Proposed Student Project: Role of Knowledge Graphs and Large Language Models in Improving Question Answering in Healthcare

Proposed Student Group Project

- Proposed student group project for the School of Information at the University of Texas – Austin
- AI in Health Class
  - Dr. Ying Ding
  - Project for group of 2-3 graduate level students
- Project Sponsors
  - Darrell Woelk, CTO, Green Room Technologies
  - Juan Sequeda, Head of AI Lab, data.world

data.world Knowledge Graphs and Large Language Models Benchmark Paper

https://arxiv.org/pdf/2311.07509.pdf

A BENCHMARK TO UNDERSTAND THE ROLE OF KNOWLEDGE GRAPHS ON LARGE LANGUAGE MODEL'S ACCURACY FOR QUESTION ANSWERING ON ENTERPRISE SQL DATABASES

TECHNICAL REPORT

Juan F. Sequeda
data.world
juan@data.world

Dean Allemang
data.world
dean.allemang@data.world

Bryon Jacob
data.world
bryon@data.world

November 14, 2023

ABSTRACT

Enterprise applications of Large Language Models (LLMs) hold promise for question answering on enterprise SQL databases. However, the extent to which LLMs can accurately respond to enterprise questions in such databases remains unclear, given the absence of suitable Text-to-SQL benchmarks tailored to enterprise settings. Additionally, the potential of Knowledge Graphs (KGs) to enhance LLM-based question answering by providing business context is not well understood. This study aims to evaluate the accuracy of LLM-powered question answering systems in the context of enterprise questions and SQL databases, while also exploring the role of knowledge graphs in improving accuracy. To achieve this, we introduce a benchmark comprising an enterprise SQL schema in the insurance domain, a range of enterprise queries encompassing reporting to metrics, and a contextual layer incorporating an ontology and mappings that define a knowledge graph. Our primary finding reveals that question answering using GPT-4, with zero-shot prompts directly on SQL databases, achieves an accuracy of 16%. Notably, this accuracy increases to 54% when questions are posed over a Knowledge Graph representation of the enterprise SQL database. Therefore, investing in Knowledge Graph provides higher accuracy for LLM powered question answering systems.

Keywords Knowledge Graphs · Large Language Models · Question Answering · SQL Databases · Benchmark · Retrieval Augmented Generation (RAG)

1 Introduction

Question answering, the ability to interact with data using natural language questions and obtaining accurate results, has been a long-standing challenge in computer science dating back to the 1960s[8, 9, 20, 10]. The field has advanced throughout the past decades [3, 22, 17], through Text-to-SQL approaches, as a means of facilitating chatting with the data that is stored in SQL databases[13, 12, 16, 1, 6, 19]. With the rise of Generative AI and Large Language Models (LLMs), the interest continues to increase. These question answering systems hold tremendous potential for transforming the way data-driven decision making is executed within enterprises.

While question answering systems have shown remarkable performance in several Text-to-SQL benchmarks [4, 2], such as Spider [18], WikiSQL[19], KaggleDBQA [11] their implications relating to enterprise SQL databases remain relatively obscure. [1] We argue that existing Question Answering and Text-to-SQL benchmarks, although valuable, are often misaligned with real-world enterprise settings:

1. these benchmarks typically overlook complex database schemas representing enterprise domains, which likely comprise hundreds of tables,

[1] A full survey on existing Text-to-SQL benchmarks is outside the scope of this work. We do believe that such survey would be beneficial for the community.

Modified Benchmark for Healthcare Information Domain

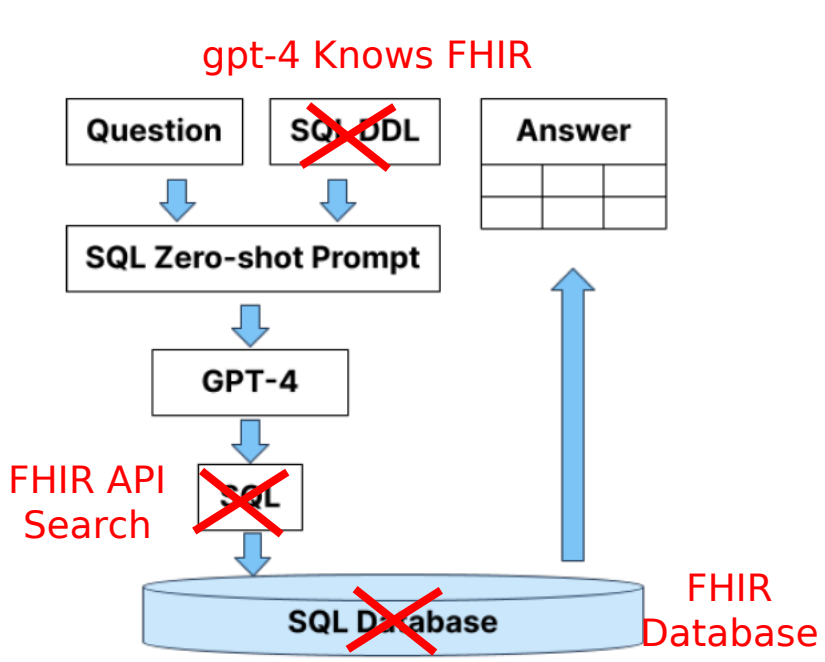


Figure 4: Question Answering System for SQL

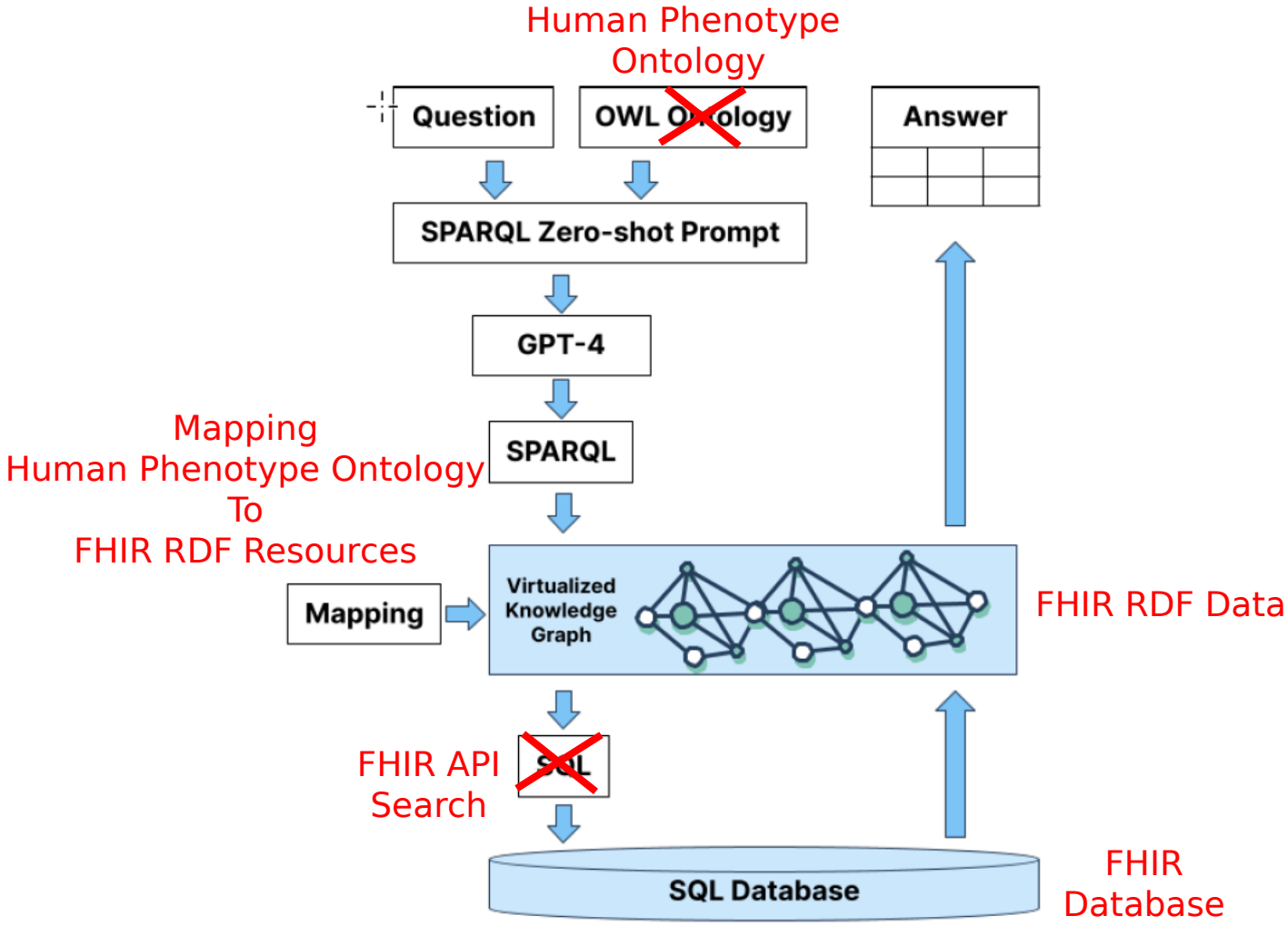
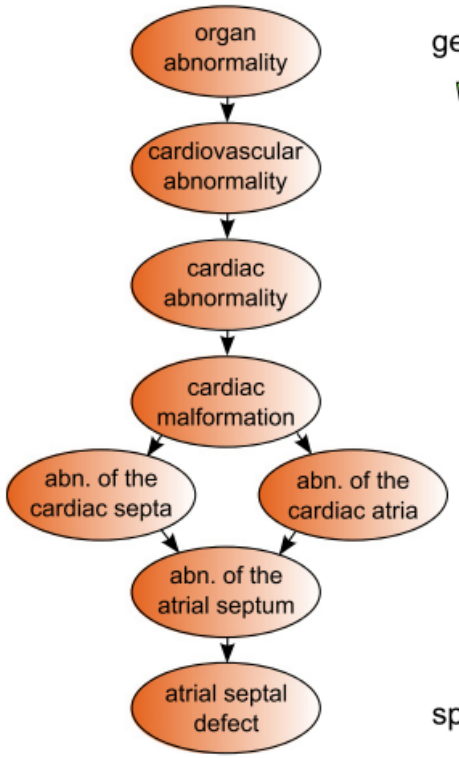


Figure 5: Question Answering System for Knowledge Graph

https://arxiv.org/pdf/2311.07509.pdf

GA4GH and Vulcan FHIR Accelerator Projects



general terms

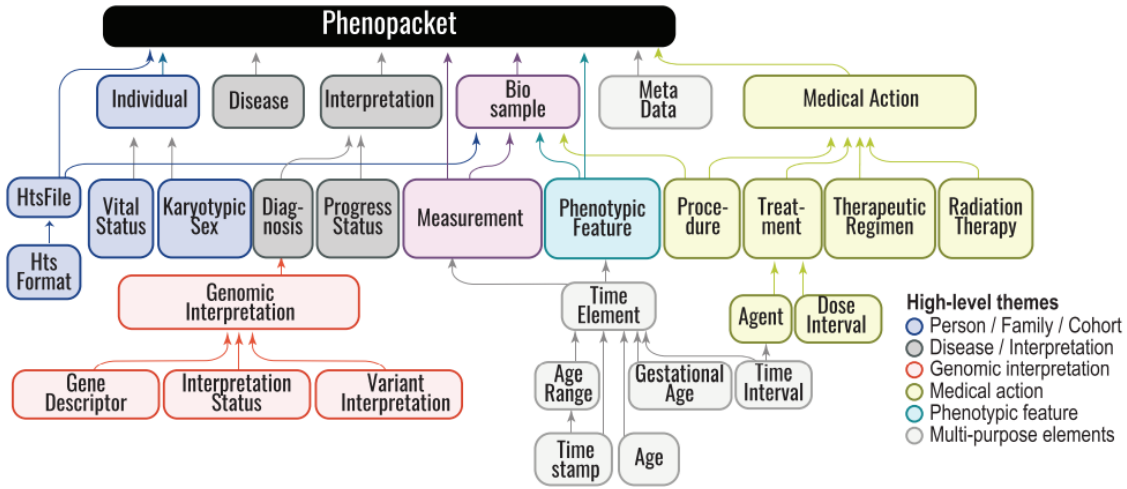
specific terms

Example from
Human Phenotype Ontology

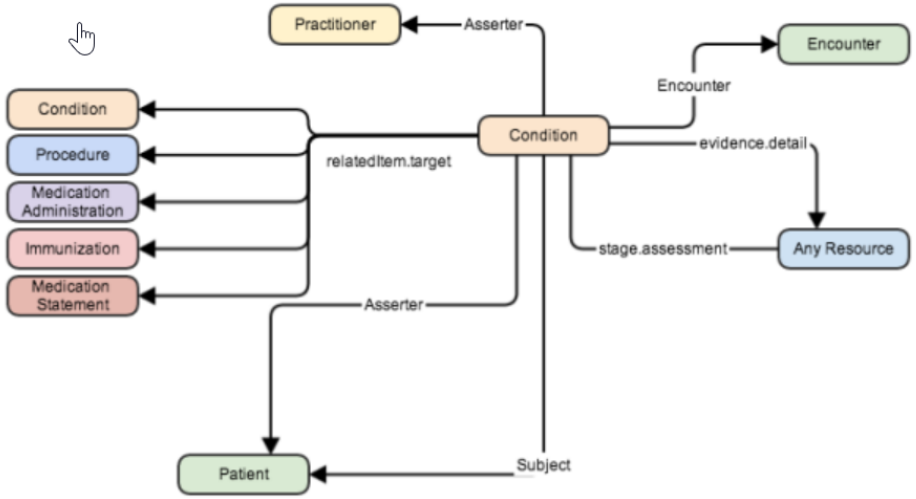
GA4GH Project

Proposed Student Project

GA4GH Phenopackets



Vulcan FHIR Accelerator Project



HL7 FHIR Resources

Questions and Issues

- Need a list of publications for students to learn about FHIR RDF
- Need a good (simple) example of the integration of an ontology with FHIR RDF data to get students started
- Need feedback on other similar research projects
- Need feedback on setting expectations for how much students can accomplish in one semester