusammengearbeitet – alle drei Team-Mitglieder sind jedoch bibliotheksnah in derselben Abteilung angesiedelt wie die intellektuelle Inhaltsererschließung auch („Wissenschaftliche Dienste“, im Programmbereich „Bestandsentwicklung und Metadaten“), was sich im Gestaltungsprozess der benötigten Funktionalitäten als ein enormer Vorteil erwies.

2. Fortgesetzte angewandte Forschung und Produktivbetrieb

2.1 Methoden

Im Vorläuferprojekt AutoIndex wurde ein prototypischer Ansatz entwickelt, der mehrere Modelle miteinander fusionierte und das Ergebnis noch einmal anhand von Regeln filterte.⁵ Etwa zur selben Zeit entstand an der finnischen Nationalbibliothek (NLF) das Open-Source-Toolkit Annif⁶, welches verschiedene Machine-Learning-Modelle zum Einsatz anbietet, aber auch das Verwenden eigener Modelle erlaubt. Schon damals bestand ein Austausch zwischen den beiden Institutionen, und 2019 waren sowohl das Toolkit als auch die Aktivitäten an der ZBW weit genug gereift, dass AutoSE Annif als Kernmodul übernehmen konnte. Dieses wurde dann mit einer Reihe weiterer Komponenten flankiert, die für einen vollen Produktivbetrieb gebraucht werden. Hierzu gehören Mechanismen für die Hyperparameteroptimierung⁷, die Automatisierung des Modelltrainings, die Qualitätskontrolle und der Anschluss an die sonstigen Erschließungsworkflows im Haus. Das AutoSE-Team arbeitet an der Weiterentwicklung von Annif mit, veranstaltet zusammen mit der NLF Tutorials dazu⁸ und berät andere Institutionen in Deutschland zum Einsatz des Toolkits.⁹

4 AutoSE, <<https://www.zbw.eu/de/ueber-uns/arbeitsschwerpunkte/automatisierung-der-erschliessung>>, Stand: 30.09.2022.

5 Toepfer, Martin; Seifert, Christin: Fusion architectures for automatic subject indexing under concept drift, in: International Journal on Digital Libraries 21, 2018, S. 169–189. Online: <<https://doi.org/10.1007/s00799-018-0240-3>>

6 Annif, <<http://annif.org/>>, Stand: 30.09.2022.

7 D.h. die Suche nach optimalen Hyperparametern. Ein Hyperparameter ist ein Parameter, der zur Steuerung des Trainings verwendet wird und dessen Wert daher vor dem Training des Modells festgelegt werden muss.

8 Die Materialien sind zum Selbststudium verfügbar unter: Annif-Tutorial, <<https://github.com/NatLibFi/Annif-tutorial>>, Stand: 30.09.2022.

9 Unter anderem auch die DNB bei der Entwicklung ihrer „Erschließungsmaschine“ (EMA). Zur EMA siehe auch Busse, Frank; Grote, Claudia; Jacobs, Jan-Helge u.a.: Erschließungsmaschine gestartet, 04.05.2022, <<https://blog.dnb.de/erschliessungsmaschine-gestartet/>>, Stand: 30.09.2022.

Die AutoSE-Software kombiniert mit Hilfe von Annif mehrere Machine-Learning-Modelle, darunter auch eine Eigenentwicklung (stwfsa)¹⁰, und wendet auf das von Annif zurückgelieferte Ergebnis diverse eigene Filter und Regeln an. Parallel dazu wird im Zuge der angewandten Forschung für AutoSE mit weiteren Machine-Learning-Ansätzen experimentiert, etwa aus dem Deep Learning – insbesondere mit Transformermodellen, welche sich besonders für eine multilinguale Verschlagwortung anbieten.¹¹ Da Experimente im Bereich Machine Learning eine beträchtliche Rechenleistung erfordern, hatte die ZBW 2020 für das Training von Modellen neue Hardware einschließlich zweier Graphikkarten erworben. Das Team sammelt und implementiert jedoch auch Ideen, wie der Ressourcenbedarf beim Training verringert werden kann, und arbeitet zusammen mit der NLF daran, sowohl Transformermodelle als auch diese Optimierungsmaßnahmen in Annif zu integrieren, um sie für eine größere Community von Einrichtungen mit unterschiedlichen Ressourcenvoraussetzungen nachnutzbar zu machen.

Was den Produktivbetrieb anbelangt, so basiert der AutoSE-Dienst auf einem Kubernetes-Cluster mit fünf Knoten (betrieben auf fünf virtuellen Maschinen) und greift für Steuerung, Überwachung, das Integrieren neuer Komponenten und das Ausrollen neuer Versionen der Software auf diverse weitere Schlüsseltechnologien zurück.¹² Die erste produktive Version wurde im Frühjahr 2021 in Betrieb genommen und der Dienst seither Schritt für Schritt ausgebaut und optimiert. Die Datenflüsse rund um AutoSE sind in Abb. 1 dargestellt; der nächste Abschnitt enthält einige Erläuterungen dazu.

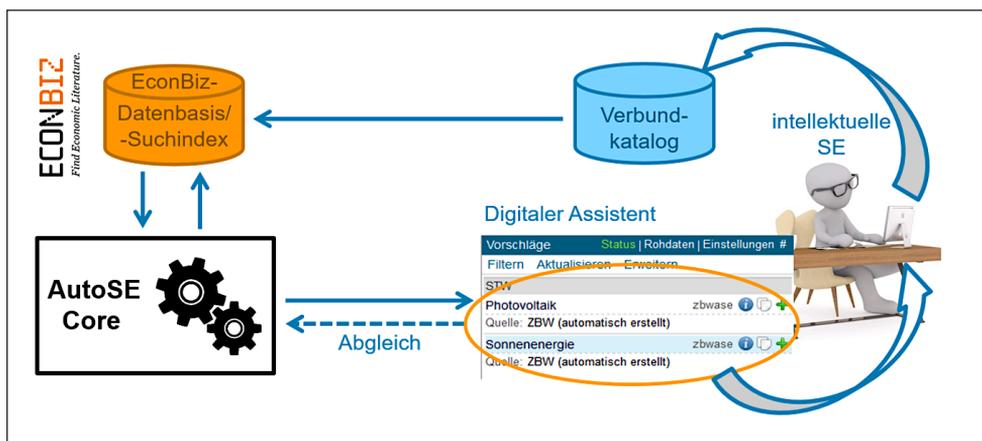


Abbildung 1: Datenflüsse rund um AutoSE

- 10 stwfsa, <<https://github.com/zbw/stwfsapy>>, Stand: 30.09.2022, und stwfsa in Annif, <<https://github.com/NatLibFi/Annif/wiki/Backend%3A-STWFSA>>, Stand: 30.09.2022.
- 11 Für eine niederschwellige Erklärung von Transformern siehe hier unter dem Stichwort „BERT“: Winkler, Christian: Wer, wie, was. Textanalyse über Natural Language Processing mit BERT, heise online, 12.08.2020, <<https://www.heise.de/hintergrund/Wer-wie-was-Textanalyse-mit-BERT-4864558.html>>, Stand: 30.09.2022.
- 12 Für weitere Details siehe Bartz, Christopher: Software Architecture for the Automatization of Subject Indexing. Vortrag bei der ELAG 2022 in Riga, Litauen (8. bis 10. Juni 2022). Foliensatz online unter: <<https://elag2022.inb.lv/programme/schedule/>>

2.2 Vollautomatisierte und maschinell unterstützte Inhaltserschließung

Der Dienst prüft stündlich die Datenbasis des Rechercheportals der ZBW – EconBiz¹³ – auf neue oder geänderte Metadatensätze, und wenn diese gewisse Kriterien erfüllen, werden sie unverzüglich mit automatisiert erzeugter Verschlagwortung versehen und zurückgeschrieben. Im Moment werden die eingesetzten Modelle auf englischsprachigen Daten trainiert und entsprechend können auch nur englischsprachige Metadaten zur Verschlagwortung herangezogen werden.¹⁴ Aus den Metadaten werden aktuell der Titel und, wenn vorhanden, Autoren-Keywords verwendet. Die Auswertung von Abstracts und Inhaltsverzeichnissen ist in Prüfung, allerdings muss für eine flächendeckende Verwendung von ersteren noch ein praktikabler Umgang mit einer relativ heterogenen Rechtslage entwickelt werden: Für viele lizenzierte Ressourcen ist die Anwendung automatisierter Methoden auf dem Textmaterial für einen produktiven Dienst im Bibliothekskontext nicht explizit erlaubt, selbst wenn sie es für Forschungszwecke ist – ein Indiz dafür, wie klein die Zahl produktiver Dienste auf Machine-Learning-Basis in der Bibliothekswelt aktuell noch ist, sonst ließe sich die rechtliche Klärung sicher kooperativ mit mehr Dringlichkeit verfolgen.

Einer der ersten Meilensteine war die Einrichtung eines Schreibzugriffs auf die EconBiz-Datenbasis im Juli 2021 – bis zu diesem Zeitpunkt wurde maschinell generierte Verschlagwortung jährlich von Hand im Batchverfahren eingespielt. In den folgenden sechs Monaten konnten bereits über 100.000 Ressourcen per Schreibzugriff mit AutoSE-Daten angereichert werden. Zusammen mit den weiterhin in unregelmäßigen Abständen eingespielten Batchdaten (zum Beispiel, um bereits mit AutoSE-Daten versehene Ressourcen mit einem verbesserten Verfahren noch einmal zu verschlagworten) enthält die EconBiz-Datenbasis nun über 1,4 Mio. automatisiert verschlagwortete Metadatensätze (Stand September 2022).

Zusätzlich werden diese Schlagwörter über eine Schnittstelle dem Digitalen Assistenten (DA-3)¹⁵ als Vorschläge für die intellektuelle Inhaltserschließung bereitgestellt. Die Plattform DA-3 zeigt zur Unterstützung des Sacherschließungsvorgangs zu einer Ressource bereits vorhandene Verschlagwortung aus Fremdquellen an – AutoSE („*zbwase*“) ist also nur eine weitere Quelle unter vielen, in der Anzeige werden AutoSE-Daten allerdings mit dem Vermerk „automatisch erstellt“ versehen (siehe auch Abb. 2). Bisher sind diese Vorschläge nur im DA-3-Profil des Bibliotheksverbundes K10plus zu sehen, dem die ZBW angehört. Die durch den wissenschaftlichen Dienst der ZBW (über den DA-3 oder über die in Ausnahmefällen weiterhin verwendete Plattform WinIBW) intellektuell erstellte Inhaltserschließung wird direkt in den K10plus-Verbundkatalog geschrieben und von dort wiederum in die EconBiz-Datenbasis gespiegelt (siehe nochmals Abb. 1). Dadurch bietet sich für das AutoSE-Team eine Möglichkeit, die Ergebnisse der maschinellen Erschließung zu überprüfen, indem es diese intellektuell vergebenen Schlagwörter abgreift und mit dem automatisiert erzeugten Output

13 EconBiz, <<https://www.econbiz.de>>, Stand: 30.09.2022.

14 Um auf eine zukünftige multilinguale Verschlagwortung hinzuwirken, wird mit Transformermodellen experimentiert, siehe Abschnitt 2.1.

15 Digitaler Assistent (DA-3), <<https://www.europider.com/de/relevancy-produkt/digitaler-assistent-da-3>>, Stand: 30.09.2022. Für mehr Informationen zur Entstehung des DA-3 siehe auch: Beckmann, Regine; Hinrichs, Imma; Janßen, Melanie u.a.: Der Digitale Assistent DA-3 – eine Plattform für die Inhaltserschließung, in: o-bib – das offene Bibliotheksjournal 6 (3), 2019, S. 1–20. Online: <<https://doi.org/10.5282/o-bib/2019H3S1-20>>

vergleicht. So ergibt sich zumindest ein binäres Qualitätsmaß im Sinne von „dieser Deskriptor wurde von der Maschine ebenfalls vorgeschlagen“ vs. „dieser Deskriptor hat bei den maschinellen Vorschlägen gefehlt“, woraus sich auch ein F1-Wert für die Summe der für eine Ressource vorgeschlagenen Deskriptoren berechnen lässt.¹⁶

Kurztitel	#	Vorschläge	Status	Rohdaten	Einstellungen	#
Nummer: 1032536500		STW				
Signature experience : art and science of customer engagement for fashion and luxury companies / edited by Stefania Saviolo		Beziehungsmarketing			zbwase	
Personen: Saviolo, Stefania [HerausgeberIn]		Quelle: ZBW (automatisch erstellt)				
Ausgabe: First edition		Konsumentenverhalten			zbwase	
Publ.: Milano : BUP, August 2018		Luxusgüter			zbwase	
ISBN: 978-88-99902-31-5, 978-88-85486-59-1		Markenführung			zbwase	
Sprache: Englisch [text]		Mode			zbwase	
Weitere Daten		GND				
		Beziehungsmarketing [Sach]	@stw-exact			
		Luxusout [Sach]	@stw-exact			

Abbildung 2: Anzeige von AutoSE-Vorschlägen im Digitalen Assistenten (DA-3)

3. Qualitätssicherung

Ein zentraler Aspekt bei der Automatisierung einer so essenziellen bibliothekarischen Kernaufgabe wie der Inhaltserschließung ist natürlich der Umgang mit dem Themenkomplex Qualität – welche Maße und Kriterien werden für sie definiert, wie werden die Werte dafür ermittelt, wie wird das angestrebte Qualitätsniveau sichergestellt und verbessert? Das AutoSE-Team arbeitet seit Beginn der Pilotphase an einem umfassenden Konzept mit mehreren verschiedenen Ansätzen, um für die von AutoSE herausgegebenen Daten ein gewisses Gesamtniveau garantieren zu können.¹⁷ Ein etablierter und auch durch das AutoSE-Team durchgängig angewandter Ansatz ist das Ermitteln von im Machine-Learning-Bereich gängigen Metriken wie etwa *precision*, *recall* und F1-Wert, um eine Orientierung aus wissenschaftlicher Perspektive für die Performanz der eingesetzten Methoden zu gewinnen und dann auf deren Optimierung hinzuwirken.

Für AutoSE ist jedoch noch ein weiterer Ansatz zur Qualitätseinschätzung, -sicherung und verbesserung von zentraler Bedeutung, sozusagen als Gegengewicht zur berechenbaren Metrik, und das

16 Der F1-Wert berechnet sich in diesem Fall aus dem harmonischen 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?) und ist damit eine Option für ein Maß der Gesamtqualität automatisiert erzeugter im Vergleich mit intellektuell erzeugter Verschlagwortung.

17 Merke: Es geht hier um die Gesamtmenge des Outputs pro Zeiteinheit. Das bedeutet nicht, dass einzelne Datensätze nicht trotzdem unpassende Deskriptoren enthalten können – das wäre angesichts der aktuell verfügbaren Trainingsdaten und Machine-Learning-Methoden ein unvernünftiger Anspruch.

ist der menschliche Input. Im Machine Learning spricht man vom *human in the loop*,¹⁸ also von einer intelligenten Verzahnung von maschinellen und menschlichen Aktivitäten zur Problemlösung. Mögliche Umsetzungen des *human in the loop* sind ein weites Feld – manche sind sehr naheliegend und scheinen daher trivial, etwa die Tatsache, dass Trainingsdaten für Anwendungen wie AutoSE in der Regel von Menschen annotiert sind, und die Tatsache, dass Wissensorganisationssysteme und Konkordanzen zwischen diesen in der Regel ebenfalls noch von Menschen erstellt und gepflegt werden. Das maschinell unterstützte Erschließen, bei dem der Mensch in letzter Instanz den Output „absegnen“ muss, fällt ebenso darunter wie das intellektuelle Bewerten von Stichproben des maschinellen Outputs. Wenn die Maschine dann zum Beispiel auf der Basis dieses Feedbacks unmittelbar neu trainiert wird, spricht man von Online-Learning. Und wenn die Maschine selbst entscheidet, an welchen Punkten sie einen Menschen interaktiv um ein Feedback zu einem bestimmten Datum bittet, spricht man von Active Learning – hier dient der Mensch sozusagen als Quelle für Trainingsdaten auf Abruf und die Maschine definiert ihren Bedarf beim Lernen selbst.

3.1 Intellektuelle Qualitätssicherung

Welche dieser Umsetzungen finden sich bereits in AutoSE wieder?¹⁹ Auch für AutoSE werden von Menschen annotierte Trainingsdaten und ein von Menschen redaktionell betreuter Thesaurus verwendet, und der Output wird über den DA-3 als Unterstützung bei der Inhaltserschließung angeboten. Was das intellektuelle Bewerten anbetrifft, so wurden bereits seit dem Vorläuferprojekt AutoIndex jährlich Reviews durchgeführt, bei denen eine Gruppe von Inhaltserschließenden an der ZBW nach ihrer Einschätzung der Qualität der automatisiert erzeugten Verschlagwortung für eine Stichprobe von rund 1000 Dokumenten gefragt wurde – und zwar sowohl für jeden Deskriptor einzeln als auch für die Verschlagwortung des jeweiligen Dokuments als Ganzes. Außerdem konnten Deskriptoren angegeben werden, die aus Sicht der Bewertenden zu einer angemessenen Verschlagwortung fehlten. Verwendet wurde hierfür eine Oberfläche namens *releasetool*, die in AutoIndex entwickelt wurde und es den Bewertenden erlaubt, die relevanten Metadaten (Titel, Keywords, Abstract) einzusehen, über einen Link auf den Volltext zuzugreifen, in der Liste der ihnen zugewiesenen Datensätze zu navigieren und den Fortschritt beim Abarbeiten zu verfolgen.²⁰

Diese Reviews waren unter anderem ein wertvolles Instrument zur Standortbestimmung im Übergang vom Vorläuferprojekt AutoIndex zur Daueraufgabe AutoSE. In Abb. 3 sind die Ergebnisse von 2019 und 2020 im Vergleich zu sehen. Grundlage des Reviews 2019 war eine Rekonstruktion der aus AutoIndex geerbten Methoden, 2020 wurden bereits im Kontext von AutoSE weiterentwickelte

18 Monarch, Robert M.; Manning, Christopher D.: Human-in-the-loop machine learning – active learning and annotation for human-centered AI. (E-Book), Manning Publications, 2021. Online: <<https://livebook.manning.com/book/human-in-the-loop-machine-learning/>>, Stand: 30.09.2022.

19 Für einen Artikel zu dieser Frage siehe auch Kasprzik, Anna: Get everybody on board and get going – the automation of subject indexing at ZBW, in: 87th IFLA World Library and Information Congress (WLIC), Satellite Meeting: Information Technology – New Horizons in Artificial Intelligence in Libraries, 2022. Online: <<https://repository.ifla.org/handle/123456789/2047>>.

20 *releasetool*, <<https://github.com/zbw/releasetool>>, Stand: 30.09.2022. Für einen Screenshot siehe z.B. Folie 19 von Kasprzik, Anna: Get everybody on board and get going – the automation of subject indexing at ZBW. Vortrag beim 87th IFLA World Library and Information Congress (WLIC), Satellite Meeting: Information Technology – New Horizons in Artificial Intelligence in Libraries am 22. Juli 2022 in Galway, Irland. Online: <<https://repository.ifla.org/handle/123456789/2047>>.

Methoden verwendet und bewertet. Der Fortschritt ist deutlich sichtbar, insbesondere in den besseren Stufen beider Bewertungskategorien. Es wurde jedoch auch offensichtlich, dass der Anteil an eindeutigen falschen Deskriptoren und unzureichend erschlossenen Dokumenten zur Erfüllung vorherrschender bibliothekarischer Qualitätsstandards für einen produktiven Einsatz zukünftig weiter verringert werden musste.²¹

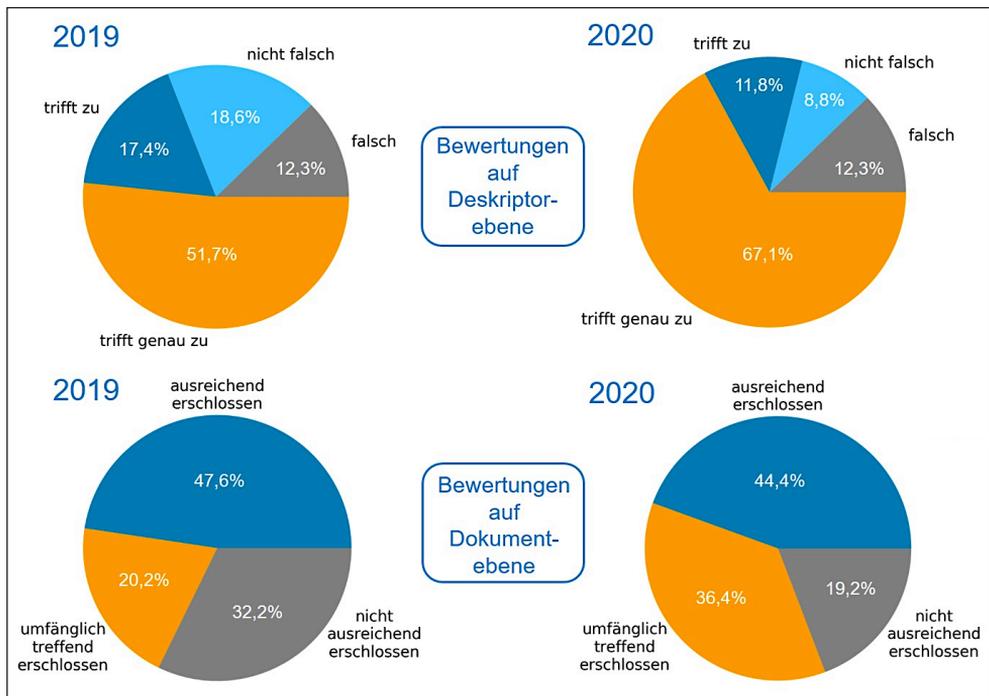


Abbildung 3: Ergebnisse der intellektuellen Reviews 2019 und 2020 im Vergleich

Außerdem waren die Reviews nützlich, um gewisse systematische Abweichungen vom gewünschten Ergebnis zu identifizieren. Zwei Beispiele: Die Bewertenden bemerkten, dass die Maschine die beiden Deskriptoren „USA“ und „Theorie“ besonders oft falsch vergab. Dies ist nicht ungewöhnlich, da diese beiden Deskriptoren mit 16% und 27% in den Trainingsdaten mit Abstand am stärksten vertreten waren, so dass die Maschine ebenfalls dazu tendierte, sie häufig zu vergeben. Eine höhere Vergabehäufigkeit bietet auch mehr Gelegenheiten dafür, „danebenzuliegen“. Als *ad-hoc*-Gegenmaßnahme blockierte das AutoSE-Team „USA“, außer entsprechende Zeichenketten („USA“, „US“, „United States of America“, etc.) kamen explizit im Textmaterial vor. Bei „Theorie“ wurde kritisiert, dass oft gleichzeitig ein engerer Begriff vorgeschlagen wurde und im Sinne der Erschließungsregeln dann nur dieser

21 Siehe Abb. 3: Der Anteil eindeutig falscher Deskriptoren war gleich geblieben und der Anteil nicht ausreichend erschlossener Dokumente hatte sich um 13 Prozentpunkte auf 19,2% verringert, was noch nicht zufriedenstellend war. Merke, dass „nicht ausreichend erschlossen“ auch bedeuten kann, dass zu viele Deskriptoren fehlen, dieser Anteil muss also nicht mit dem Anteil eindeutig falscher Deskriptoren korrelieren.

vergeben werden sollte. Deshalb ließ sich das Team von den Bewertenden eine Blacklist mit spezifischeren Theoriebegriffen aus den Wirtschaftswissenschaften geben, die fortan die Vergabe von „Theorie“ blocken sollten.²² Solche „handgebastelten“ Filterregeln sollten allerdings nicht überhandnehmen, da eine große Anzahl davon nicht gut zu pflegen ist. Auch sollte nach zukünftigen Trainingsvorgängen regelmäßig geprüft werden, ob die Maschine diese Regeln mittlerweile selbst gelernt hat.

Reviews wie die bisher durchgeführten bedeuten allerdings auch einen gewissen Aufwand seitens des wissenschaftlichen Dienstes: Die Bewertenden müssen dafür eine separate Plattform (das *release-tool*) nutzen und ihre zu bewertende Liste über ca. vier Wochen im Jahr zusätzlich zu ihren Routineaufgaben abarbeiten. Folglich kann bei diesem Vorgehen immer nur eine recht kleine Stichprobe bewertet werden – wünschenswert wäre natürlich eine größere und kontinuierlich gesammelte Basis von Bewertungsdaten.

Nach einem Austausch mit der DNB zu diesem Thema konkretisierte das Team diese Idee in der Spezifikation eines Bewertungstools für den DA-3 und trat damit an den Betreiber Eurospider heran, welcher sie umsetzte und im Frühjahr 2022 für das K10plus-Profil freischaltete, siehe Abb. 4. Nach ein paar Testphasen sind die Sacherschließenden der ZBW nun in ihrer täglichen Arbeit angehalten, nach der Erschließung einer Ressource im DA-3 jedes Mal auch die AutoSE-Vorschläge zu bewerten, wenn vorhanden, und zwar wie bei den separaten Reviews sowohl die einzelnen Deskriptoren als auch die AutoSE-Gesamtverschlagwortung des Dokuments. Da laut den Bewertenden der größte Aufwand bei den separaten Reviews darin bestand, die Ressource zumindest mental selbst zu erschließen und die Vorschläge dann damit abzugleichen, stellt die Bewertung nach dem Erschließungsvorgang über den DA-3 keinen großen Mehraufwand mehr dar. Die so entstehenden Daten greift das AutoSE-Team ab, wertet sie aus, visualisiert das aktuelle Qualitätslevel und dessen Verlauf in einem internen User-Interface und bezieht diese Erkenntnisse mit ein, wenn es darum geht, Teilziele für die zukünftige Methodenentwicklung abzustecken.

²² Bemerkenswert ist dabei, dass die Theoriebegriffe auf der Liste bisher keine Unterbegriffe von „Theorie“ sind – es wäre in Absprache mit der Thesaurusredaktion eine Überlegung wert, ob der STW entsprechend umstrukturiert werden kann, damit AutoSE die Hierarchiebeziehungen ausnutzen kann, statt sich einer Blacklist bedienen zu müssen.

DA-3-Profil: „k10plus“

Tools > Bewertung Einstellungen #

Bewertung abschicken 7/7

Gesamtbewertung

Quelle zbwase ++ + o - |x

STW

Arbeitsverhalten	zbwase	++	+	o	-	x
Arbeitszufriedenheit	zbwase	++	+	o	-	x
Mitarbeiterbindung	zbwase	++	+	o	-	x
Umweltbewusstsein	zbwase	++	+	o	-	x
Umweltmanagement	zbwase	++	+	o	-	x
Verhalten in Organisationen	zbwase	++	+	o	-	x

Abbildung 4: Bewertungstool im DA-3

3.2 Maschinelle Qualitätssicherung

Die im vorigen Abschnitt beschriebene intellektuelle Qualitätskontrolle wird seit Mitte 2022 im Produktivbetrieb durch eine machine-learning-basierte Qualitätskontrolle namens *qualle* ergänzt. Die Idee zum Ansatz und erster prototypischer Code dazu stammt aus dem Vorläuferprojekt Auto-Index; für den produktiven Einsatz musste dieser Code allerdings in Python reimplementiert werden.²³ Bei *qualle* handelt es sich um eine aus Trainingsdaten erlernte Abschätzung der zu erwartenden Qualität der maschinell generierten Gesamtverschlagwortung für eine Ressource. Genauer: *qualle* schätzt den voraussichtlichen Wert einer Qualitätsmetrik – für AutoSE aktuell den *recall*-Wert. In die Abschätzung werden die Konfidenzwerte für die einzelnen, bereits generierten Deskriptoren einbezogen sowie diverse Merkmale wie etwa die Länge des Textstücks und wie viele Sonderzeichen es enthält,²⁴ und als zentrale Heuristik auch einen Vergleich der vorhergesagten Anzahl von Deskriptoren mit der tatsächlichen Anzahl der vorgeschlagenen Deskriptoren.²⁵

Die mit *qualle* gefilterten Ergebnisse (d.h. Verschlagwortungen, für den *qualle* einen bestimmten Mindest-*recall* vorhersagt) wurden im Frühjahr vor der Produktivnahme einem intellektuellen Review unterzogen und für mindestens gleich gut befunden wie der vorige Output. Der *qualle*-Filter stellte insbesondere insofern eine Verbesserung gegenüber der bisher verwendeten, intellektuell postulierten Filterregel „min2VB“ dar, als er feingranularer wirkt. Laut der Regel „min2VB“ musste die Gesamtverschlagwortung für eine Ressource mindestens zwei Deskriptoren aus den beiden STW-Subthesauri für die wirtschaftswissenschaftlichen Kernbereiche BWL und VWL

23 *qualle*, <<https://github.com/zbw/qualle>>, Stand: 30.09.2022.

24 Z.B. Fragezeichen, was eine Frage statt einer Aussage markiert und damit auf ein für die Maschine schwieriger zu verschlagwortendes Textfragment hinweisen könnte.

25 Für mehr Details siehe Toepfer, Martin; Seifert, Christin: Content-Based Quality Estimation for Automatic Subject Indexing of Short Texts Under Precision and Recall Constraints, in: Méndez, Eva, Crestani, Fabio, Ribeiro, Cristina u.a. (Hg.): Digital Libraries for Open Knowledge. TPD 2018, Cham, 2018 (LNCS 11057). Online: <https://doi.org/10.1007/978-3-030-00066-0_1>.

enthalten, sonst wurde sie verworfen. Diese relativ grobe Heuristik vernachlässigt also andere mit den Wirtschaftswissenschaften assoziierte Bereiche, während *qualle* einfach aus den Trainingsdaten lernt, was eine gute Verschlagwortung ist, ohne nach Subthesauri zu unterscheiden. Der Produktivbetrieb enthält nun also eine Qualitätssicherungsschicht, die darin besteht, dass für alle Deskriptoren einzeln eine Konfidenzwertschwelle gilt und auf die verbleibende Gesamtverschlagwortung anschließend noch der *qualle*-Filter angewendet wird. Während diese Maßnahmen auf einer rein technischen Berechnung von Metriken beruhen, bietet sich auch hier ein Anknüpfungspunkt für eine Arbeitsteilung im Sinne des *human in the loop*: Zukünftig könnten Datensätze für Ressourcen, die den *qualle*-Filter nicht passieren und damit von einer Maschine voraussichtlich nur schlecht zu erschließen sind, direkt an menschliche Erschließende weitergeleitet werden.

4. Fazit und Ausblick

Aus dem in der Pilotphase Erlebten lassen sich insbesondere die beiden folgenden *lessons learned* formulieren: Zum einen war es ein zentraler Schritt, die Automatisierung der Inhaberschließung aus dem Projektstatus zu lösen und zur Daueraufgabe zu erklären. Das hat für die Verbindlichkeit und die Ressourcen gesorgt, die notwendig waren, um die Ergebnisse aus der angewandten Forschung über den Prototypen hinaus in einen produktiven zu Dienst überführen und für diesen ein tragbares Betriebsmodell auszuarbeiten. Hierfür sind insbesondere unterschiedliche Expertisen und damit eine entsprechende Ausdifferenzierung der Personalressourcen und Rollen grundlegend: angewandte Forschung, Softwareentwicklung, IT-Administration und fachliche Leitung. Den Erfahrungen in AutoSE zufolge sind drei Personen die absolute Minimalbesetzung für ein solches Team; die gute Praxis empfiehlt möglichst eine Dopplung von einzelnen Kompetenzen für mehr Ausfallsicherheit.

Zum anderen ist es absolut essenziell, Automatisierungsmaßnahmen dieser Art mit den erschließenden Personen der Einrichtung abzustimmen. Sie sind sowohl Partner als auch Stakeholder in einem solchen Unterfangen, und damit sollte eine solche Abstimmung eine Selbstverständlichkeit sein. Außerdem lohnt es sich, bei ihnen für Transparenz bezüglich der eingesetzten Methoden zu sorgen: Regelmäßige Aktivitäten zur Information der Erschließenden erhöhen nicht nur die Akzeptanz und verbessern die Zusammenarbeit in der aktuellen Phase, sondern bereiten auch den Boden für zukünftige disruptivere Umorganisationen des Erschließungsworkflows auf dem Weg zu einer tiefgreifenderen Verzahnung von Mensch und Maschine im Sinne des *human in the loop* (siehe Abschnitt 3).

Im verbleibenden Jahr der Pilotphase soll für AutoSE noch ein Webinterface fertiggestellt werden, das es den Stakeholdern erlaubt, über ein Demo mit dem produktiven Backend des Dienstes zu interagieren und verschiedene Statistiken zur Performanz und Informationen zu den verwendeten Methoden abzurufen. Letztere werden ständig weiterentwickelt und ausgeweitet, beispielsweise in Richtung einer Nutzung von Abstracts und Inhaltsverzeichnissen oder der Erschließung von Ressourcen in weiteren Sprachen. Schlussendlich sollen die Erfordernisse eines nachhaltigen Betriebsmodells dokumentiert werden, um den langfristigen Ressourcenbedarf aufzeigen und abdecken zu können. Dies ist wichtig auch im Hinblick auf eine perspektivische Erweiterung zur Extraktion und Aufbereitung von Textmaterial aus Volltexten, um weitere Bereiche der Inhalts- und auch Formalserschließung zu automatisieren und so auch den Input für AutoSE zu optimieren.

AutoSE ist ein Praxisbeispiel für den erfolgreichen Aufbau eines produktiven Dienstes. Aus den an der ZBW gemachten Erfahrungen lässt sich aber auch ableiten, welche Bedingungen mindestens gegeben sein müssen, um Machine-Learning-Lösungen an einer Institution für die Inhaltserschließung einsetzen zu können, und nicht jede Institution kann diese Bedingungen herstellen – oder muss ihre Prioritäten entsprechend verschieben. Wenn eine fortschreitende Automatisierung in der Bibliothekswelt tatsächlich in Teilen mit anspruchsvolleren Methoden aus der Künstlichen Intelligenz erfolgen soll, besteht sowohl an einzelnen Institutionen, aber auch institutionsübergreifend in Bezug auf strategische Allianzen und die gemeinsame Nutzung von Ressourcen noch ein deutlicher Handlungsbedarf.²⁶

Literaturverzeichnis

- Bartz, Christopher: Software Architecture for the Automatization of Subject Indexing. Vortrag bei der ELAG am 08.06.2022 in Riga, Litauen. Online: <<https://elag2022.lnb.lv/programme/schedule/>>, Stand: 30.09.2022.
- Beckmann, Regine; Hinrichs, Imma; Janßen, Melanie u.a.: Der Digitale Assistent DA-3 – eine Plattform für die Inhaltserschließung, in: o-bib – das offene Bibliotheksjournal 6 (3), 2019, S. 1–20. Online: <<https://doi.org/10.5282/o-bib/2019H3S1-20>>.
- Busse, Frank; Grote, Claudia; Jacobs, Jan-Helge u.a.: Erschließungsmaschine gestartet, 04.05.2022, <<https://blog.dnb.de/erschliessungsmaschine-gestartet/>>, Stand: 30.09.2022.
- Monarch, Robert M.; Manning, Christopher D.: Human-in-the-loop machine learning – active learning and annotation for human-centered AI. (E-Book), Manning Publications, 2021. Online: <<https://livebook.manning.com/book/human-in-the-loop-machine-learning/>>, Stand: 30.09.2022.
- Kasprzik, Anna: Get everybody on board and get going – the automation of subject indexing at ZBW [Artikel], in: 87th IFLA World Library and Information Congress (WLIC), Satellite Meeting: Information Technology – New Horizons in Artificial Intelligence in Libraries, 2022. Online: <<https://repository.ifla.org/handle/123456789/2047>>.
- Kasprzik, Anna: Get everybody on board and get going – the automation of subject indexing at ZBW [Folien]. Vortrag beim 87th IFLA World Library and Information Congress (WLIC), Satellite Meeting: Information Technology – New Horizons in Artificial Intelligence

26 Für eine weitere Diskussion zum Thema siehe auch die folgenden Quellen:
Seeliger, Frank; Puppe, Frank; Ewerth, Ralph u.a.: Zum erfolgversprechenden Einsatz von KI in Bibliotheken – Diskussionsstand eines White Papers in progress, in: b.i.t.online 24 (2 und 3), 2022, S. 173–178 (Teil 1) und S. 290–299 (Teil 2). Online: <<http://hdl.handle.net/11108/488>> und <<http://hdl.handle.net/11108/490>>. ZBW Mediatalk: KI in wissenschaftlichen Bibliotheken, Teil 1: Handlungsfelder, große Player und die Automatisierung der Erschließung, 17.08.2022, <<https://www.zbw-mediatalk.eu/de/2022/08/ki-in-wissenschaftlichen-bibliotheken-teil-1-handlungsfelder-grosse-player-und-die-automatisierung-der-erschliessung/>>, Stand: 30.09.2022. ZBW Mediatalk: KI in wissenschaftlichen Bibliotheken, Teil 3: Voraussetzungen und Bedingungen für den erfolgreichen Einsatz, 31.08.2022, <<https://www.zbw-mediatalk.eu/de/2022/08/ki-in-wissenschaftlichen-bibliotheken-teil-3-voraussetzungen-und-bedingungen-fuer-den-erfolgreichen-einsatz/>>, Stand: 30.09.2022.

in Libraries am 22. Juli 2022 in Galway, Irland. Online: <<https://repository.ifla.org/handle/123456789/2047>>.

- Seeliger, Frank; Puppe, Frank; Ewerth, Ralph u.a.: Zum erfolgversprechenden Einsatz von KI in Bibliotheken – Diskussionsstand eines White Papers in progress, in: b.i.t.online 24 (2 und 3), 2022, S. 173–178 (Teil 1) und S. 290–299 (Teil 2). Online: <<http://hdl.handle.net/11108/488>> und <<http://hdl.handle.net/11108/490>>.
- Toepfer, Martin; Seifert, Christin: Fusion architectures for automatic subject indexing under concept drift, in: International Journal on Digital Libraries 21, 2018, S. 169–189. Online: <<https://doi.org/10.1007/s00799-018-0240-3>>.
- Toepfer, Martin; Seifert, Christin: Content-Based Quality Estimation for Automatic Subject Indexing of Short Texts Under Precision and Recall Constraints, in: Méndez, Eva; Crestani, Fabio; Ribeiro, Cristina u.a. (Hg.): Digital Libraries for Open Knowledge. TPDL 2018, Cham, 2018 (LNCS 11057). Online: <https://doi.org/10.1007/978-3-030-00066-0_1>.
- Tochtermann, Klaus; Kasprzik, Anna: Auf Augenhöhe mit Forschungspartnern aus der Wissenschaft – Anwendung von Künstlicher Intelligenz in der ZBW, in: BuB – Forum Bibliothek und Information 74 (6), 2022, S. 306–311. Online: <<https://pub.zbw.eu/dspace/bitstream/11108/526/2/2022-Kasprzik-Tochtermann-Augenh%C3%B6he.pdf>>.
- Winkler, Christian: Wer, wie, was. Textanalyse über Natural Language Processing mit BERT, heise online, 12.08.2020, <<https://www.heise.de/hintergrund/Wer-wie-was-Textanalyse-mit-BERT-4864558.html>>, Stand: 30.09.2022.
- ZBW Mediatalk: KI in wissenschaftlichen Bibliotheken, Teil 1: Handlungsfelder, große Player und die Automatisierung der Erschließung, 17.08.2022, <<https://www.zbw-mediatalk.eu/de/2022/08/ki-in-wissenschaftlichen-bibliotheken-teil-1-handlungsfelder-grosse-player-und-die-automatisierung-der-erschliessung/>>, Stand: 30.09.2022.
- ZBW Mediatalk: KI in wissenschaftlichen Bibliotheken, Teil 3: Voraussetzungen und Bedingungen für den erfolgreichen Einsatz, 31.08.2022, <<https://www.zbw-mediatalk.eu/de/2022/08/ki-in-wissenschaftlichen-bibliotheken-teil-3-voraussetzungen-und-bedingungen-fuer-den-erfolgreichen-einsatz/>>, Stand: 30.09.2022.