

Maximilian Kähler, DNB

Erschließung und KI –

Ergebnisse aus vier Jahren KI-Projekt an der DNB

Inhaltsverzeichnis

- 1. Einführung Automatische Erschließung**
- 2. Status Quo der maschinellen
Beschlagwortung**
- 3. Welche „KI-Ansätze“ gibt es aktuell**
- 4. Vergleichende Evaluation**

Automatische Inhaltserschließung in der DNB



Automatische **Klassifizierung** von Online- und ausgewählten Printpublikationen mit DDC-Sachgruppen und DDC-Kurznotationen



Automatische **Indexierung** von Online- und ausgewählten Printpublikationen mit normierter Terminologie GND

Dafür verwenden wir das Open-Source-Toolkit **annif**

Poley, C etal. (2025). Automatic Subject Cataloguing at the German National Library.

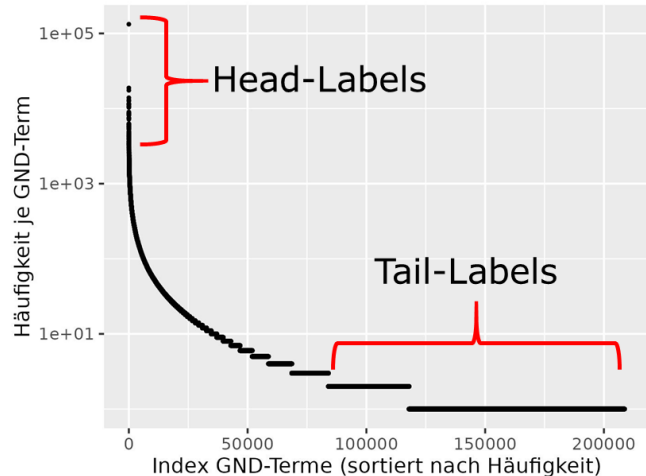
<https://doi.org/10.53377/lq.19422>

Suominen, O. Etal. (2022). Annif and Finto AI: Developing and Implementing Automated Subject Indexing.

<https://doi.org/10.4403/jlis.it-12740>

Warum ist maschinelle Beschlagwortung mit der GND so schwer?

Long-Tail Charakteristik



*GND gesamt**
1.428.353 Schlagwörter

*mind. einmal oder häufiger im Bestand der DNB vernüpft***
208.688 Schlagwörter

nicht im Bestand der DNB vernüpft
1.219.665 Schlagwörter

*Anm.: Nur Katalogisierungslevel 1 oder z und aus dem Teilbestand s

**Anm.: Bezieht sich nur auf die Verknüpfung mit *deutschsprachigen* Publikationen in unserem Trainings-Datensatz

Wie kann „KI“ den Status-Quo verbessern?

- A. Erweiterung bestehender Ansätze durch **verbesserte Text-Präsentation** (unter Einsatz von Transformer-Modellen)
- B. Entwicklung neuer Verfahren auf Basis generativer KI (Einsatz großer Sprachmodelle)

A Weiterentwicklung der „alten“ Ansätze durch Transformer

Embedding-basiertes Matching (EBM)¹:

Das String-basierte Matching-Verfahren im lexikalischen Ansatz wird durch einen Abgleich von transformer-basierten Text-Embeddings ersetzt oder ergänzt

XR-Transformer²:

Die Bag-of-Words basierte TFIDF-Vektorisierung der Texte wird durch Embeddings eines feinabgestimmten Transformermodelles ersetzt oder ergänzt.

¹Prototyp aus dem DNB-KI-Projekt:

<https://github.com/deutsche-nationalbibliothek/ebm4subjects>

²Zhang et al. (2021). *Fast Multi-Resolution Transformer Fine-tuning for Extreme Multi-label Text Classification*.

<https://arxiv.org/abs/2110.00685v2>

B Neue Welten: Beschlagwortung mit großen Sprachmodellen

- Schlüsselproblem: wie bringe ich einem LLM ein großes Ziel-Vokabular wie die GND bei?
- Bisheriger Ansatz^{1,2}: LLM-Ensemble mit Few-Shot-Prompting
 - Generiere freie Schlagwortvorschläge mit einem LLM
 - Erzeuge ein Mapping auf die GND durch ein (kleineres) Encoder-Modell
 - Wiederhole die Prozedur für verschiedene Modelle und Prompts

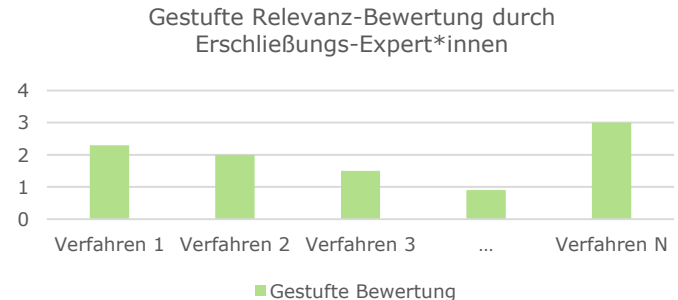
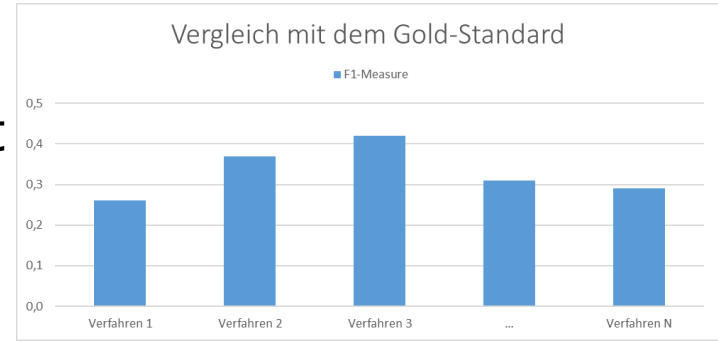
¹Kluge, L., & Kähler, M. (2024).
Few-Shot Prompting for Subject Indexing of German Medical Book Titles.
<https://aclanthology.org/2024.konvens-main.16>

²Prototyp aus dem DNB-KI-Projekt:
https://github.com/deutsche-nationalbibliothek/semEval25_1lmensemble

Vergleich der verschiedenen Ansätze

Unsere zwei Säulen der Evaluation

1. Maschinelle Vorschläge werden mit RSWK-Gold-Standard abgeglichen
2. Inhaltserschließer*innen bewerten die Nützlichkeit der maschinellen Vorschläge aus Retrieval-Perspektive



Vergleich von 5 Titel-Modellen am konkreten Beispiel

Beispiel-Titel: „Forschungsmoral in der qualitativen Sozial- und Gesundheitsforschung
Konflikte – Reflexion – Expertise“

GND-Term (Fett RSWK-Gold-Standard)	Omikujii	XR-Transformer	Embedding-bas. Matching	MLLM (lexikal. Matching)	LLM-Ensemble
Forschung	0,075	0,056	-	-	-
Sozialer Konflikt	0,078	-	-	-	-
Qualitative Sozialforschung	0,097	0,053	0,431	-	0,567
Medizin	0,115	0,052	-	-	-
Gesundheitswesen	0,141	0,050	-	-	-
Empirische Sozialforschung	-	-	-	-	0,127
Wissenschaftsethik	-	-	0,073	-	0,573
Qualitative Methode	-	0,050	0,125	-	-
Forschungsprozess	-	-	-	-	-
Gesundheitswissenschaften	-	-	-	-	0,089
Moralisches Handeln	-	-	-	-	-
Konflikt	-	-	0,059	0,076	-
Forschungsmethode	-	-	0,085	-	-
Expertise	-	-	-	0,139	-
EXPERTIS	-	-	-	0,139	-

True positives

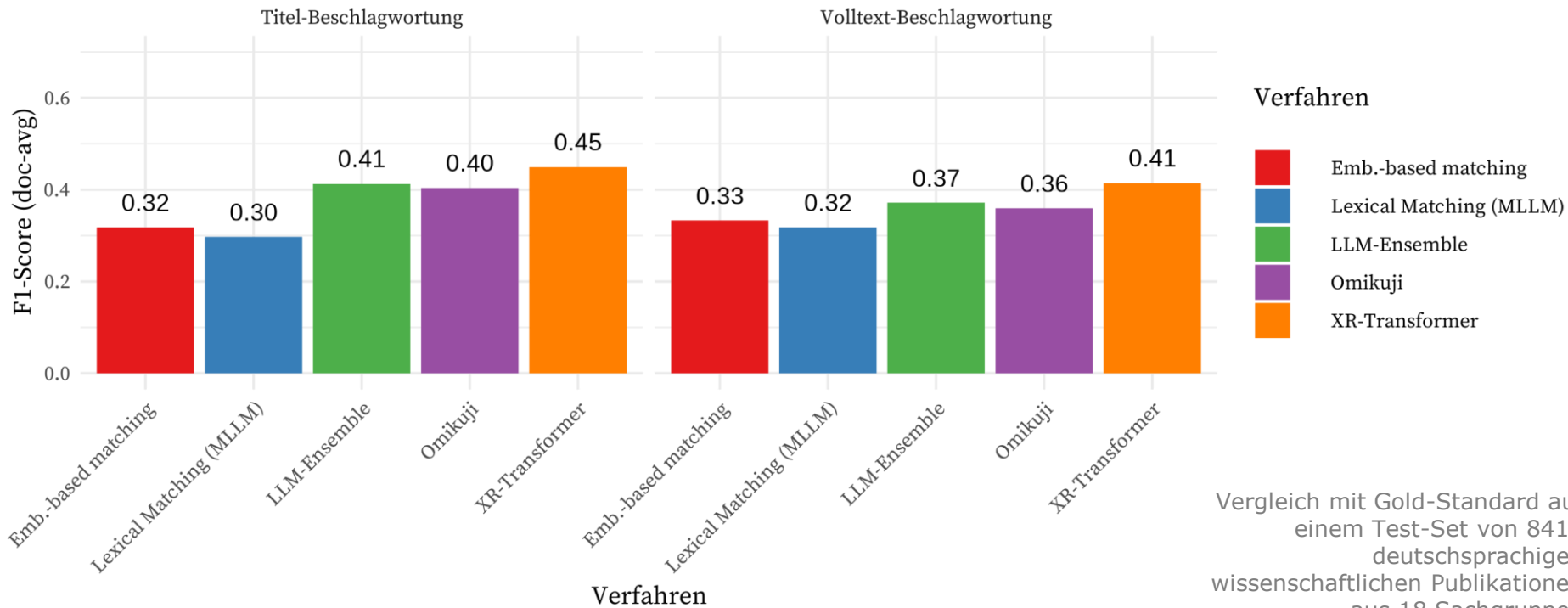
False negatives

False positives

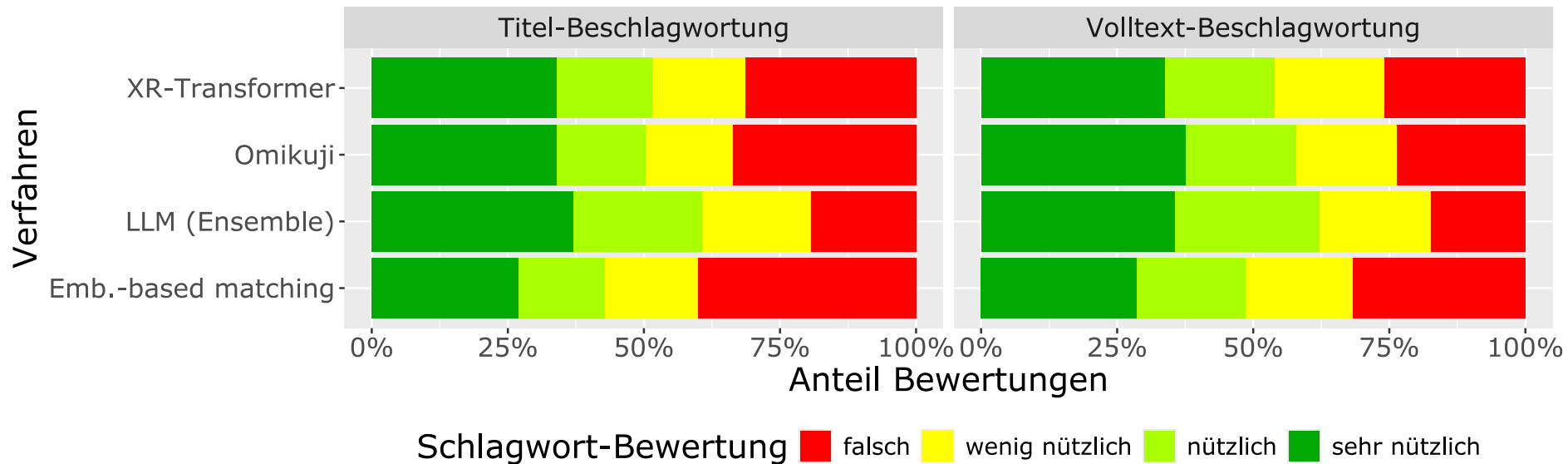
True negatives

Zahlenwerte sind Confidence-Scores der jeweiligen Verfahren (nicht normiert)

Ergebnisvergleich - Gesamt-Performance

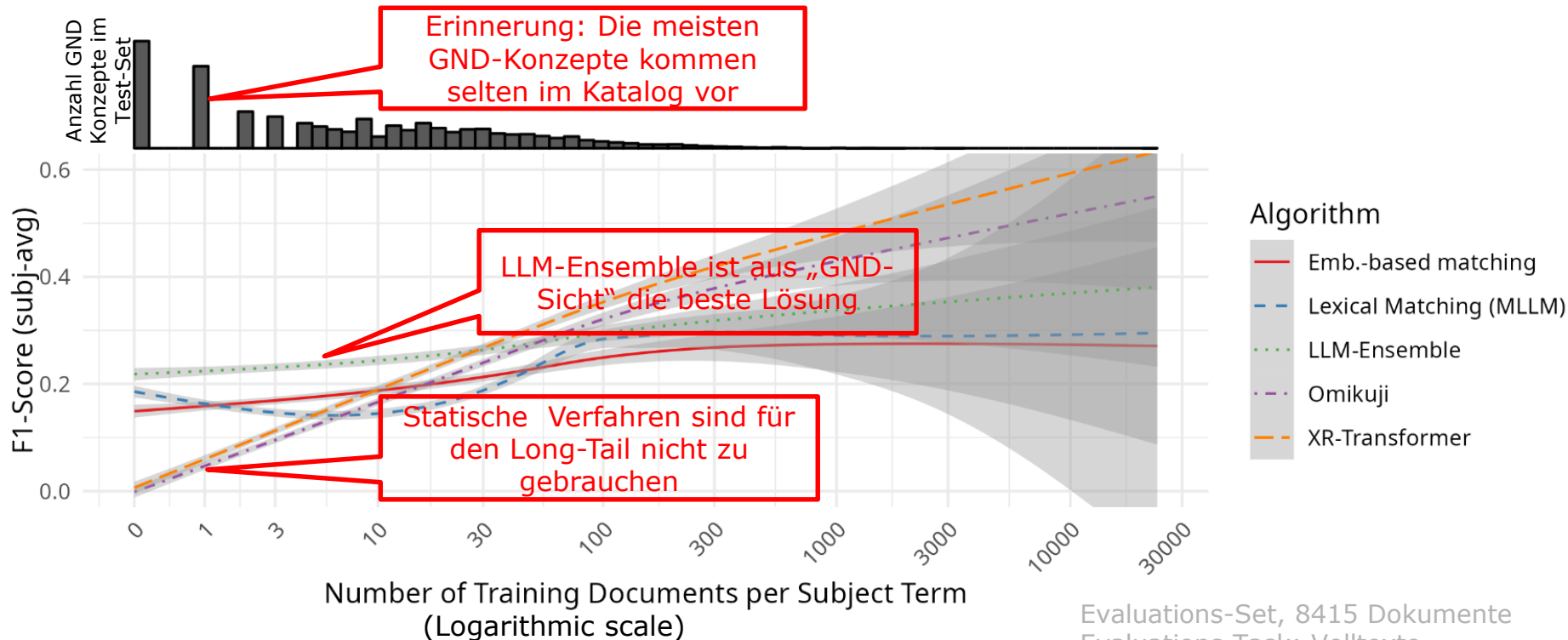


Gestufte Schlagwortbewertungen



Jedes Verfahren wurde auf jeweils einem Test-Set von ca. 1100 dt. wiss. Texten evaluiert. Bewertet wurden die Top-5 Vorschläge pro Dokument und Verfahren.

Die "GND-Sicht" auf die Qualität der Beschlagwortung



Fazit: KI-Einsatz in der Inhaltserschließung

- Transformer-Modelle können existierende Erschließungsverfahren verbessern
- Der Einsatz von Sprachmodellen bringt einen qualitativen Sprung, ABER
- Automatische Indexierung bleibt ein komplexes Problem und es ist wichtig, dass sich auch Kultur-Institutionen in die Forschung einbringen
- Kein Beschlagwortungsverfahren ist allein die beste Lösung
- Stärken und Schwächen können in einem Ensemble geschickt kombiniert werden
- Kosten-Nutzen Trade-Off kann für „schlankere“ Verfahren (MLLM, EBM) günstiger sein als für LLMs

Vielen Dank!

Projekt-Team: Lisa Kluge, Katja Konermann, Maximilian Kähler

Fragen und Anregungen an:

Maximilian Kähler

m.kaehler@dnb.de

 @mfakaehler@openbiblio.social

Mehr Infos zum KI-Projekt:

<https://www.dnb.de/ki-projekt>

<https://blog.dnb.de/ki-projekt-gewinnt-best-paper-award/>