Anna Kasprzik, Klaus Tochtermann

Auf Augenhöhe mit Forschungspartnern aus der Wissenschaft

Anwendung von Künstlicher Intelligenz in der ZBW

Bereits im Jahr 2017 wurde von 2,5 Millionen neuen Publikationen pro Jahr ausgegangen, was zum Teil auf den jährlichen Zuwachs von vier bis fünf Prozent publizierenden Forschenden zurückzuführen ist[1]. In dem für die ZBW – Leibniz-Informationszentrum Wirtschaft relevanten Fachgebiet der Wirtschaftswissenschaften hat sich die Zahl der Publikationen im Web of Science allein in den Jahren 2000 bis 2017 verdoppelt. Diese von Jahr zu Jahr steigende Menge an verfügbaren Publikationen führt dazu, dass es für Bibliotheken nahezu unmöglich ist, jede einzelne Publikation von Hand mit sorgfältig kuratierten Metadaten zu versehen, was jedoch für eine Bereitstellung zum Zwecke einer sinnvollen Nachnutzung notwendig ist. Daher erfährt die Automatisierung auch durch die Anwendung von Methoden aus der Künstlichen Intelligenz (KI) eine zunehmend bedeutende Rolle in Bibliotheken. Das Verständnis von KI orientiert sich im Rahmen dieses Beitrags an der Definition aus dem Gabler Wirtschaftslexikon: *Künstliche Intelligenz (KI) beschäftigt sich mit Methoden, die es einem Computer ermöglichen, solche Aufgaben zu lösen, die, wenn sie vom Menschen gelöst werden, Intelligenz erfordern*[2].

Zu solchen Aufgaben gehören in der ZBW unter anderen die Erschließung von Publikationen, die Suche und das Auffinden von Publikationen sowie deren bedarfsorientierte Strukturierung und die Analyse von Zusammenhängen zwischen unterschiedlichen Publikationen. In diesem Kontext greift die ZBW neueste Erkenntnisse, Verfahren und Werkzeuge der KI auf, prüft sie auf deren Praxistauglichkeit und übersetzt sie in eigene Anwendungen beziehungsweise nutzt sie für deren Weiterentwicklung.

Dieser Beitrag ist wie folgt aufgebaut: Zunächst wird in Kapitel 1 ein Überblick über unterschiedliche Anwendungsbereiche gegeben, in denen in der ZBW KI-Methoden eingesetzt werden. Kapitel 2 eröffnet anhand von zwei Beispielen tiefere Einblicke in die Anwendung und den Nutzen von KI-Methoden in den Themenfeldern inhaltliche Sacherschließung und Analyse von Wissenschaftsdynamiken. Kapitel 3 schließt diesen Beitrag mit einer kurzen Conclusio ab.

1. Künstliche Intelligenz in der ZBW – ein Überblick

Die Anwendung von KI-Methoden hat in der ZBW eine lange Historie, auch wenn erst seit wenigen Jahren tatsächlich nutzenstiftende Ergebnisse erzielt und in den Arbeitsalltag in der ZBW überführt wurden. Während man in frühen Jahren (circa 2010) primär auf externe KI-Expertise zurückgegriffen und im Wesentlichen von Seiten der ZBW inhaltliche Domänenkompetenz beigetragen hatte, wurde in der jüngsten Vergangenheit und mit dem Ausbau der Forschung in der ZBW eigene KI-Expertise aufgebaut. Unterschiedliche Anwendungsfelder werden nachfolgend kurz vorgestellt, bevor im anschließenden Kapitel zwei dieser Felder näher beleuchtet werden.

Gegenwärtig untersucht die ZBW die Verwendung künstlicher neuronaler Netze, das heißt eine Simulation natürlicher neuronaler Netze, für Bausteine der Literatursuche. Gegenstand der Untersuchungen sind insbesondere die Einsatzmöglichkeiten und die beeinflussenden Faktoren (zum Beispiel Struktur des Netzwerks von Personen, Publikationen, Institutionen et cetera oder Titel von Publikationen) auf die entwickelten Modelle. Die bisherigen Entwicklungen für Literatur-Empfehlungssysteme können beispielsweise auf Basis der zitierten Werke auf möglicherweise fehlende Zitationen hinweisen oder auf Basis bereits erschlossener Deskriptoren weitere Deskriptoren empfehlen.

In einem weiteren Themenfeld setzt die ZBW Methoden des maschinellen Lernens – einem Untergebiet der Künstlichen Intelligenz – ein, um Wortvektoren zur verbesserten Beantwortung von Anfragen in Suchmaschinen zu generieren. Wortvektoren sind Zahlenvektoren, die die Bedeutung eines Wortes darstellen. Sie können mithilfe von maschinellem Lernen aus großen Datenmengen von unstrukturiertem Text, wie



Die Automatisierung erfährt durch die Anwendung von Methoden aus der Künstlichen Intelligenz (KI) eine zunehmend bedeutende Rolle – auch in Bibliotheken. Foto: greenbutterfly – stock.adobe.com

etwa Publikationen, erzeugt werden. Dies ermöglicht den Vergleich der Ähnlichkeit zweier Worte anhand eines geometrischen Vergleichs ihrer Wortvektoren. So lassen sich durch das Messen von Abstand beziehungsweise des Winkels zwischen zwei Wortvektoren semantische Ähnlichkeiten und Zusammenhänge zwischen den Worten ableiten, die durch die Wortvektoren repräsentiert werden. Mit der Erweiterung zu sogenannten Wortmatrizen kann zudem die Reihenfolge der Worte berücksichtigt werden, welche bei Wortvektoren meist vernachlässigt wird.

Neben der Literatursuche innerhalb eines Repositories wurden entsprechende Verfahren erfolgreich eingesetzt, um Suchanfragen über mehrere verschiedene Repositorien auszuweiten. Beispielsweise können somit ausgehend von aufgefundenen Publikationen aus einem Repository für Wirtschaftswissenschaften thematisch verwandte Publikationen aus einem weiteren Repository für Sozialwissenschaften aufgefunden werden (Hajra & Tochtermann 2017).

Die inhaltliche Sacherschließung, also das Annotieren von Publikationen mit semantischen Angaben, ist eine Kernaufgabe der ZBW, um das Auffinden relevanter Literatur im Bestand zu erleichtern. Aufgrund der digitalen Publikationsflut ist es kaum mehr möglich, alle Publikationen intellektuell zu annotieren, sodass über Automatisierungsstrategien nachgedacht werden muss. Aus Sicht des maschinellen Lernens ist die Sacherschließung eine sogenannte Multi-Label-Classification-Herausforderung, das heißt jeder Publikation können mehrere Labels beziehungsweise in diesem Kontext inhaltliche Schlagwörter zugewiesen werden. Mit dem Aufschwung der KI in den letzten Jahren stehen inzwischen immer mehr Methoden, auch als Open-Source-Software, zur Lösung dieser Aufgabe bereit.

Die bestehenden Metadatenbestände an der ZBW eignen sich schließlich auch hervorragend für die Anwendung von maschinellen Lernverfahren, um Repräsentationen in Netzwerken aus bibliografischen Metadaten abzuleiten. Das Ziel ist es, eine Repräsentation für die Konzepte eines kontrollierten Vokabulars zu erlernen. Als Datengrundlage wird neben den Metadaten von Publikationen, auf die Netzwerkstruktur zwischen Autorinnen und Autoren, Publikationen und Institutionen zurückgegriffen. Ein besonderer Anwendungsschwerpunkt liegt auf der Analyse von Wissenschaftsdynamiken, das heißt der Konvergenz beziehungsweise Diffusion von unterschiedlichen wissenschaftlichen Themenfeldern.

Im nachfolgenden Kapitel 2 wird mit der Automatisierung der Sacherschließung die Anwendung von KI-Methoden in einem traditionellen Aufgabenfeld von Bibliotheken vertieft. Zudem wird in einem zweiten Anwendungsbereich beschrieben, wie die KI-basierte Analyse von bibliografischen Daten für wissenschaftliche Fragestellungen, im konkreten der Analyse von Wissenschaftsdynamiken, verwendet werden können.

2. Künstliche Intelligenz in der ZBW – Zwei Anwendungsfelder

2.1 Automatisierung der Sacherschließung (AutoSE)

Die ZBW hatte bereits in verschiedenen Vorläuferprojekten seit der Jahrtausendwende Möglichkeiten zur Automatisierung der Sacherschließung ausgelotet, anfangs noch mit externen Partnern und mit kommerziellen Werkzeugen. Seit circa 2016 wird an der ZBW in diesem Themenfeld eigene angewandte

Forschung mit dem Ziel betrieben, praktikable Lösungen für eine automatisierte oder maschinengestützte Inhaltserschließung zu entwickeln. Hierfür werden Methoden des maschinellen Lernens herangezogen. Die bis Ende 2018 entwickelten Forschungsprototypen waren allerdings nicht für den direkten Einsatz in der Praxis geeignet, sondern mussten noch in produktive Dienste für den durchgängigen Betrieb überführt und mit den anderen Abläufen zur Metadatenverarbeitung in der ZBW sinnvoll verzahnt werden.

Daher wurden 2020 eine Pilotphase für den Aufbau einer geeigneten Architektur zur Integration solcher Dienste ausgerufen und zusätzliche Ressourcen für die notwendige Softwareentwicklung bereitgestellt. Sowohl die angewandte Forschung als auch die Architekturentwicklung sind direkt bei den wissenschaftlichen Diensten der ZBW angesiedelt und profitieren von einem engen Austausch mit den Verantwortlichen für die intellektuelle Erschließung. Die Pilotphase soll 2024 abgeschlossen werden, die Automatisierung der Sacherschließung (»AutoSE«) ist jedoch an der ZBW aus dem Projektstatus heraus und in den Status einer Daueraufgabe überführt worden, und die Lösungen dafür sollen auf der Basis der eigenen angewandten Forschung kontinuierlich weiterentwickelt werden.

Im Vorläuferprojekt (AutoIndex, bis 2018) wurde ein Prototyp für einen Fusion-Ansatz entwickelt, der mehrere Methoden kombinierte und das Ergebnis noch einmal anhand von Regeln filterte (Toepfer & Seifert 2018-1). Parallel dazu entstand an der finnischen Nationalbibliothek das Open-Source-Toolkit Annif, das verschiedene Machine-Learning-Algorithmen zur Automatisierung der Inhaltserschließung anbietet, es aber auch erlaubt, eigene Modelle zu integrieren.

Zu Beginn der Pilotphase übernahm das AutoSE-Team Annif als eine Art Steckrahmen für verschiedene *State-of-the-Art*-Modelle – einschließlich einer für den an der ZBW gepflegten Standard-Thesaurus Wirtschaft (STW)[3] optimierten Eigenentwicklung[4] – und flankierte diesen mit Mechanismen für wissenschaftliches Experimentieren, Parameteroptimierung, Qualitätskontrolle und den Anschluss an die Erschließungsworkflows im Haus. Parallel dazu wird mit Ansätzen aus dem Deep Learning experimentiert, insbesondere mit Transformermodellen, welche besonders vielversprechend sind für eine multilinguale Erschließung. Inzwischen arbeitet die ZBW an der Weiterentwicklung von Annif mit, prüft regelmäßig, ob sich Erkenntnisse aus dem AutoSE-Kontext als neue Funktionalitäten in Annif integrieren lassen, veranstaltet zusammen mit der finnischen Nationalbibliothek Tutorials und berät zum Einsatz von Annif.

Eine erste Version des AutoSE-Dienstes ging 2021 produktiv. Der Output wird bisher an zwei Stellen nachgenutzt: Der Dienst prüft stündlich die Datenbasis des ZBW-Rechercheportals EconBiz[5] auf neue einschlägige Metadatensätze, verschlagwortet diese und schreibt sie zurück. Bisher sind die Modelle ausgerichtet auf englische Texte, und genutzt werden bei der Verarbeitung aktuell nur der Titel und, wenn vorhanden, von Autorinnen und Autoren vergebene Keywords – die Nutzung von Abstracts ist in Prüfung. Allein im ersten Halbjahr nach Produktivgang wurden auf diese Weise über 100 000 Ressourcen automatisiert verschlagwortet.

Außerdem wird der Output über eine Schnittstelle dem Digitalen Assistenten (DA-3)[6] als Vorschläge zur Verfügung gestellt, also der Plattform, die an der ZBW zur Unterstützung der intellektuellen Sacherschließung genutzt wird. Sobald eine Publikation intellektuell verschlagwortet und das Ergebnis über den DA-3 in den Verbundkatalog und von dort wieder in die EconBiz-Datenbasis überführt worden ist, greift das AutoSE-Team diese Verschlagwortung ab und vergleicht sie mit dem eigenen automatisierten Erzeugnis, um fortlaufend Feedback über die Qualität des AutoSE-Outputs zu sammeln.

Ein zentrales Thema ist die Qualitätskontrolle. Hierfür wird an einem umfassenden Konzept mit verschiedenen Ansätzen gearbeitet, um ein möglichst hohes Gesamtniveau garantieren zu können. Auf der technischen Seite wird dazu mit diversen im maschinellen Lernen üblichen Metriken gearbeitet und es werden entsprechende Schwellwerte ermittelt. Dem Verschlagwortungsvorgang werden diverse Filter nachgeschaltet – seit 2022 wird auch eine maschinell gelernte Abschätzung der zu erwartenden Qualität der Gesamtverschlagwortung pro Metadatensatz produktiv angewendet. Das Verfahren namens *qualle* stützt sich auf die Konfidenzwerte für die einzelnen Schlagwörter und weitere Heuristiken (zum Beispiel Textlänge). *qualle* basiert auf einem in (Toepfer & Seifert 2018-2) beschriebenen Prototypen – für den produktiven Einsatz musste der Code allerdings komplett neu geschrieben werden.[7]

Eines der wichtigsten Elemente in der Qualitätssicherung bleibt jedoch die Zusammenarbeit zwischen Maschine und Mensch – hier spricht man vom *human in the loop*.[8] Mögliche Umsetzungen reichen von der simplen Tatsache, dass Trainingsdaten von Menschen annotiert sind (das trifft auch für AutoSE zu) und dass Wissensorganisationssysteme und Konkordanzen meist noch von Menschen erstellt und gepflegt werden (wie der STW und dessen Mappings auf andere Vokabulare), über maschinenunterstützte Erschließung (wie das Anbieten von AutoSE-Vorschlägen im DA-3) und diverse Arten des intellektuellen Feedbacks bis hin zu Ansätzen wie Online-Learning, wo sich die Maschine (zum Beispiel auf der Basis solchen Feedbacks) umgehend selbst neu trainiert, und Active Learning, wo die Maschine Menschen punktuell und interaktiv um die Annotation oder Bewertung eines Datenpunkts ersuchen kann.

---

1 http://blog.cdnsciencepub.com/21st-century-science-overload/

2 https://wirtschaftslexikon.gabler.de/definition/kuenstliche-intelligenz-ki-40285

3 https://zbw.eu/stw/version/latest/about.de.html

4 https://github.com/zbw/stwfsapy

5 https://www.econbiz.de/

6 https://www.eurospider.com/de/relevancy-produkt/digitaler-assistent-da-3

7 https://github.com/zbw/qualle

8 Für eine Einführung siehe Monarch & Manning (2021)

Über den Abgleich der AutoSE-Vorschläge mit der intellektuellen Sacherschließung hinaus erfolgten qualitative Prüfungen bisher circa einmal jährlich in intellektuellen Reviews, bei denen eine Gruppe von Sacherschließungsexpertinnen und -experten an der ZBW für eine Stichprobe von rund 1 000 Dokumenten Bewertungen abgeben – auf der Ebene einzelner Deskriptoren und auf der Ebene der Gesamtverschlagwortung. Hierfür wurde eine in Projekt AutoIndex entwickelte Oberfläche verwendet, über die die Bewertenden die relevanten Metadaten und über einen Link den Volltext einsehen, in den Datensätzen navigieren und im Zuge einer Bewertung auch fehlende Deskriptoren ergänzen konnten.

Während der Abgleich mit der intellektuellen Sacherschließung im Alltagsbetrieb nur ein binäres Feedback zurückliefert (»stimmt überein«/»stimmt nicht überein«) und der Aufwand eines Reviews auf einer getrennten Plattform die Durchführung nur in größeren Abständen zulässt, fehlte noch eine Möglichkeit, im Erschließungsbetrieb fortlaufend gestuftes Feedback einzuholen. Im Frühjahr integrierte die DA-3-Anbieterin Eurospider ein im AutoSE-Kontext konzipiertes Bewertungstool. Seitdem können Sacherschließende der ZBW im DA-3 bei jeder Ressource, die sie erschließen und für die AutoSE Schlagwörter bereitstellt, zu diesen Vorschlägen eine Bewertung abgeben – und sind auch dazu angehalten, um für AutoSE einen größtmöglichen Datenbestand an gestuftem Feedback aufzubauen. So werden Erkenntnisse gewonnen, wie der AutoSE-Output aktuell angenommen wird, bekommt gegebenenfalls Hinweise auf systematische Fehler und kann damit die eingesetzten Methoden weiter verbessern.

Erfahrungen aus der bisherigen Pilotphase haben Folgendes gezeigt: Regalfertige maschinelle Erschließungssysteme gibt es (aktuell noch) nicht – verfügbare Open-Source-Lösungen müssen mit verschiedenen Expertisen begleitet und angepasst werden. Zur Sicherung der Qualität und zur Förderung der Akzeptanz empfiehlt sich eine möglichst enge Zusammenarbeit mit dem Anwendungsbereich und das Einbeziehen von Erschließungsexpertinnen und -experten. Und schließlich: Das Projektformat hinter sich zu lassen, hat sich gelohnt – die Automatisierung der Inhaltserschließung ist eine Daueraufgabe, die Bibliotheken noch viele Jahre begleiten wird. Ein darauf ausgerichteter Produktivbetrieb braucht ein langfristig angelegtes und sorgfältig ausgearbeitetes Betriebsmodell und muss mit entsprechenden Ressourcen (Personal, Soft- und Hardware) unterfüttert werden.

ANZEIGE



2.2 Analyse von Wissenschaftsdynamiken

Bibliografische Daten von Bibliotheken eignen sich hervorragend für die Anwendung und Weiterentwicklung maschineller Lernverfahren. Von besonderer Bedeutung sind Daten der inhaltlichen Sacherschließung, vor allen Dingen dann, wenn diese auf einem kontrollierten Vokabular und im besten Fall einem Thesaurus basieren. In diesen Kontext lässt sich das Projekt Q-Aktiv einordnen, das von der ZBW gemeinsam mit der ZB MED – Informationszentrum Lebenswissenschaften und dem Lehrstuhl für Technologiemanagement der Universität Kiel durchgeführt wurde. Das Ziel des Projekts war es, die Methoden zur Prognose von Dynamiken und Wechselwirkungen zwischen Forschung, Technologieentwicklung und Innovation zu verbessern. Der Fokus des Vorhabens lag unter anderem auf der Entwicklung eines Instruments zur Beobachtung der Entstehung neuer wissenschaftlicher Themenfelder sowie auf der Analyse von Konvergenzprozessen innerhalb und zwischen wissenschaftlichen Disziplinen.

Eine besondere Herausforderung bestand in der Berücksichtigung von inhaltlichen Veränderungen im Literatur-Korpus über die Zeit. Denn nur so konnten Konvergenzprozesse erkannt und analysiert werden. Für diesen Zweck wurden große Mengen (> 10 Millionen) bibliografischer Daten genutzt und entsprechend der obigen Fragestellung ausgewertet. Aufgrund dieser großen Datenmenge sind maschinelle Lerntechniken erforderlich, einschließlich Text-Mining-Methoden, die sich für das Erforschen der Entwicklung von Technologiefeldern als erfolgreich erwiesen haben und meist auf der Analyse von Wortvorkommen, Zitaten, Zusammenfassungen et cetera basieren. Die Untersuchung der wissenschaftlichen Konvergenz als Konzept des Verschiebens und Zusammenführens von zuvor getrennten Themen- und Forschungsfeldern erfordert zunächst die Identifizierung und Unterscheidung einzelner Forschungsfelder. Eine solche Analyse erfordert den Einsatz von Techniken, wie sie häufig für das Clustering von Patenten und wissenschaftlichen Veröffentlichungen genutzt werden (Niemann et al. 2017).

Im Rahmen des Projekts wurde das entwickelte Instrument zur Analyse von Forschungsdynamiken in drei Fallbeispielen untersucht: Cholesterol, Algenforschung, COVID-19. Die ZBW übernahm dabei die Anwendung und Weiterentwicklung von KI-Methoden, ZB MED stellte ihre Daten bereit und bereitete diese auf beziehungsweise reicherte sie an und die Universität Kiel führte die Analysen mit dem entwickelten Instrument durch und bewertete die Ergebnisse.

Exemplarisch wird kurz auf Daten und Methoden der Fallstudie zu COVID-19 (Galke et al. 2021) eingegangen. Sie beschäftigte sich mit wissenschaftlichen Artikeln sowie Preprints des Jahres 2020 und untersuchte die interdisziplinären Anteile. Als Datenbasis wurden Daten aus dem ZB MED Knowledge Environment und Pre-Prints aus der ZB MED Projektdatenbank COVID-19-Hub genutzt, die mit Informationen aus den Datenbanken ORCID, Wikidata und Crossref (referenzierte Literatur und ISSN) angereichert und aufbereitet wurden. Methodisch wurden von der ZBW Begriffe/Konzepte der inhaltlichen Sacherschließung verwendet, mit denen die einzelnen Publikationen der digitalen Literatur-Korpora annotiert sind, um darauf aufbauend Ähnlichkeiten zwischen den Konzepten zu errechnen. Mit Techniken des maschinellen Lernens (im Speziellen des Deep Learning) lassen sich so niedrig-dimensionierte kontinuierliche Vektoren erstellen, die die einzelnen Konzepte zur Beschreibung einer Publikation darstellen. Zwischen diesen Konzeptrepräsentationen kann schließlich ein Wert für das Ähnlichkeitsmaß berechnet werden. Die Ähnlichkeit zwischen zwei Konzepten wird dabei anhand der Kosinusähnlichkeit der Repräsentationen (Vektoren von reellen Zahlen) gemessen.

Die im Projekt durchgeführten Analysen mit dem entwickelten Ähnlichkeitsindikator, die durchgeführten qualitativen Validitätstests sowie quantitativen Tests ergaben, dass der Indikator für die Messung von wissenschaftlichen Dynamiken, inklusive Konvergenz- und Divergenzprozessen gut geeignet ist. Dieses Maß liefert kontextuell wertvolle Einblicke in die wissenschaftliche Dynamik in intra- und interdisziplinären Forschungsbereichen und kann für die Identifizierung der wissenschaftlichen Konvergenz in Forschungsfeldern eingesetzt werden. Außerdem lassen sich durch die Anwendung des entwickelten Instruments auf wissenschaftliche Publikationsdaten zukunftsrelevante Themen in Wissensbereichen frühzeitig erkennen.

Die Bewertung der Ergebnisse lassen sich wie folgt kurz zusammenfassen (Melnychuk et al. 2021): Mittels des in Q-Aktiv entwickelten Instruments und den zugrundeliegenden bibliografischen Daten der beteiligten Bibliotheken lassen sich Themen identifizieren, die inhaltlich miteinander verbunden sind beziehungsweise sich »aufeinander zubewegen«, obwohl sie möglicherweise zu unterschiedlichen Themenfeldern oder gar Disziplinen gehören. Auf dieser Basis können zum Beispiel Entscheidungen über die Organisation von großen Forschungsorganisationen getroffen werden, sei es mit dem Ziel, eine Homogenität einzelner Einrichtungen zu erreichen oder im Interesse der interdisziplinären Forschung die Zusammenarbeit entfernter Wissensgebiete zu fördern. Außerdem ermöglicht das Instrument eine Verfolgung wissenschaftlicher Dynamiken und damit die Erkennung zukunftsrelevanter Themen wissenschaftlicher Felder.

3. Conlusio

Künstliche Intelligenz spielt in Bibliotheken eine zunehmend wichtige Rolle. Dabei lassen sich zwei große Anwendungsbereiche von KI identifizieren. Zum einen können, wie in Abschnitt 2.1 dargestellt, KI-Verfahren eingesetzt werden, um klassische Aufgaben von Bibliotheken, wie die Erschließung von Publikationen, zu unterstützen. Dies ist insbesondere vor dem Hintergrund des eingangs beschriebenen schnellen jährlichen Zuwachses an Publikationen von entscheidender Bedeutung, damit Bibliotheken ihrem Auftrag der Sammlung, Erschließung und Bereitstellung von Publikationen zukünftig weiterhin gerecht werden können.

Die traditionell sehr hohe Qualität der in Bibliotheken generierten Metadaten liefert die Grundlage für den zweiten Anwendungsbereich von KI in Bibliotheken. Die großen bibliografischen Datenbestände, oft erstellt unter Verwendung von kontrolliertem Vokabular oder gar einem Thesaurus, bieten eine exzellente Grundlage zur Anwendung und Weiterentwicklung von KI-Verfahren, die diese Daten vor dem Hintergrund unterschiedlicher Fragestellungen analysieren. Am Beispiel von Wissenschaftsdynamiken wurde dies in Abschnitt 2.2 dargestellt. Durch diesen Anwendungsbereich können Bibliotheken nicht nur als anerkannter Partner im Wissenschaftssystem Dienstleistungen für die Wissenschaft erbringen, sondern auf Augenhöhe mit Forschungspartnern aus der Wissenschaft gemeinsam Projekte durchführen, die ohne Bibliotheken als Partner nicht möglich wären.

4. Literatur

Arben Hajra, Klaus Tochtermann: Linking science: approaches for linking scientific publications across different LOD repositories. International Journal of Metadata, Semantics and Ontologies – Olney, Bucks: Inderscience, ISSN 1744-263X, Vol. 12 (2017), No. 2/3, pp. 124–141, doi:10.1504/IJMSO.2017.10011833

Helen Niemann, Martin Moehrle, Jonas Frischkorn (2017). Use of a new patent text-mining and visualization method for identifying patenting patterns over time: Concept, method and test application. Technological Forecasting and Social Change, 115, 210–220. doi.org/10.1016/j.techfore.2016.10.004

Lukas Galke, Eva Seidlmayer, Gavin Lüdemann, Lisa Langnickel, Tetyana Melnychuk, Konrad U. Förstner, Klaus Tochtermann, Carsten Schultz (2021), COVID-19++: A Citation-Aware Covid-19 Dataset for the Analysis of Research Dynamics, Big Data Analytics In: 2021 IEEE International Conference on Big Data (Big Data), 15–18 Dec. 2021 New York City: IEEE, 2021, pp. 4350–4355, doi:10.1109/BigData52589.2021.9671730 .

Martin Toepfer, Christin Seifert (2018-1): Fusion architectures for automatic subject indexing under concept drift. Int J Digit Libr 21, 169–189. https://doi.org/10.1007/s00799-018-0240-3

Martin Toepfer, Christin Seifert (2018-2): Content-Based Quality Estimation for Automatic Subject Indexing of Short Texts Under Precision and Recall Constraints. In: Méndez, E., Crestani, F., Ribeiro, C., David, G., Lopes, J. (eds) Digital Libraries for Open Knowledge. TPDL 2018. LNCS, vol 11057. Springer, https://doi.org/10.1007/978-3-030-00066-0_1

Robert M. Monarch, Christopher D. Manning: Human-in-the-loop machine learning: active learning and annotation for human-centered AI. Manning Publications, 2021.

Tetyana Melnychuk, Lukas Galke, Eva Seidlmayer, Konrad Förstner. Klaus Tochtermann, Carsten Schultz (2021). Früherkennung wissenschaftlicher Konvergenz im Hochschulmanagement. In: Hochschulmanagement – Bielefeld: Universitäts Verlag Webler, ISSN 1860-3025, 2021, No. 1/2021, pp. 24–28



Prof. Dr. Klaus Tochtermann: Nach wissenschaftlichen Stationen in Dortmund, Texas (USA), Ulm und Graz (Österreich) ist Klaus Tochtermann seit 2010 Direktor der ZBW – Leibniz-Informationszentrum Wirtschaft in Kiel und Hamburg. Er hat außerdem eine Professur für Digitale Informationsinfrastrukturen im Institut für Informatik der Christian-Albrechts-Universität zu Kiel. Sein aktueller Forschungsschwerpunkt liegt im Themenfeld Open Science, mit einem besonderen Schwerpunkt auf dem Management von Forschungsdaten. In diesem Themenfeld engagiert er sich in der wissenschaftspolitischen Beratung, sowohl auf nationaler als auch auf europäischer Ebene. (Foto: Sven Wied)



Dr. Anna Kasprzik koordiniert seit 2019 die Automatisierung der Sacherschließung (AutoSE) an der ZBW – Leibniz-Informationszentrum Wirtschaft. Sie hat Linguistik, Informatik und Psychologie studiert und 2012 im Bereich der Theoretischen Informatik promoviert. Nach einem Bibliotheksreferendariat am KIM Konstanz folgten Tätigkeiten in einem IT-Projekt beim Bibliotheksverbund Bayern und in der Forschung und Entwicklung an der TIB Hannover. Ihre Interessen umfassen verschiedene Ansätze aus der Künstlichen Intelligenz, inbesondere semantische Technologien, Machine-Learning-Verfahren und die Frage, wie diese beiden Herangehensweisen für die Inhaltserschließung sinnvoll verzahnt werden können.