ZBW Publikationsarchiv

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

Kasprzik, Anna

Contribution to Periodical — Published Version

Intelligente Technologien in Bibliotheken: Wie KI die Inhaltserschließung revolutioniert – Die Transformation der Inhaltserschließung durch KI-Technologien und deren Auswirkungen auf Bibliotheken. Ein Interview mit Anna Kasprzik (ZBW)

dpr Bibliotheken

Suggested Citation: Kasprzik, Anna (2024) : Intelligente Technologien in Bibliotheken: Wie KI die Inhaltserschließung revolutioniert – Die Transformation der Inhaltserschließung durch KI-Technologien und deren Auswirkungen auf Bibliotheken. Ein Interview mit Anna Kasprzik (ZBW), dpr Bibliotheken, ISSN 2512-9368, Digital Publishing Report GmbH, Nördlingen, Iss. Mai, pp. 29-31

This Version is available at:
http://hdl.handle.net/11108/622

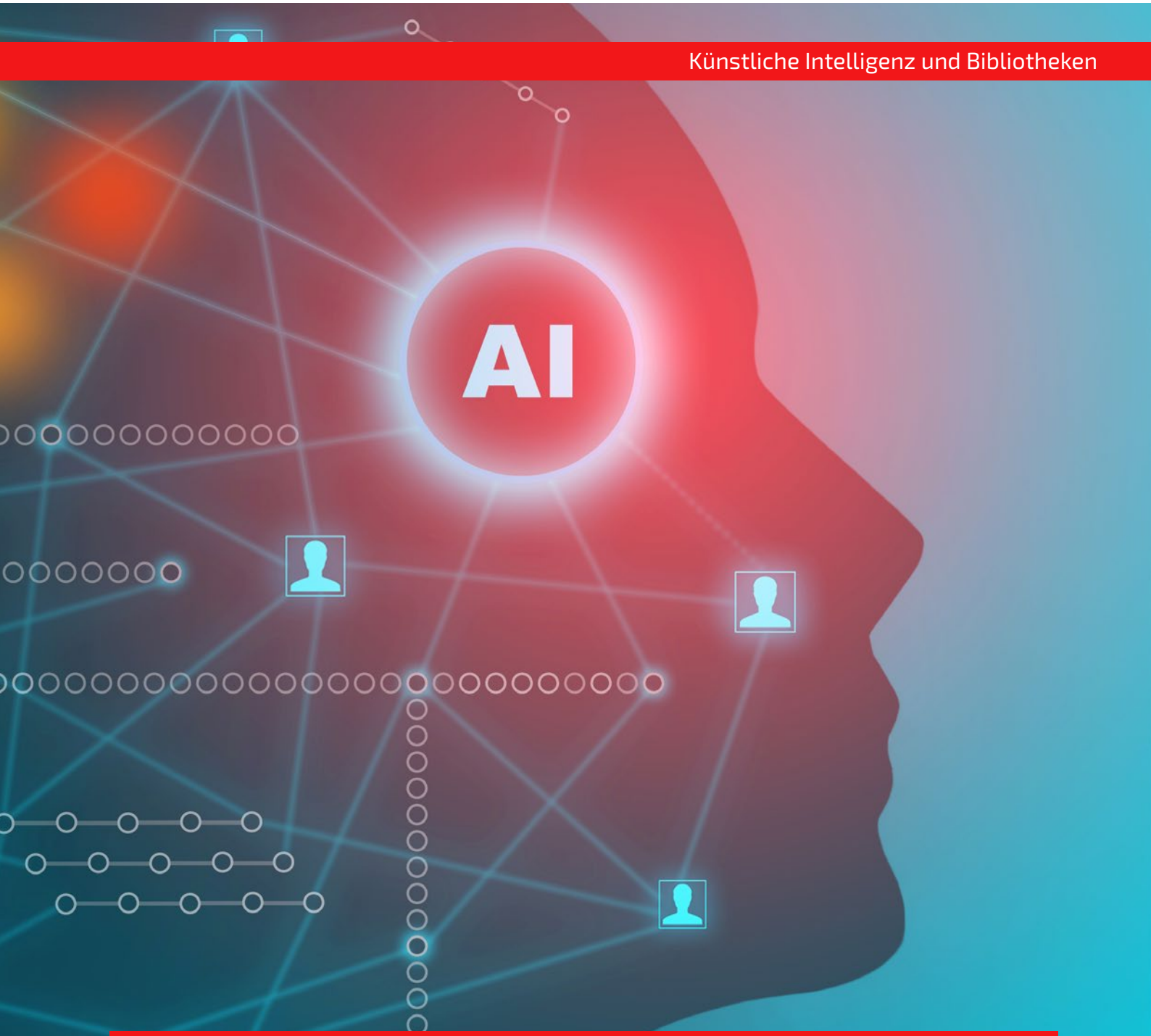
Kontakt/Contact
ZBW – Leibniz-Informationszentrum Wirtschaft/Leibniz Information Centre for Economics
Düsternbrooker Weg 120
24105 Kiel (Germany)
E-Mail: info@zbw.eu
https://www.zbw.eu/de/ueber-uns/profil-der-zbw/veroeffentlichungen-zbw

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

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

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Mitglied der Leibniz-Gemeinschaft



Intelligente Technologien in Bibliotheken: Wie KI die Inhaltserschließung revolutioniert

Die Transformation der Inhaltserschließung durch KI-Technologien und deren Auswirkungen auf Bibliotheken. Ein Interview mit Anna Kasprzik (ZBW)

In einer Zeit, in der künstliche Intelligenz die Landschaft der Informationstechnologie grundlegend verändert, kommt auch der Bibliothekswelt eine Schlüsselrolle bei der Implementierung von KI-Methoden zu. Dr. Anna Kasprzik, zuständig für die Koordination der Automatisierung der Sacherschließung an der ZBW – Leibniz-Informationszentrum Wirtschaft, teilt in einem Interview Einblicke in die Anwendung von maschinellem Lernen zur Optimierung der Inhaltserschließung. In einer Zeit, in der künstliche Intelligenz die Landschaft der Informationstechnologie grundlegend verändert, auch der Bibliothekswelt eine Schlüsselrolle bei der Implementierung von KI-Methoden zu.

An welchen Stellen wird derzeit maschinelles Lernen aktuell in der Bibliothekspraxis eingesetzt?

Maschinelles Lernen ließe sich prinzipiell an vielen Stellen in Bibliotheken einsetzen – im Wesentlichen kann man an allen Stellen, die sich zur Automatisierung anbieten, auch prüfen, ob sich Methoden aus dem maschinellen Lernen für den jeweiligen Anwendungsfall eignen – etwa für OCR, Extraktion von Metadaten für die Formalerschließung und anderen Textelementen aus Volltexten, für die Inhaltserschließung.

Soweit ich das überblicken kann, glaube ich allerdings, dass viele Institutionen hier mit produktiven Implementierungen in der Regel noch relativ weit am Anfang stehen und sich über Projekte vorsichtig herantasten oder auch die entsprechenden Ressourcen gar nicht zur Verfügung haben.

Welche spezifischen Herausforderungen gab es bei der Entwicklung und Implementierung von Machine-Learning-Lösungen an der ZBW, und wie wurden diese gelöst?

Dr. Anna Kasprzik promovierte nach einem Studium der Linguistik, Informatik und Kognitionspsychologie in der Theoretischen Informatik und legte anschließend das Bibliotheksreferendariat ab. Es folgten Tätigkeiten in einem IT-Projekt beim Bibliotheksverbund Bayern und in der Forschung und Entwicklung an der TIB Hannover. Anna Kasprzik koordiniert seit 2018 die Automatisierung der Sacherschließung (AutoSE) an der ZBW – Leibniz-Informationszentrum Wirtschaft und befasst sich schwerpunktmäßig mit verschiedenen Ansätzen aus der Künstlichen Intelligenz, insbesondere Wissensorganisationssystemen und Machine Learning, und der Frage, wie diese unterschiedlichen Herangehensweisen für die Inhaltserschließung sinnvoll verzahnt werden können.

Copyright: ZBW; Foto: Hagen Immel.

An der ZBW wurde seit fast zwei Jahrzehnten über Möglichkeiten zur Automatisierung der Inhaltserschließung nachgedacht – es gab zwei Vorläuferprojekte mit externen Partnern und kommerzieller Software. Da die Ergebnisse in beiden Fällen den Qualitätsansprüchen der Bibliothek aber nicht genügten, wurde in einer Neuorientierungsphase um 2014 herum der Entschluss gefasst, die notwendige angewandte Forschung bezüglich geeigneter Machine-Learning-Methoden im Rahmen einer Promotionsstelle im eigenen Haus durchzuführen. Die anschließend entwickelten prototypischen Ansätze fanden jedoch weiterhin keine Anwendung in der Praxis – erst ein strategischer Statuswechsel vom Projekt zur Daueraufgabe brachte den Wandel, da daraufhin eine weitere Stelle für den Aufbau der benötigten Softwarearchitektur geschaffen wurde. Ab 2020 wurde der produktive Dienst („AutoSE“) in enger Abstimmung mit den Inhaltserschließenden im Haus sukzessive aufgebaut, er läuft seit 2021 und hat seither bereits ca. 1,7 Mio. Ressourcen verschlagwortet, das ist etwa ein Viertel des ZBW-eigenen Bestandes.

Welche Vorteile hat der Einsatz von Machine Learning für die automatisierte Inhaltserschließung an der ZBW gebracht?

Über die vollautomatisierte Inhaltserschließung konnten wir große Mengen von Ressourcen inhaltserschließen, die sonst gar nicht erschlossen worden wären. Und da wir unseren Output parallel dazu als Schlagwortvorschläge zur Verfügung stellen, unterstützen wir so auch die intellektuelle Inhaltserschließung. Erschließen-

de an der ZBW berichten durchaus von einem Geschwindigkeits- und Komfortzuwachs, seit die Anknüpfung an die Erschließungsplattform „Digitaler Assistent“ umgesetzt wurde. Außerdem ist durch die erstmalige Implementierung einer modular aufgebauten Automatisierungsarchitektur samt Schnittstellen zu anderen Metadatensystemen die Einführung weiterer methodischer Neuerungen für diesen Bereich in Zukunft beträchtlich einfacher geworden, da wir dadurch sozusagen einen infrastrukturellen Fuß in der Tür haben.

Wie werden sich die Anforderungen an das Fachpersonal in Bibliotheken verändern, wenn KI-Lösungen implementiert werden, etwa im Bereich der Erschließung?

Sie werden einerseits ein gewisses Maß an Kontrolle abgeben und sich die konzeptionelle Aufgabe einer hochqualitativen Erschließung nun mit Menschen aus anderen Fachbereichen – etwa der Informatik und der Softwareentwicklung – teilen müssen und sich andererseits daran gewöhnen müssen, dass ihre Expertise im Bereich der Wissensorganisation über neue, ungewohnte Kanäle und Arbeitsabläufe in den Erschließungsvorgang einfließen wird.

Meiner Einschätzung nach stehen die Erschließenden diesem Wandel jedoch mehrheitlich eher positiv gegenüber und sind sich bewusst, wie wertvoll beispielsweise ihre Qualitätsbeurteilungen für die kontinuierliche Weiterentwicklung und Verbesserung der maschinellen Methoden sind.

Wie wirken sich Anwendungen wie ChatGPT, die auf Large Language Models aufsetzen, auf die traditionellen Methoden der Inhaltserschließung und Datenverwaltung in Bibliotheken aus?

Anwendungen wie ChatGPT wecken in einigen Kreisen Hoffnungen auf eine ganz neue Form der Suche, bei der Nutzende ihre inhaltlichen Fragen in natürlicher Sprache stellen können und eine ausformulierte Antwort zurückbekommen statt sich durch eine Trefferliste wühlen zu müssen. Entsprechend wurde auch schon prognostiziert, dass eine derartige Interaktion zwischen Nutzenden und Volltexten über LLMs sowohl die Inhaltserschließung als auch die gezielte Metadatenerstellung an sich obsolet machen werde. Das halte ich zumindest für fraglich, solange diese Anwendungen nicht mit etablierten Wissensbasen untersetzt sind, da LLMs ja nur darauf trainiert sind, möglichst natürlich klingenden Text zu generieren, aber in ihrer Reinform keinen Mechanismus enthalten, der den Output zum Beispiel auf Konsistenz (in sich und mit bereits bekannten Informationen) prüft. Ansätze für solche Mechanismen gibt es bereits, z.B. Retrieval-Augmented Generation, und diese wiederum beruhen teilweise auch auf der Nutzung von Metadaten im weitesten Sinne. Hier haben Bibliotheken eine Chance, relevant zu bleiben, müssen sich dazu jedoch einerseits auf ihre Kernaufgabe „Metadaten“ zurückbesinnen, diese andererseits aber in neue Formen gießen, die als Input für diese Modelle geeignet sind. Das erfordert noch einen beträchtlichen Innovationsschub.

Welche Überlegungen sollten bei der Implementierung von KI-Technologien in Bibliotheken noch berücksichtigt werden?

Die neuen Modelle sind extrem ressourcenhungrig – nicht nur beim Training, sondern auch im Betrieb. Bibliotheken stehen generell für Nachhaltigkeit, und daher sollten wir uns schon sehr genau überlegen, welche Einsatzszenarien einen derartigen Energieverbrauch und CO2-Ausstoß überhaupt rechtfertigen und was der umweltschonendste Weg ist, diese umzusetzen.