

Data Querying using Natural Language

Problem

WHERE keyword LIKE '%cardiac arrest%'

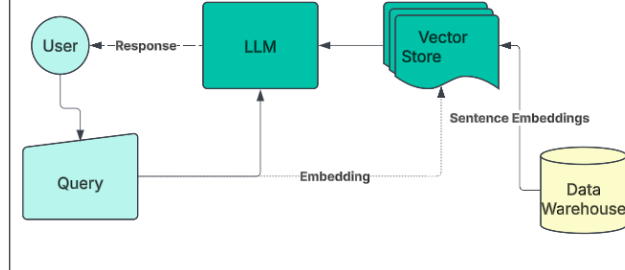
The current process for reporting on the health information linked to NIHR RDN portfolio is **time-consuming, labour-intensive**, and prone to **human error**. We run queries in SQL Server to retrieve information. This presents a word-for-word return based on exact string matches and loses records with similar meaning i.e. **semantic context** is lost.

Solution

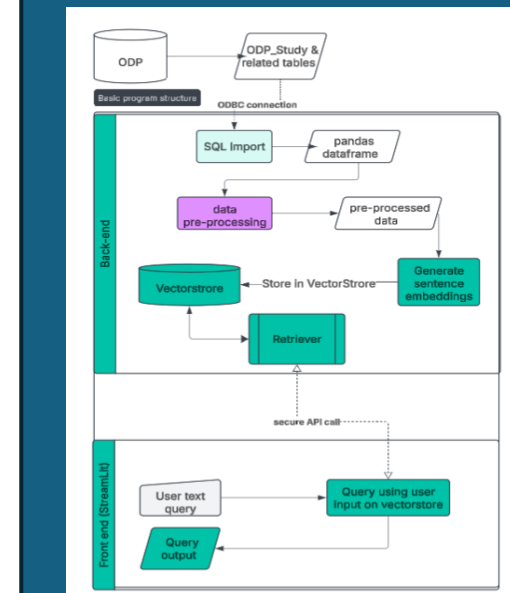
A 'Google' search for data warehouses, the semantic search and RAG tool allows users to input query related a healthcare research and get relevant information linked to the studies(part of NIHR RDN portfolio). This automated approach aims to significantly reduce the time and effort required for manual verification, ensuring more efficient and reliable outcomes.

[Github repository](#)

UML Diagram



Data Flow



Components



Sample output

NIHR: Semantic Search using RAG

Welcome to NIHR Research Semantic Search - Your repository for NIHR supported research across the UK

Please Note:

- For best results, ensure your prompts relevant health data.
- Please do not upload personally identifiable information.

We're here to help you discover healthcare research around the UK. Feel free to ask questions to inform results from the NIHR RDN.

Hello! Ask me about cancer research.

Summarise cancer research.

The given context provides details on two distinct research studies focusing on different aspects of cancer, specifically Non-small cell Lung, Colorectal, and Breast Cancer (Tumour Angiogenesis) and Colorectal Cancer (Molecular pathology).

In the first study titled "Tumour Angiogenesis," researchers investigate the correlation between tumor angiogenesis in those types of cancer, radiology-pathology, and prognostic correlation. The study is observational, encompassing primary cancers, and conducted within tertiary care settings with a managing specialty of Cancer, specifically Colorectal Cancer.

Outcomes

- Democratization of data analysis
- More exhaustive results
- Results summarisation
- Data Exploratory Analysis
- Operational and performance efficiencies
- Curated Portfolios

Author: Abhinav Jindal (Jin)

abhinav.jindal912@gmail.com | www.linkedin.com/in/abhinav-jindal/

Note of thanks:

Athanasios Anastasiou, NIHR CCF Sr. Data Scientist

Mark Sainsbury, NIHR NETSCC Data Science Lead

Sammi Rosser, HSMA Co-lead

Alex Barnard, NIHR RDN Head of Data & Analytics & others