

Masterarbeit

Knowledge graph and ontology extraction for enterprise knowledge and data management

Ausgangslage und Motivation

Wissensmanagement sowie die Darstellung und Vermitteln von fachlichen Konzepten sind in der Praxis für die APG hochrelevant (Onboard, Offboarding, übergreifende Zusammenarbeit, gegenseitiges Verständnis, fachliche Modelle im Zuge des qualitativen Datenmanagements etc.) Aktuell sind viele Informationen in diversen (Text-)Quellen verstreut gespeichert (Confluence, Word-Dokumente, Mails, PDFs etc.)

Problemstellung

Es ist daher unterstützend, wenn es Methoden gibt mit denen man aus Textinformationen die wichtigsten Konzepte/Entitäten sowie deren Attribute und Zusammenhänge bzw. Relationen zueinander extrahieren kann, um diese in graphischer Form aufbereitet zu dokumentieren

Methoden der Knowledge-Graph-Extraktion bzw. Ontologie-Extraktion existieren bereits lange, eine systematische Analyse und Gegenüberstellung gibt es aktuell nicht

Mit dem Aufkommen von Large Language Modells hat die systematische Extraktion von semantischer Information aus Texten weiteren Aufschwung erfahren, was in der Masterarbeit explizit als methodischer Ansatz jedenfalls Berücksichtigung finden soll

Der entstehende Graph soll dann Grundlage für nachfolgende Verfeinerungen sein sowie eine Erstmodellierung im Datenkatalog (Fachliches Datenmodell)

Forschungsfrage(n)

1. Welche Methoden kommen für die Extraktion von Wissensgraphen bzw. Ontologien in Frage und wie unterscheiden sich diese?
2. Welche Kriterien sind geeignet die Qualität solcher Methoden zu vergleichen und wie schneiden die selektierten Methoden bzgl. dieser Kriterien ab?
3. Welche Verbesserungen und neue Ansätze ergeben sich durch den Einsatz von Large Language Models in diesem Zusammenhang?
4. Inwiefern sind die so extrahierten Graphen bzw. Ontologien im praktischen Einsatz qualitativ gut genug, um das Wissensmanagement im unternehmerischen Kontext zu unterstützen?
5. Welche Limitationen und absehbare Weiterentwicklungen sind absehbar in diesem Gebiet?

Gewünschte Methodik

- Literaturrecherche
- Programmierung
- Benchmarking und Gegenüberstellung von relevanten Methoden
- Prototyp mit Basisfunktionalität sowie
- Ergebnisaufbereitung in strukturierter Form für den APG-Datenkatalog