

# Projekt: Natürlich-sprachiger Sucheinstieg

## Unternehmen/Organisation & Ansprechpartner

Als Deutsche Nationalbibliothek sind wir die zentrale Archivbibliothek Deutschlands. Wir sammeln, dokumentieren und archivieren alle Medienwerke in Schrift, Bild und Ton, die seit 1913 in und über Deutschland oder in deutscher Sprache veröffentlicht werden. Ob Bücher, Zeitschriften, CDs, Schallplatten, Karten oder Online-Publikationen – wir sammeln ohne Wertung, im Original und lückenlos.



Kontaktperson: IT, Tracy Arndt, t.arndt@dnb.de

## Problemstellung & Motivation

Die Durchsuchung des umfangreichen Datenbestands der Deutschen Nationalbibliothek (DNB) erfordert oft ein detailliertes Verständnis der vorhandenen Ressourcen sowie der zugrunde liegenden Metadatenorganisation. Zahlreiche, komplexe Verknüpfungen zwischen den Datensätzen können mit herkömmlichen Suchmethoden nur unzureichend abgebildet werden. Zwar ermöglichen SPARQL-Abfragen eine präzise Erfassung dieser Beziehungen, doch stellen sie für viele Nutzer eine beträchtliche Hürde dar.

Neben den Publikationsdaten betreibt die DNB auch die Gemeinsame Normdatei (GND), in der zentrale Entitäten aus dem kultur- und geisteswissenschaftlichen Bereich – wie Personen, Werke, Orte, Sachbegriffe und Körperschaften – standardisiert erfasst werden.

Ein möglicher Ansatz zur Überwindung dieser Herausforderungen ist der Einsatz von Graph-RAG in Verbindung mit SPARQL-Abfragen. Dieser Ansatz kombiniert graphbasierte Datenstrukturen mit Sprachmodellen, um narrative Zusammenhänge in komplex vernetzten Daten zu erkennen. Durch den Einsatz von Graph-RAG können nicht nur die bestehenden Verbindungen zwischen den Datensätzen besser verfolgt werden, sondern es eröffnet sich auch das Potenzial, bisher ungenutzte Beziehungsmuster zu identifizieren.

## Zielsetzung

Ziel ist es einen niedrigschwelligen, natürlich-sprachigen Sucheinstieg zu schaffen, der es erlaubt die Verknüpfungen innerhalb der Daten zu entdecken. Das Einbeziehen externer Wissensgraphen, bspw. der dbpedia, zur Beantwortung der Anfragen wäre eine optionale Anforderung.

## Technologien & Rahmenbedingungen

Die Daten liegen sowohl im bibliothekarischen Datenformat MARC21, als auch in RDF vor. Im Laufe des Jahres 2025 wird ein SPARQL Endpunkt zur Verfügung gestellt werden können.

## Weitere Informationen (beliebig viele Seiten)

GND Beispiel

<https://explore.gnd.network/gnd/11850553X>

