



Contribution ID: 76

Type: **Poster**

Knowledge graph embedding based missing metadata prediction

Wednesday 26 February 2025 19:40 (20 minutes)

In scientific research, effective data management is crucial, especially when handling experimental data. The increasing volume and complexity of data collected in experimental settings necessitate rigorous methodologies to ensure that such data remains findable, accessible, interoperable, and reusable (FAIR). These requirements are seamlessly met by the RDF graphs, which is a type of knowledge graph. For example, Chair of Fluid Systems at TU Darmstadt developed a Metadata Database for the sensors based on a sensor information model. Physical properties like sensitivity, bias, measurement range, sensor actuation range, and other attributes such as identifier, manufacturer, and location are stored in RDF graphs.

However, metadata records accompanying legacy data may be incomplete for various reasons, such as adherence to outdated standards, omission of essential parameters, redactions for confidentiality, and errors. Consequently, measure 5 of the NFDI4Ing Task Area “Alex” initiative focuses on reconstructing incomplete metadata to ensure its continued utility. This issue has been a topic of discussion and research in the biomedical field for several years. Numerous methods, including natural language processing techniques like Named Entity Recognition, are being explored to extract metadata from document abstracts or titles. However, challenges remain. For instance, metadata may be dispersed across multiple documents, making it difficult to locate, and some metadata may not be recorded at all. Moreover, the semantic relationships between different data samples are overlooked. In response to these challenges and the growing trend of using RDF graphs to store metadata, we are employing knowledge graph embedding methods to predict the missing metadata.

I want to participate in the youngRSE prize

Primary author: CONG, Longwei

Presenter: CONG, Longwei

Session Classification: Poster and Demo Session together with Reception

Track Classification: Research Software: AI and ML in a research context