

## FRAGEN BEANTWORTEN ÜBER WISSENSGRAPHEN (KGQA)



Der Zugang zum Wissen in einem Unternehmen kann sehr fragmentiert sein. Die Informationen können in Form von relationalen Datenbanken, Textnotizen, Webtabellen und so weiter vorliegen. Ein System zur Beantwortung von Fragen über Wissensgraphen kann einen einzigen Zugangspunkt zu allen relevanten Informationen bieten, die im Wissensgraphen eines Unternehmens gespeichert sind, ohne dass technische Kenntnisse (zum Beispiel SQL-Kenntnisse) erforderlich sind. KGQA-Systeme können das Auffinden von Informationen in großen Datenmengen erleichtern und effizienter gestalten.

### FÜR FOLGENDE HERAUSFORDERUNGEN

- Auffinden von Sachinformationen aus dem Wissensgraphen des Unternehmens
- Schneller Zugriff auf mehrere Informationsquellen (zum Beispiel Informationen über Kunden, Lieferungen, Personal, Teile und Serviceanfragen)
- Auffinden von Informationen über frühere Vorgänge (etwa: Wann wurde eine bestimmte Kundenmaschine gewartet, mit welchen Teilen, wer hat die Reparatur durchgeführt, wann wurde die Reparatur durchgeführt, welche Probleme sind aufgetreten)

### DER ANWENDUNGSFALL

- Ein Mitarbeiter hat einen Informationsbedarf, zum Beispiel „Wann wurde die Maschine XYZ zuletzt repariert?“
- Er gibt diese Frage in das System ein.
- Das System identifiziert die betreffende Maschine im Wissensgraphen des Unternehmens und sucht nach der Beziehung, die den Reparaturen an einer Maschine entspricht.
- Es findet fünf Reparaturen, die für die Maschine protokolliert wurden (die Beziehung im Wissensgraphen ist „serviced\_on“) und gibt die jüngste, nach Datum geordnet, zurück.
- Ähnliche Fragen können zu jeder Maschine aus dem Wissensgraphen des Unternehmens gestellt werden.
- Weitere Informationsbedürfnisse können ebenfalls beantwortet werden - zum Beispiel. „Wer hat die Reparatur an Maschine XYZ durchgeführt?“, „Welche Ersatzteile wurden für die Reparatur an Maschine XYZ verwendet?“
- Solange die relevanten Informationen im Wissensgraphen vorhanden sind, ist das System in der Lage zu lernen, wie man sie richtig beantwortet.

### DIE LÖSUNG IM DETAIL

- Für das KGQA-System wird ein trainiertes GNN-basiertes Modell verwendet, das in der Lage ist, aus dem Wissensgraphen Antworten auf die von dem Benutzer/der Benutzerin in natürlicher Sprache formulierten Fragen zu geben.
- Das System identifiziert zunächst die Entitäten in der Frage (auf der Grundlage eines für den Wissensgraphen des Unternehmens spezifischen Entitätsverknüpfungstools).
- Die identifizierten Entitäten werden verwendet, um die Beziehungen zu identifizieren, auf die sich die Frage bezieht.
- Bei komplexeren Fragen erfordert der Schritt der Identifizierung der Beziehungen mehrere Iterationen.
- Die vom System vorhergesagte Antwort wird dann dem Nutzer/der Nutzerin übermittelt.

### STATUS IM PROJEKT

Das aktuelle Prototypsystem beantwortet Fragen auf Englisch zu einem allgemeinen Wissensgraphen (Wikidata)<sup>1</sup>. Es erreicht eine Rate von 70 % richtiger Antworten bei einfachen Fragen.

### VORAUSSETZUNGEN

- Das System erfordert einen bestehenden Wissensgraphen, um zu funktionieren.
- Das System setzt ein vorhandenes Tool zur Verknüpfung von Entitäten voraus, das in der Lage ist, Namen von Entitäten zu erkennen, die spezifisch für den Wissensgraphen des Unternehmens sind).
- Das System kann nur Fragen zu Entitäten und Beziehungen beantworten, die Teil eines bestimmten Wissensgraphen sind.
- Das System kann nur faktoide Fragen beantworten - d.h. Fragen, die im Wissensgraphen eine präzise Antwort haben.
- Das System muss im Voraus trainiert werden; in der Regel ist seine Leistung umso besser, je mehr Fragen es sieht; je nach Größe des Wissensgraphen kann die Mindestanzahl von Beispielen zwischen 1.000 und 10.000 Fragen liegen.

<sup>1</sup> [https://www.wikidata.org/wiki/Wikidata:Main\\_Page](https://www.wikidata.org/wiki/Wikidata:Main_Page)



## VERFÜGBARKEIT

- Der Code für das Prototypsystem wird Ende 2022 als Open-Source-Paket über GitHub zur Verfügung gestellt.
- Die trainierten Modelle sind sprach- und domänenspezifisch - zum Beispiel wird ein Modell, das nur auf dem Allzweck-Wissensgraphen trainiert wurde, bei einem ganz anderen Wissensgraphen (etwa mit anderen Relationen und Entitätstypen) schlechter abschneiden.
- Modelle, die für die Englische Sprache trainiert wurden, sind für Input auf Deutsch nicht nutzbar. Diese Modelle müssen für die Deutsche Sprache erneut trainiert werden.

## SPEZIFIKATION

	Inputdaten	Preprocessing	Datenspeicher	Algorithmen	Interfaces
High-Level-Beschreibung	<p>Für das Training: Fragen in natürlicher Sprache + die entsprechende Antwort aus der KB</p> <p>Für die Anwendung: eine Frage in natürlicher Sprache</p>	Identifizierung von Entitäten aus dem Wissensgraphen in der Frage und ihrer zugehörigen Beziehungen (erfordert einen geeigneten Entity Linker)	<p>SPARQL-basierte Speicherung für den Wissensgraphen (z. B. Blazegraph)</p> <p>Platz für die Speicherung von Entity-Indizes</p>	GCNs (Graph Convolutional Networks) zum Finden der Antwort auf die Frage anhand der Hintergrunddaten; SentenceBERT für die Darstellung von Fragen/Entitäten	REST-API, Eingabe der Frage (JSON), Ausgabe von Antwort und Kontext (JSON),
Konfigurierbarkeit	Verbindungseinstellungen für Knowledge Graph, Entity Linker API	Fragen, Entitäten und zugehörige Beziehungen mit SentenceBERT darstellen	Idealerweise sollten die Daten über REST-API-Aufrufe verfügbar sein	Die Hyperparameter des Modells (Anzahl der GCN-Schichten, Größe der Schichten) können so angepasst werden, dass sie dem aktuellen Wissensgraphen am besten entsprechen	Die Antwort kann in ein einfaches Textformat oder in andere Formate umgewandelt werden.
Technische Umsetzung	Eingabe aus JSON-Dateien (Train) oder direkt aus einem Textfeld (Use)	<p>REST-API für die Verknüpfung von Entitäten</p> <p>REST-API für die Abfrage des Wissensgraphen nach Entitäten und Beziehungen</p>	Blazegraph für den Wissensgraphen eventuell SQLite für andere Indizes	Python, PyTorch, DGL, Zugang zu GPUs, die für Training und Nutzung erforderlich sind	Flask API
Spezifisches Beispiel aus dem Schnellboot	Unternehmensspezifischer Wissensgraph und Fragen zu dem darin gespeicherten Wissen	Die Benutzerdaten werden vorverarbeitet	Die Benutzerdaten werden mit geeigneten Speichermethoden gespeichert	Beantwortung von Fragen mithilfe des Wissensgraphen	Einfache Schnittstelle zur Eingabe der Frage und Auflistung der Antwort

