Künstliche Intelligenz bei EBSCO

Frank Bertsch und Roman Piontek

Einführung

EBSCO Information Services ist ein führender Anbieter wissenschaftlicher Online-Inhalte und Suchtechnologien für forschungsnahe, schulische wie öffentliche Bibliotheken sowie für medizinische Einrichtungen, Unternehmen und Behörden weltweit. EBSCO hat nun damit begonnen, unter anderem mehr und mehr KI einzusetzen, um die Arbeit mit Forschungsmaterialien zu vereinfachen.

Verantwortungsvoller Einsatz von KI bei EBSCO

EBSCO hat sich zur Aufgabe gesetzt, KI stufenweise so einzuführen, dass wissenschaftliche und ethische Standards sowie „Best Practices" bestmöglich eingehalten werden. Hier sind einige dieser Grundsätze und Praktiken:

Qualität und Genauigkeit

EBSCO stellt sicher, dass zum Einsatz kommende KI-Modelle auf qualitativ hochwertigen und zuverlässigen Daten beruhen. Dies wird durch den Einsatz von „Retrieval-Augmented Generation" (RAG) und „Knowledge Graphs" erreicht. Diese Zusatztechnologien tragen dazu bei, die Präzision zu erhöhen und die Fehlerrate, zumal die sogenannter „Halluzinationen", zu verringern.

Transparenz

Dies ist ein zentraler Aspekt des KI-Einsatzes. EBSCO legt offen, wie die KI-Modelle funktionieren, wie sie ausgewählt wurden, wie die Qualität bewertet wird und welche Daten verwendet werden. All dies findet sich in der Dokumentation zu den einzelnen KI-Features, dort im jeweiligen Abschnitt „Transparenz".

Ethische Normen

EBSCO hält sich an ethische Standards und Richtlinien, um sicherzustellen, dass KI-Technologien verantwortungsvoll eingesetzt werden. Dazu gehört auch die minutiöse Einhaltung erforderlicher Datenschutzbestimmungen.

Zusammenarbeit mit Experten

EBSCO arbeitet stets eng mit Bibliothekaren und Fachreferenten zusammen, um sicherzustellen, dass die KI-Routinen dem neuesten Stand von Forschung und Praxis genügen. Diese fortwährende Zusammenarbeit trägt dazu bei, Qualität und Relevanz KI-gestützter Funktionalitäten kontinuierlich zu verbessern.

Umweltbewusstsein

EBSCO berücksichtigt u. a. auch die Umweltauswirkungen von KI-Technologien, da diese unweigerlich entsprechenden Ressourceneinsatz erfordern. Das Unternehmen ist bestrebt, die Energieeffizienz der KI-Modelle zu optimieren und den ökologischen Fußabdruck somit zu minimieren.

Bildung und Ausbildung

EBSCO bietet darüber hinaus Ressourcen und Schulungen an, um Bibliotheken und Forschende bei der verantwortungsvollen KI-Nutzung zu unterstützen. Dazu gehört die breite Vermittlung von Fähigkeiten im Umgang mit KI sowie die Förderung eines kritischen Verständnisses der Technologien.

EBSCOs aktuelle KI-Funktionalitäten und Roadmap

EBSCO hat einstweilen drei neue KI-Funktionen eingeführt: „AI Insights" und „Natural Language Search" (NLS) sowie „Dyna-AI". Diese Funktionalitäten wurden mit dem Bestreben entwickelt, den Forschungsprozess zu erleichtern und die Nutzerfreundlichkeit zu verbessern.

„AI Insights":

Diese Funktion generiert aus Artikeln auch und vor allem bei Abwesenheit von „Abstracts" kurze, prägnante Zusammenfassungen mit etwa fünf thematischen Punkten, die es erleichtern, die fachliche Relevanz eines Artikels auf einen Blick beurteilen zu können.

Artikelvolltexte werden einem „Large Language Model" (LLM) zeitgleich mit der Sucheingabe (gemäß den jeweiligen rechtlichen Vorgaben) übermittelt, das den Text daraufhin als Grundlage für die KI-Ergebnisse verwendet.

Die Zusammenfassungen entstehen auf Basis des Volltextes und verwenden eine Methode namens „Retrieval-Augmented Generation" (RAG), um faktische Ungenauigkeit weitestgehend zu vermeiden, ohne dass das KI-Modell dafür extra trainiert werden müsste. Die KI-Ausgabe ist dennoch stets mit einem entsprechenden Vermerk versehen, da hier stets „semantische Vorsicht" geboten ist.

EBSCO verwendet dafür folgende Beurteilungsmatrix:

Genauigkeit: Sind die Informationen anhand des Artikelinhaltes überprüfbar korrekt?

- Terminologie: Stimmt die Terminologie mit dem Artikelinhalt überein?
- Thematik: Kommen Hauptthemen des Artikels nachweislich zur Sprache?
- Stil: Bleibt der Stil des Ausgangsdokumentes gewahrt?
- Nützlichkeit: Ist die entstandene Zusammenfassung dem Forschungsprozess dienlich?

„Natural Language Search“ (NLS)

Diese Funktionalität ermöglicht die Suche in natürlicher Sprache, so dass man u. U. auch dann erfolgreich suchen kann, falls nicht von vornherein die exakte Terminologie bekannt sein sollte. NLS zerlegt die Suchanfrage in „Noun Phrases“, ordnet nach Möglichkeit Synonyme zu und passt die Anfrage im Hintergrund an.

Die Suchanfrage wird an ein LLM übermittelt, welches sie interpretiert, natürlichsprachige Eingaben analysiert und daraufhin die erweiterte Anfrage an die jeweilige EBSCO-Suchmaschine zurückübermittelt.

Für die Darstellung der Ergebnisse verwendet EBSCO die firmeneigene Relevanzgewichtung. Weitere Details zur Funktionsweise unserer Suchtechnologien finden Sie hier: https://www.ebsco.com/de-de/wissenschaftliche-bibliotheken/produkte/ebsco-discovery-service/hochwertige-daten-praezise-suche.

Die neue Suchabfrage wird im Suchfenster auch in boolescher Form angezeigt, um eine Brücke zwischen der natürlichen Sprache des Nutzers und fortgeschritteneren Suchmethoden zu schlagen.

„Dyna-AI“:

Dyna-AI liefert Medizinern Informationen, die am Ort der Behandlung zwecks fundierter Entscheidungen benötigt werden. Dyna-AI bezieht die Informationen ausschließlich aus den expertengepflegten und evidenzbasierten Inhaltsmengen aus „DynaMedex“, „DynaMed“ und „Dynamic Health“, wobei die jeweilige Herkunft samt Quellenlinks angegeben wird. Mit Hilfe von KI und RAG liefert Dyna-AI somit rasche Antworten auf natürlichsprachige Anfragen medizinischen Personals – ohne jedoch Anspruch darauf zu erheben, die klinische Beurteilung zu ersetzen.

Weitere geplante Entwicklungen und Ideen

Die nachstehenden Überlegungen befinden sich allesamt noch in einem frühen Stadium, und es steht derzeit somit noch nicht fest, in welcher Abfolge und zu welchem Zeitpunkt sie in Produkte inkorporiert werden.

Zukunft der Empfehlungen

Die Empfehlungsdienste von EDS und EBSCOhost werden in naher Zukunft einer Neugestaltung unterliegen. Erste Nutzertests zeigen, dass die Überarbeitung der zugrunde liegenden Funktionen das Potenzial hat, die Relevanz der vorgeschlagenen Inhalte deutlich zu erhöhen. Wir untersuchen auch Graphalgorithmen, die es uns ermöglichen würden, eine sogenannte „Multi-Hop-Analyse“ durchzuführen, um möglicherweise Inhalte zu empfehlen, die nicht nur unmittelbaren, sondern u. U. auch entfernteren Bezug zur Ursprungssuchanfrage aufweisen, also mitunter zwei oder drei „Sprünge“ entfernt sind.

Auf diese Weise könnten komplexe oder indirekte Beziehungen zwischen zunächst teils diversen Themenblöcken zutage gefördert werden.

Ein künftiger KI-Assistent für die Literatursuche

Dieser befindet sich ebenfalls noch im Planungsstadium. EBSCO befasst sich derzeit damit, die Funktionalitäten zu bewerten und Nutzerfeedback dazu zu erheben.

Das Hauptaugenmerk wird darauf liegen, Forschende bei der Auswertung und Analyse einer breiteren Literatursammlung zu unterstützen. Es werden dabei keine originären Literaturrezensionen entstehen, da dies im Widerspruch zum Forschungsauftrag stünde: eigene und kritische Auswertungen vor einem bestimmten Theoriehintergrund vorzunehmen.

Es sollen jedoch Werkzeuge entstehen, die diese Auswertung zumindest beschleunigen und übersichtlicher gestalten können. Hier einige der Ideen, wie ein solcher Literatursuchassistent an Gestalt gewinnen könnte:

Es könnte z. B. hilfreich sein, bestimmte Aspekte einer vorausgewählten Gruppe von Inhalten hervorzuheben. Beispielsweise können unterschiedliche kombinierte F- und Effektivitätsmaße aus einer Vielzahl von Studien oder die Stichprobengröße berücksichtigt werden bzw. verwendete Methoden kontrastiv verglichen werden.

Bei paralleler Auswertung zahlreicher Arbeiten ist die Herausforderung häufig, den nötigen Überblick zu behalten. Die genaue Herkunft einer kopierten Textstelle oder eines Studienergebnisses kann mit zunehmender Zeit zuweilen nicht mehr exakt nachzuvollziehen sein. Mittels eines KI-Systems, dem man schlichtweg die Frage stellen könnte, woher ein bestimmtes Faktum oder eine Textstelle stammt, ließe sich mehr Struktur in die Fülle zu bearbeitender Informationen bringen. Ziel ist dabei die bequemere Gestaltung von Recherchen.

Zukünftige Anreicherung unserer Suchmaschine

Um sicherzustellen, dass das Suchen innerhalb von Fachgebieten jeweilige Spezifika einbezieht, könnte es hilfreich sein, die Abfrage a priori auf bestimmte Inhaltstypen oder Zeiträume einzugrenzen. Mitunter gibt es auf Fragen gar eine ganze Reihe praktischer Antworten. Methodisches Zusammentragen von Nachweisen aus wissenschaftlicher Literatur zwecks gezielter Anreicherung von Suchergebnissen böte sich an. Diese geplante Weiterentwicklung wird wahrscheinlich an die bereits zur Sprache gekommene „NLS“ anknüpfen.

Chatten mit einem Dokument

Es kommen weithin bereits diverse Anwendungen zum Einsatz, die es ermöglichen, mit einem Dokument regelrecht zu „chatten“. EBSCO plant, dieses Paradigma wie folgt zu modifizieren:

Der übliche Grund, einen Artikel nicht nur zu überfliegen sondern konzentriert durchzulesen, ist, ihn genau

zu analysieren, verwendete Methoden zu verstehen und Parallelliteratur zu erfassen. Aufgrund geradezu exponentiell ansteigender Artikelfülle ist diese Vorgehensweise nur bei einem Bruchteil gegenwärtig vorhandener Literatur möglich.

Daher prüft EBSCO die Möglichkeit, einen KI-Assistenten zu entwickeln, der bei der Aufschlüsselung und Kategorisierung relevanter Informationen aus wissenschaftlicher Literatur hilfreich zur Seite stehen kann.

Erstellung und Anreicherung von Metadaten

Die allgegenwärtige Arbeit mit Metadaten (insbesondere die komplexe Arbeit mit MARC-Datensatzfeldern) ist ein weiteres potentielles KI-Feld, etwa im FOLIO-Kontext. Ein gemeinhin thematisierter Aspekt ist z. B. die Unterstützung bei der Katalogisierung mehrerer Exemplare und letztlich die Durchführung von „Entity Resolutions" bzw. „Record Linkage", um Suchergebnisse besser zu bündeln und mehrschichtige Datenbeschreibung zu sichern.

Zusammenführen und Deduplizieren

Hierunter fällt nicht allein die klassische Deduplikation, sondern auch die Identifizierung eng verwandter Datensätze: die Grenzen zwischen FRBRisierung und Deduplikation sind in bestimmten Fällen fließend und stets von der Forderung nach verbesserter Datenqualität begleitet. Einzelheiten zu diesen Punkten werden derzeit noch intern erarbeitet.

Bestandsentwicklung

Gut nachvollziehbare Bestandsentwicklung spielt bekanntlich eine zentrale Rolle bei der Übersicht dessen, was und wie eine Bibliothek bereits zur Verfügung stellen kann. Falls etwa ein neuer Studiengang oder ein neues Fachgebiet (wie etwa „KI" selbst) angeboten oder ausgebaut wird, ist es eruierenswert, inwieweit der aktuelle Bestand bereits den sich anbahnenden Anforderungen entspricht und an welchen Stellen Ergänzungen erforderlich sein werden. Zunehmend KI-gestützte Bestandsaufschlüsselung könnte in diesem Zusammenhang hilfreich eingesetzt werden.

Erwerbung

Ein weiteres Beispiel stets wachsender Informationsfülle ist gewiss auch der Erwerbungsprozess, der sich mitunter komplex gestalten kann – bei Erneuerungen, aber ebenso und insbesondere im Falle von Neuerwerbungen. KI könnte dabei helfen, alle erforderlichen Schritte und Prozeduren zwecks erhöhter Effizienz und Präzision bestmöglich zu verzahnen.

Urheberrecht und KI

Die zunehmend rascher werdende Entwicklung von KI hat weitreichende Auswirkungen auf viele Bereiche, sehr sichtbar auch auf die Arbeit von und mit Bibliotheksdienstleistern. Inhaltsanbieter stehen zwar vor neuen Chancen, aber auch vor Herausforderungen, nicht zuletzt im Hinblick auf das Urheberrecht. Wie eingangs erwähnt, hält sich EBSCO bewusst an eine Reihe externer sowie interner Grundsätze und Regeln. Dazu gehört die enge Zusammenarbeit mit Urheberrechtsinhabern. Die genaue Einhaltung von Lizenzvereinbarungen und weiterer rechtlicher Vorgaben soll lückenlos sichergestellt werden.

Generative KI-Systeme werfen durchaus komplexe urheberrechtliche Fragen auf. Da KI-Modelle vornehmlich auf Basis möglichst großer Datenmengen trainiert werden, die unter anderem geschützte Werke enthalten können, stellt sich die natürliche Frage, unter welchen Bedingungen diese Daten überhaupt weiterverwendet werden dürfen. Wer besitzt die Rechte an den durch KI erstellten Zusammenfassungen oder Antworten? Wie können Bibliotheken innovative KI-Tools anbieten und dennoch Urheberrechtsverletzungen ausschließen?

Ein notwendiger Schritt ist der Abschluss eindeutiger Lizenzvereinbarungen mit Rechteinhabern, die verstärkte urheberrechtskonforme Nutzung von Open-Access-Materialien als Trainingsdaten und die Entwicklung interner Richtlinien für den Umgang mit KI und Urheberrecht.

Schulungen für Bibliotheken und Nutzerschaft sind unerlässlich, um das Bewusstsein für dieses Thema zu schärfen. Eine enge Zusammenarbeit zwischen Bibliotheken, Verlagen, Bibliotheksdienstleistern und KI-Entwicklern ist zudem notwendig, um tragfähige Lösungen zu entwickeln.

Zusammenfassung

KI wird künftig zweifellos immer mehr dazu beitragen, den Alltag etwas weniger mühsam zu gestalten. Zugleich erhalten Bibliotheken und Forschende mehr Einblicke und Daten, die Entscheidungsfindungen begünstigen, und gewinnen somit wertvolle Zeit, um sich vermehrt den wichtigen Fragen zu widmen – weniger der Routinevorbereitung.

EBSCO entwickelt stets forschungs- und kundenorientiert, um den Recherche- und Datenverwaltungsprozess für alle Beteiligten angenehm, hilf- und aufschlussreich sowie effektiv zu gestalten. Und wir verpflichten uns überdies, dies auf verantwortungsvolle Weise zu tun.

Frank Bertsch (SaaS Director)
Roman Piontek (SaaS Director)
EBSCO Information Services GmbH
Volkmarstraße 18, 12099 Berlin, Deutschland