

Master thesis

Knowledge graph and ontology extraction for enterprise knowledge and data management

Initial situation and motivation

Knowledge management and the presentation and communication of technical concepts are highly relevant for APG in practice (onboarding, offboarding, cross-divisional collaboration, mutual understanding, technical models in the course of qualitative data management, etc.).

A lot of information is currently stored in various (text) sources (Confluence, Word documents, emails, PDFs, etc.).

Problem definition

It is therefore helpful if there are methods that can be used to extract the most important concepts/entities as well as their attributes and connections or relations to each other from text information in order to document them in graphical form

Methods of knowledge graph extraction or ontology extraction have existed for a long time, but there is currently no systematic analysis and comparison

With the emergence of large language models, the systematic extraction of semantic information from texts has experienced a further upswing, which should be explicitly considered as a methodological approach in the master's thesis in any case

The resulting graph should then form the basis for subsequent refinements and initial modelling in the data catalogue (technical data model)

Research question(s)

1. Which methods can be used to extract knowledge graphs or ontologies and how do they differ?
2. Which criteria are suitable for comparing the quality of such methods and how do the selected methods perform with regard to these criteria?
3. What improvements and new approaches result from the use of Large Language Models in this context?
4. To what extent are the graphs or ontologies extracted in this way of sufficient quality in practical use to support knowledge management in a business context?
5. What limitations and foreseeable further developments are foreseeable in this area?

Desired methodology

- Literature research
- Programming
- Benchmarking and comparison of relevant methods
- Prototype with basic functionality and
- Preparation of results in structured form for the APG data catalogue