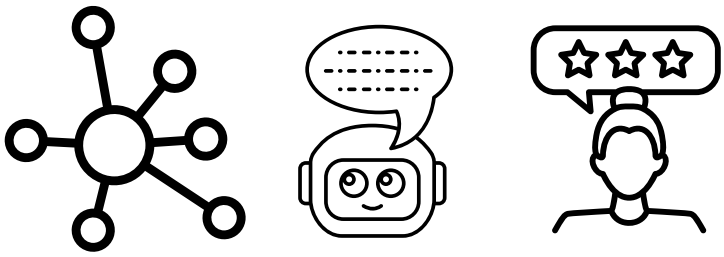
----- English version below -----

Ausschreibung Masterarbeit

Studie zur Evaluierung von LLM-basierten Text2SPARQL-Ansätzen und Vergleich mit DINa

DINa

Data Insight Navigator



Ausgangslage

Das LLM-basierte Assistenzsystem DINa wurde entwickelt, um natürlichsprachliche Anfragen automatisiert in SPARQL-Queries zu übersetzen und gezielt Daten aus Datenbanken abzurufen. Als Text2SPARQL-Methode (also Übersetzung von natürlichem Text in die Abfragesprache SPARQL) unterstützt DINa insbesondere datengetriebene Prozesse im ESG-Reporting (Nachhaltigkeitsberichte) sowie bei der Befüllung und Abfrage von Digitalen Produktpässen (DPPs).

Ziel des Systems ist es, zentrale Herausforderungen der nachhaltigen Transformation zu adressieren, insbesondere die Verfügbarkeit**,** Zugänglichkeit und Nutzbarkeit relevanter Nachhaltigkeitsdaten. DINa soll insbesondere kleine und mittlere Unternehmen (KMU) dabei unterstützen, künftige regulatorische Anforderungen an die Datenerhebung und -bereitstellung zu erfüllen.

Das System basiert auf zwei wesentlichen Komponenten:

1. der präzisen Auswahl des relevanten semantischen Kontexts, und
2. der darauf aufbauenden automatischen Generierung von SPARQL-Abfragen zur gezielten Extraktion von Nachhaltigkeits- und Produktinformationen.

Mehr Infos zu DINa auf: https://dina.tmdt.info/

Problemstellung

DINa ist dabei aber nicht der einzige Text2SPARQL Ansatz, der bisher entwickelt wurde. Um die Effektivität von DINa bewerten zu können, müssen die Ergebnisse unseres Verfahrens mit anderen

Ansprechpartner

Jakob Deich | **Tel.:** +49 202 439 1712 | **E-Mail:** deich@uni-wuppertal.de

verglichen werden. Bisher fehlt es allerdings an anwendungsnahen Datensätzen auf denen ein Vergleich möglich ist und die Stärken von DINa relevant werden.

Vorgehensweise und Erwartete Ergebnisse

Im Rahmen der Arbeit soll zunächst basierend auf einem geeigneten Datensatz ein Fragenkatalog mit Ground-Truth Antworten (Antworten die das System produzieren sollte) erstellt werden, welcher genutzt werden kann, um die verschiedenen Ansätze vergleichen zu können. Dazu werden etablierte oder neue Metriken benötigt, welche es ermöglichen die Antwortqualität zu messen. Daraufhin sollen geeignete, open-source Text2SPARQL-Ansätze identifiziert, implementiert und mithilfe des zuvor festgelegten Fragenkatalogs evaluiert werden. Abschließend können nun die Ergebnisse mit den Ergebnissen unseres Ansatzes verglichen werden.
Das erwartete Ergebnis ist eine Evaluation der verschiedenen Ansätze mit wissenschaftlichen Methoden. Die Arbeit leistet damit einen Beitrag zur realitätsnahen und zukunftsrelevanten Evaluation von Text2SPARQL-Systemen im Kontext Nachhaltigkeit.

Contact Person

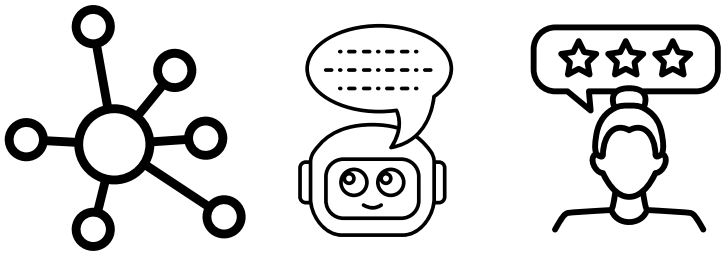
Jakob Deich | **Tel.:** +49 202 439 1712 | **E-Mail:** deich@uni-wuppertal.de

Master's thesis

Study on the evaluation of LLM-based Text2SPARQL approaches and comparison with DINa

DINa

Data Insight Navigator



Initial Situation

The LLM-based assistance system DINa was developed to automatically translate natural language queries into SPARQL queries and to retrieve specific data from databases. As a Text2SPARQL method (i.e. translation of natural text into the SPARQL query language), DINa supports data-driven processes in ESG reporting (sustainability reports) and in filling and querying digital product passports (DPPs) in particular.

The aim of the system is to address key challenges of sustainable transformation, in particular the availability, accessibility and usability of relevant sustainability data. DINa is intended to support small and medium-sized enterprises (SMEs) in particular in meeting future regulatory requirements for data collection and provision.

The system is based on two main components:

1. the precise selection of the relevant semantic context, and

2. the automatic generation of SPARQL queries for the targeted extraction of sustainability and product information.

More information about DINa at: https://dina.tmdt.info/

Contact Person

Jakob Deich | **Tel.:** +49 202 439 1712 | **E-Mail:** deich@uni-wuppertal.de

Problem Definition

However, DINa is not the only Text2SPARQL approach that has been developed so far. In order to evaluate the effectiveness of DINa, the results of our method must be compared with others. So far, however, there is a lack of application-oriented data sets on which a comparison is possible.

Methods and Expected Results

As part of the work, a question catalog with ground truth answers (correct answers that the system should produce) is to be created based on a suitable data set, which can be used to compare the different approaches. This requires established or new metrics that make it possible to measure the response quality. Suitable, open-source Text2SPARQL approaches should then be identified, implemented and evaluated using the previously defined questionnaire. Finally, the results can now be compared with the results of our approach.
The expected result is a scientific evaluation of the different approaches. The work thus makes a contribution to the realistic and future-oriented evaluation of Text2SPARQL systems in the context of sustainability.

Contact Person

Jakob Deich | **Tel.:** +49 202 439 1712 | **E-Mail:** deich@uni-wuppertal.de