

STELLAR

STAR-based Text-to-sEmantic LLM Argumentation Retrieval

Programm / Ausschreibung	Seal of Excellence, Seal of Excellence Ausschreibung 2026	Status	laufend
Projektstart	01.11.2026	Projektende	30.04.2028
Zeitraum	2026 - 2028	Projektlaufzeit	18 Monate
Barwert der Projektförderung	€ 150.000		
Keywords	Large Language Models; Semantic Data Modeling; Data Provenance; Digital Cultural Heritage		

Projektbeschreibung

In einer Zeit konkurrierender Wahrheitsansprüche und schwindenden Vertrauens in Informationen ist die Fähigkeit, Wissen in strukturierter, herkunftsnachweisbarer und transparenter Form darzustellen, dringender denn je. Dennoch beschränken sich ausgefeilte Argumentationsmodelle nach wie vor auf Demonstrationen in kleinem Maßstab: Ihre Befüllung erfordert mühsame manuelle Arbeit, während theoretisch unzureichende „Single-Truth“-Datenbanken nach wie vor die einzige praktikable Option für den groß angelegten Einsatz darstellen.

Dieses Projekt befasst sich mit diesem Engpass, indem es drei ausgereifte, aber bislang voneinander getrennte Technologien kombiniert: semantische Datenmodellierung, große Sprachmodelle (LLMs) und formale Argumentationsdarstellung. Aufbauend auf dem ERC-Projekt RELEVEN (2021–2026), das das STAR-Modell (STructured Assertion Record) zur Erfassung wissenschaftlicher Aussagen, Evidenzketten und konkurrierender Standpunkte entwickelt hat, werden wir LLM-Workflows entwerfen und testen, die komplexe Aussagestrukturen automatisch aus Primär- und Sekundärquellen extrahieren. Das Ergebnis wird ein Prototyp sein, der wissenschaftliche Publikationen aufnimmt und formale semantische Graphen ausgibt, die den dokumentierten ontologischen Zielmustern von RELEVEN entsprechen.

Das Durchbruchpotenzial ist zweifach: Erstens soll gezeigt werden, dass reichhaltiges, herkunftsgesichertes semantisches Wissen in großem Maßstab generiert werden kann; zweitens soll eine neue Fähigkeit geschaffen werden, ganze Argumente als Daten zu modellieren, wodurch konkurrierende wissenschaftliche Standpunkte rechnerisch handhabbar werden. Obwohl der Proof of Concept auf gut verstandenem historischem Material basiert – was strenge Basistests ermöglicht –, lässt sich die Methode auf den Journalismus, das Recht und jeden Bereich übertragen, der eine transparente, überprüfbare Argumentation aus Quellen erfordert. Open-Source-Tools und eine Dokumentation bewährter Verfahren werden die Wirkung auf breiterer Ebene maximieren.

Abstract

In an era of competing truth-claims and eroding trust in information, the ability to represent knowledge in structured, provenanced, and transparent forms has never been more urgent. Yet sophisticated argumentation models remain confined to small-scale demonstrations: populating them requires painstaking manual work, while theoretically inadequate "single truth" databases remain the only practical option at scale.

This project addresses that bottleneck by combining three mature but hitherto disconnected technologies: semantic data modelling, large language models (LLMs), and formal argumentation representation. Building on the ERC project RELEVEN (2021–2026), which developed the STAR (STructured Assertion Record) model for capturing scholarly assertions, evidence chains, and competing viewpoints, we will design and test LLM workflows that automatically extract complex assertion structures from primary and secondary sources. The outcome will be a prototype that ingests scholarly publications and outputs formal semantic graphs conforming to RELEVEN's documented ontological target patterns.

The breakthrough potential is twofold: first, demonstrating that rich, provenanced semantic knowledge can be generated at scale; second, establishing a new capacity to model whole arguments as data, making competing scholarly viewpoints computationally tractable. Though the proof of concept operates on well-understood historical material — enabling rigorous baseline testing — the method generalises to journalism, law, and any domain requiring transparent, auditable reasoning from sources. Open-source tools and best-practice documentation will maximise broader impact.

Projektpartner

- Universität Wien Institut für Geschichte