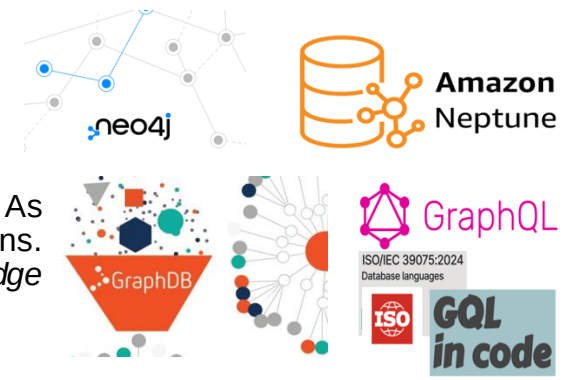
University of Stuttgart
Institute for Artificial Intelligence

Analytic Computing Group

Proseminar Sommer 2025

Introduction to Graph Database

Dr. Ratan Bahadur Thapa
ratan.thapa@ki.uni-stuttgart.de



Target Group

The seminar is designed for students interested in emerging *graph database* technologies and their applications. As part of the seminar, students will learn about, design, and implement graph databases for real-world applications. Although no prior experience in database design and implementation is required, basic experience with *knowledge graphs*, *relational databases*, or *data management courses* would be beneficial.

The Topic

In today's data-driven landscape, real-world applications are increasingly tasked with not only managing larger volumes of data but also deriving actionable insights from existing datasets. In this context, the relationships between data points become equally important as the stored individual data points. To fully exploit these data relationships, we need a database technology that treats relationship information also as a core entity along with individual data points. Graph databases provide such a capability. Traditional relational database management systems are suboptimal for handling data relationships due to their rigid schemas. In contrast, by design, graph databases offer a more efficient means of storing data relationships and provide greater flexibility in scaling data models or adapting to changing business requirements.

The evolution of graph databases began with the realization that traditional relational databases were not optimal for handling application scenarios that require processing complex relationships. Early systems, such as Network databases (1960s–1970s), were designed to manage hierarchical or networked data, but these were rigid and lacked flexibility. Later, *Object-oriented databases* (1980s) tried to overcome some of the limitations but still fell short for large-scale, connected data. The need for graph databases became more apparent with the rise of the internet, social networks, and big data analytics in the 2000s. As web data became more connected (for example, *social networks, supply chains*, *spread of viral infections, etc*), there was a growing need for a database model that could efficiently handle these complex connections. This led to the rise of modern graph databases such as *Neo4j, Amazon Neptune, IBM Graph, Dgraph*, *OrientDB, RedisGraph, TigerGraph, MemGraph, ArangoDB*, and more, which directly modeled relationships as first-class citizens in the database schema. To unify and streamline the efforts of existing graph databases and their query languages, the *Graph Query Language* (GQL) was released in April 2024 by ISO/IEC. GQL draws inspiration from existing query languages such as *Cypher, G-Core, PGQL, GSQL,* and others, and is the first newly ISO-certified standard database query language since the standardization of *Structured Query Language* (SQL) in 1987.

The seminar will be held in English and will provide an overview of existing graph database technologies, with a particular focus on their support for database schema design and query capabilities. Students will learn to model and implement graph database schemas and explore the core graph analytics features of graph query languages, particularly those associated with the widely adopted language *(Open)Cypher* and the newly issued ISO standard, GQL.

Procedure

Following introductory lectures on the topic, including a live demonstration of an example project, participants—either individually or as part of a small group—will design a graph database for their selected project (e.g*., ridesharing, food delivery applications, etc*) and write a technical report. The report will then be peer-reviewed by fellow participants. Subsequently, participants—either individually or as part of a small group—will deliver a 20-minute presentation of their project and submit their finalized project report as part of the final assessment.