

Knowledge Graph Question Answering



Wie mit GPT aus natürlicher Sprache Datenabfragen werden von Fabian Liebig

1. Motivation

Stärken von W3C Knowledge Graphen:

- Semantische Verbindung von Daten.
- Konnektivität zu anderen Graphen.
- Dynamische Datenverarbeitung.
- Maschinelle Interpretierbarkeit.

Hierfür sind jedoch SPARQL Queries für die Abfrage der Informationen nötig, was für manche Nutzer*innen schwierig und zeitaufwändig sein kann.

Aber wie komme ich an die ID des schwersten Moleküls?

SPARQL 1.1 Abfrage

```
PREFIX chembl: <http://rdf.ebi.ac.uk/terms/chembl#>
PREFIX sio: <http://semanticscience.org/resource/>

SELECT ?chemblID ?property ?value
WHERE {
  ?molecule chembl:chemblID ?chemblID .
  ?molecule sio:SIO_000008 ?property .
  ?property sio:SIO_000300 ?value .
  FILTER(STRENGTHS(str(?property), "full_mwt"))
}
ORDER BY DESC(?value)
LIMIT 1
```



2. User Interface

Willkommen! Ich bin hier, um deine Fragen mit Hilfe von Wissensgraphen zu beantworten. Wie kann ich dir heute helfen?

Was kann ich dich fragen?

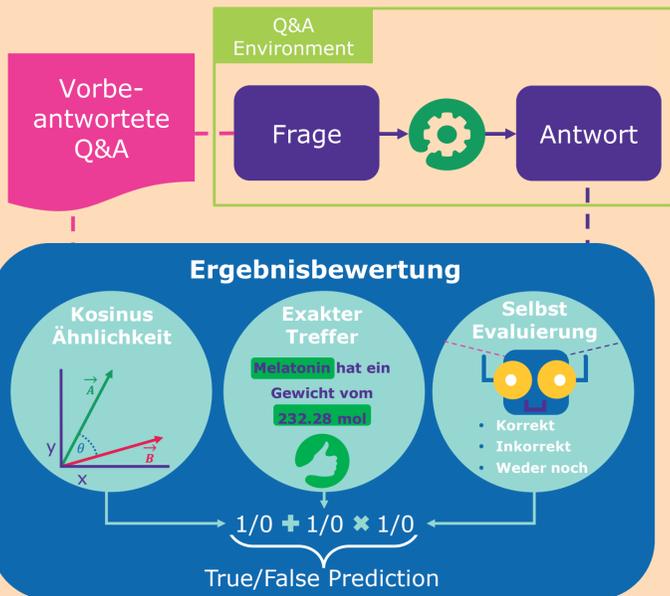
Sie können mich nach Informationen zu Bioaktivität, Wirkstoffen und Molekülen in ChEMBL oder nach Genexpressionsdaten und vergleichender Genomik aus Bgee fragen. Was möchten Sie wissen?

Erfüllt COc1ccc2[nH]cc(CCNC(C)=O)c2c1 die "Rule-of-Three"-Kriterien? Senden

3. Evaluierung

Ergebnismessungen zu vordefinierten Fragen und Antworten erlauben:

- Das Erkennen von Varianz zwischen Stichproben.
- Den Vergleich unterschiedlicher Herangehensweisen und Prompts.
- Eine Leitlinie während der Entwicklung.
- Der F1-Score gibt hierbei die Trefferquote und den Anteil der relevanten Instanzen unter den als relevant erkannten Instanzen an.

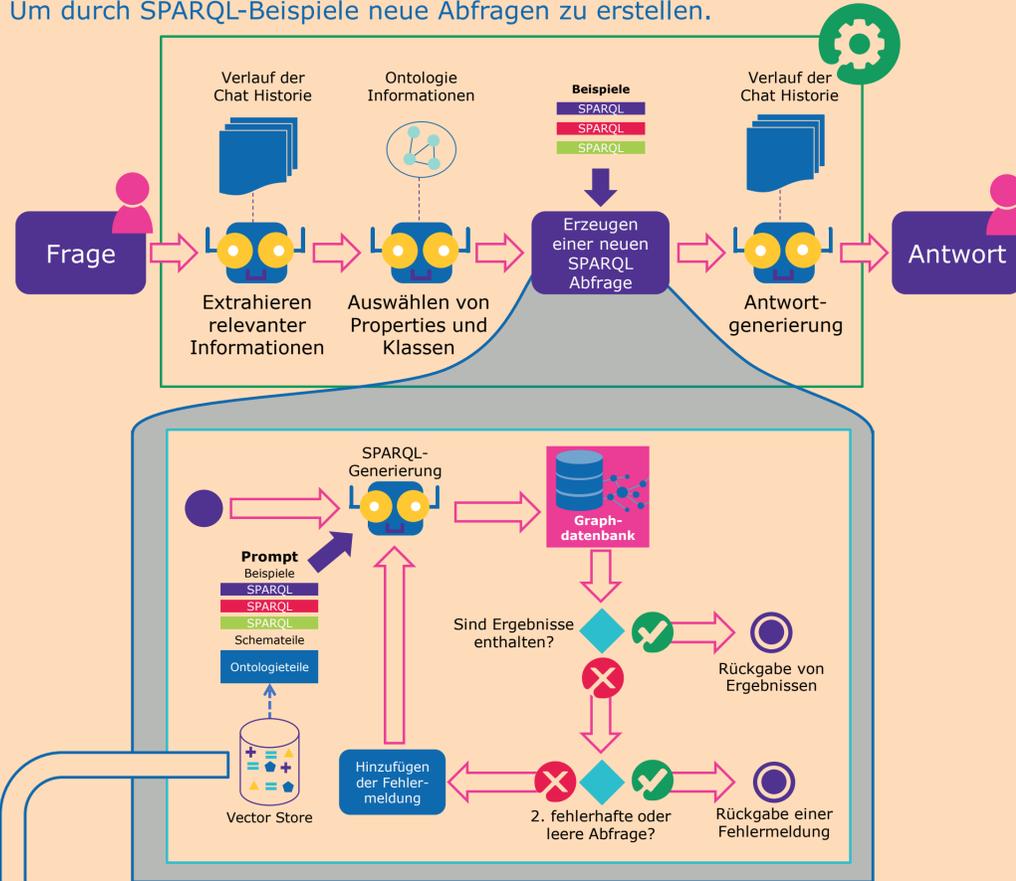


4. Programmstruktur

Der Generative Pre-Trained Transformer extrahiert zuerst Informationen aus:

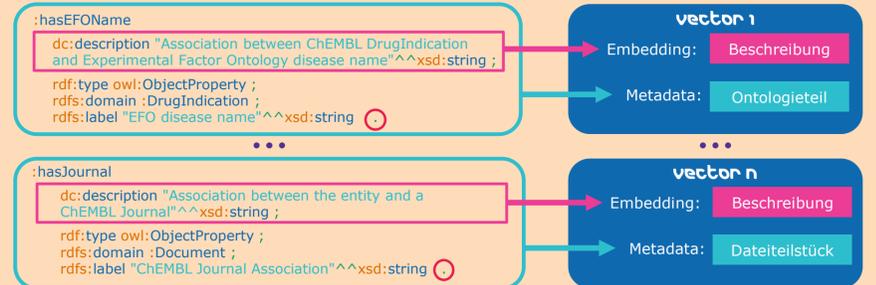
- Dem Chatverlauf.
- Graph Objektnamen, Klassen und Properties.

Um durch SPARQL-Beispiele neue Abfragen zu erstellen.



Semantische vektorisierung

- Ontologie-Komponenten in einen Vektorraum projizieren.
- Relevante Teile basierend auf der Frage identifizieren.
- SPARQL-Prompt mit den Ontologieteilen anreichern.



5. Ergebnis

- Optimierte SPARQL-Beispiele zu den gestellten Fragen werden gut Adaptiert.
- An „Halluzinationen“ angepasste Ontologien sind leichter nutzbar.
- Die GPTs profitieren unterschiedlich von Informationen im Prompt.

