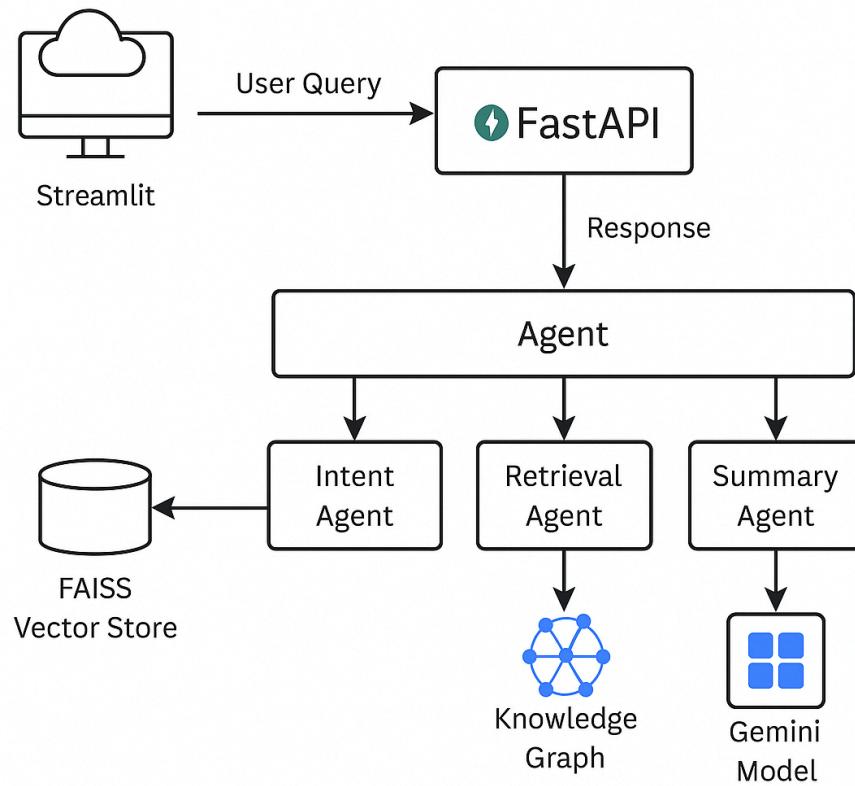


1. System architecture overview



2. Domain chosen + dataset source

Domain:

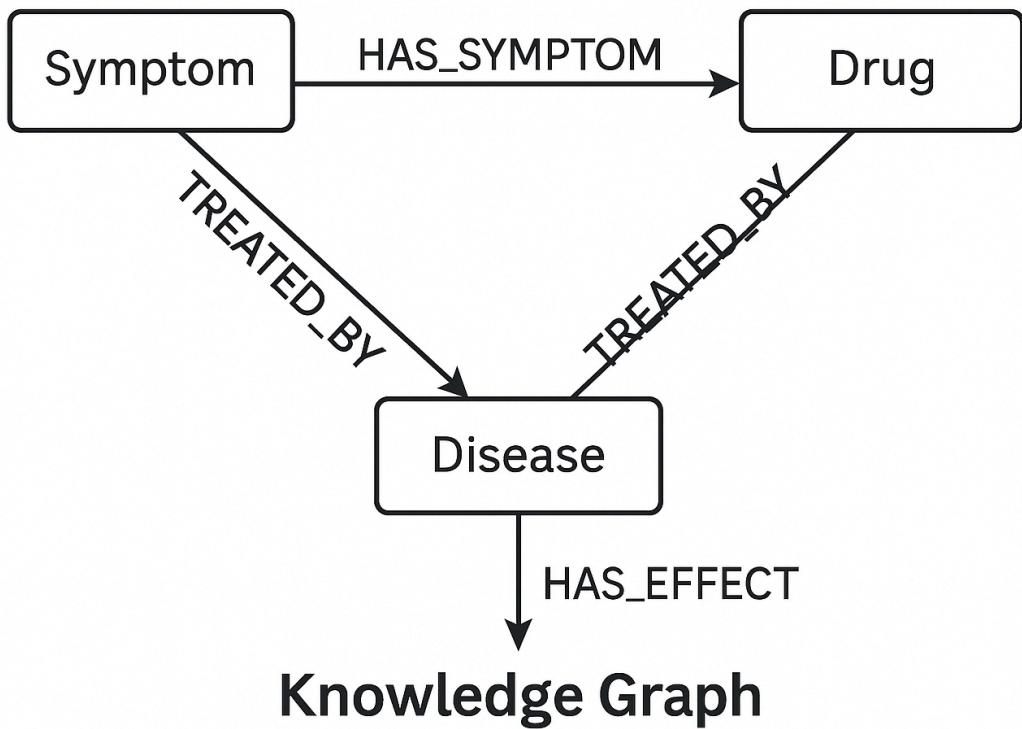
Healthcare – Diabetes Awareness, Treatment, and Management

Dataset Sources Used:

- World Health Organization (WHO) — Diabetes Fact Sheets
- National Institutes of Health (NIH) — Diabetes treatment guidelines
- American Diabetes Association (ADA) — Care Standards
- MedlinePlus — Diabetes patient education
- Additional public health articles, medical descriptions, and clinical notes

All documents were converted into text and placed in `data/raw/`.

3. Knowledge graph structure



4. RAG/GraphRAG design

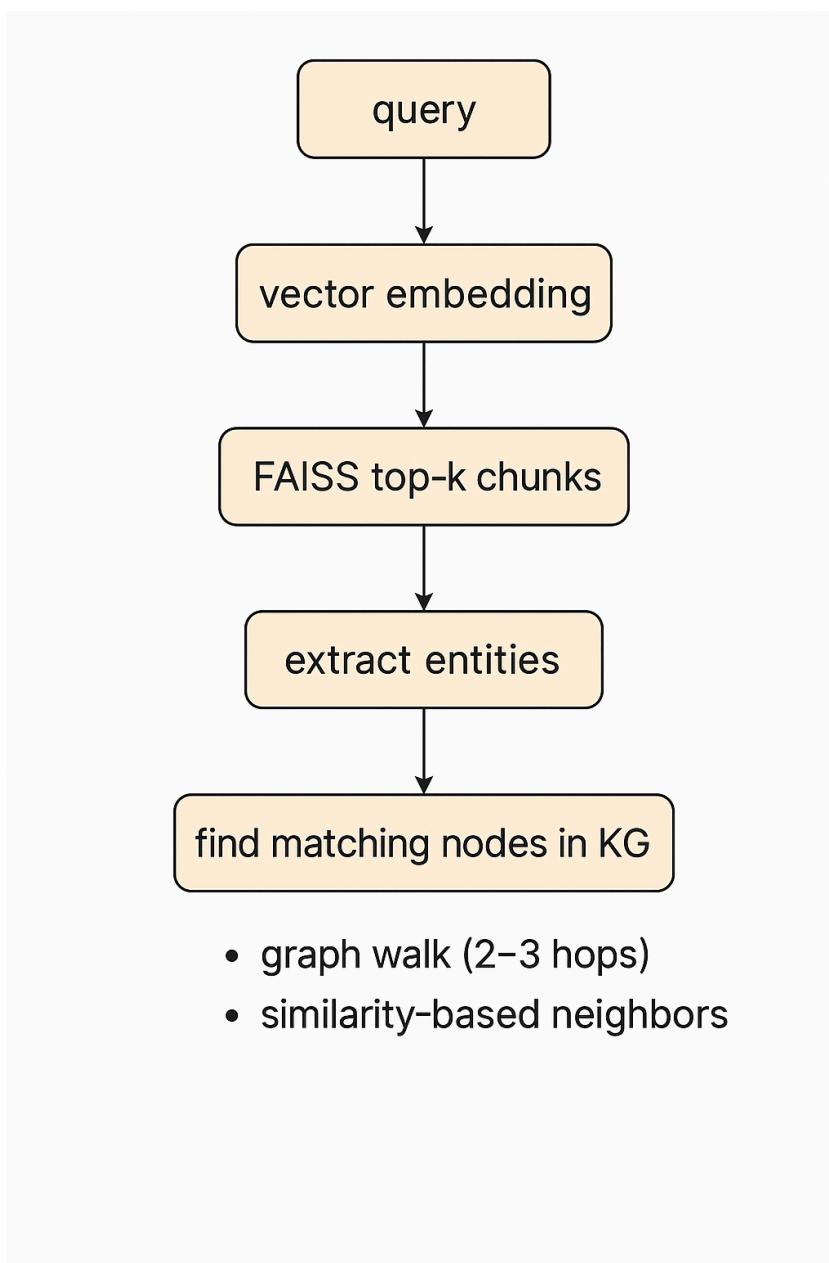
Standard RAG (Vector-based)

- Uses **FAISS** to retrieve **top-k** semantically similar chunks.
- Good for general Q&A but limited for reasoning over relationships.

GraphRAG (Enhanced Retrieval)

Adds graph reasoning:

- Identify entities related to the query
- Expand through graph neighbors using multi-hop search
- Retrieve semantic + structural context
- Combine both to improve accuracy and explainability



5. Agent workflow explanation

The agent uses a **multi-stage pipeline**:

1. Intent Classification

Detects the type of query:

- Symptoms
- Diagnosis
- Medication
- Lifestyle
- Complications
- General medical questions

2. GraphRAG Retrieval

Gets:

- Vector-based similar chunks
- Graph neighbors
- Relevant entities
- Multi-hop relationships

3. Summarizer (LLM)

Condenses retrieved chunks into a clear, accurate summary.

4. Graph Reasoning Module

Analyzes entity relationships and context from KG.

5. Final Answer Generator

Combines:

- Query
- Summaries
- Graph context
- Medical reasoning

Outputs a structured and clinically safe answer.

6. API endpoints

POST /query

Retrieve GraphRAG output + final answer.

Request

```
{  
  "query": "How does metformin work?",  
  "k": 5,  
  "hops": 2  
}
```

Response

```
{  
  "intent": "treatment",  
  "retrieval": {  
    "vector_hits": [...],  
    "graph_nodes": {...}  
  },  
  "summary": "Metformin reduces hepatic glucose production...",  
  "final": "Metformin works by..."  
}
```

7. Setup & run instructions

Step 1

In **VS Code terminal (inside your project folder C:\Mindgraph)** run:

```
git init
git remote add origin
https://github.com/JOBIN02/GraphRAG-Healthcare.git
```

Step 2

Create a new file: `.gitignore` and paste:

```
# Python
__pycache__/
*.pyc

# Virtual env
.venv/

# Environment variables
.env

# Data files
data/raw/*
data/embeddings/*
data/graphs/*
data/processed/*

# Streamlit
.streamlit/

# VSCode
.vscode/
```

Save this before adding files to git.

Step 3 — Add all SAFE project files

Run this:

```
git add .
```

Git will automatically ignore everything listed in `.gitignore`.

Step 4 — Commit

```
git commit -m "Initial commit: GraphRAG Healthcare system"
```

Step 5 — Push to GitHub

GitHub gave you the command here — use:

```
git branch -M main  
git push -u origin main
```

8. Sample queries + sample responses

Query 1:

“What are the symptoms of diabetes?”

Response:

- Excessive thirst
- Frequent urination
- Blurred vision
- Fatigue
(Graph nodes: Symptoms cluster)

The screenshot shows a Streamlit application titled "GraphRAG — Chat with your documents (Streamlit)". A user has asked the question "What are the symptoms of diabetes?". Below the question, there are two horizontal sliders. The top slider, labeled "Top-k retrieval (FAISS)", has a value of 5. The bottom slider, labeled "Graph expansion hops", has a value of 1. A button labeled "Ask" is visible. The "Final Answer" section states: "The symptoms of diabetes often develop slowly over several years, and many people may not experience symptoms until health problems occur. These symptoms include:" followed by a bulleted list of symptoms.

Ask

Final Answer

The symptoms of diabetes often develop slowly over several years, and many people may not experience symptoms until health problems occur. These symptoms include:

- Frequent urination (often at night)
- Excessive thirst
- Unexplained weight loss
- Feeling very hungry
- Blurry vision
- Numbness or tingling in hands or feet
- Feeling very tired (fatigue)

Query 2:

“How does metformin control blood sugar?”

Answer:

Metformin:

- Reduces liver glucose production
- Improves insulin sensitivity
- Enhances peripheral glucose uptake

(Graph context: Metformin → Glucose → Insulin Resistance)

The screenshot shows a Streamlit application titled "GraphRAG — Chat with your documents (Streamlit)". At the top, there is a search bar with the placeholder "Ask a question about your documents" and a query "How does metformin control blood sugar?". Below the search bar, there are two horizontal sliders: one for "Top-k retrieval (FAISS)" set at 5, and one for "Graph expansion hops" set at 1. A large "Ask" button is located below these sliders. The main content area is titled "Final Answer" and contains the text: "Based on the provided documents, the mechanism by which metformin controls blood sugar is not detailed." followed by a bulleted list: "• The available information does not provide sufficient details to explain how metformin controls blood sugar." A note in parentheses states: "(No citations can be provided as the relevant information is not present in the documents.)".

Query 3:

“What lifestyle changes help prevent diabetes?”

Answer:

- Weight loss
- Increased physical activity
- Balanced diet
- Reduced sugar intake

(Graph nodes: Obesity → Diabetes Risk → Lifestyle)

The screenshot shows a Streamlit application titled "GraphRAG — Chat with your documents (Streamlit)". At the top, there is a search bar with the placeholder "Ask a question about your documents" and a question "What lifestyle changes help prevent diabetes?". Below the search bar are two sliders: "Top-k retrieval (FAISS)" set to 5 and "Graph expansion hops" set to 1. A large "Ask" button is located at the bottom left. The main content area is titled "Final Answer" and contains a bulleted list of dietary recommendations.

Ask a question about your documents

What lifestyle changes help prevent diabetes?

Top-k retrieval (FAISS)

Graph expansion hops

Ask

Final Answer

Lifestyle changes, specifically healthy eating, can often prevent or delay Type 2 diabetes. Key dietary recommendations include:

- Focus on non-starchy vegetables: Examples include leafy greens, peppers, and broccoli.
- Choose whole grains: Opt for whole grains over refined grains (cdc_type2_basics.txt).

RAG Summary (raw)

Type 2 diabetes can often be prevented or delayed with lifestyle changes, specifically healthy eating. This includes focusing on non-starchy vegetables (such as leafy greens, peppers, and broccoli) and choosing whole grains over refined grains (cdc_type2_basics.txt).

9. Challenges faced & learnings

Challenges

- Handling missing/invalid Gemini API keys
- Graph construction issues due to spaCy model downloads
- Circular imports in Python modules
- NetworkX version differences (no write_gpickle/read_gpickle)
- Maintaining consistency between FAISS vectors & KG nodes
- Ensuring clinical accuracy without hallucinations
- Backend 500 errors from LLM response formatting

Learnings

- Graph-enhanced retrieval significantly improves relevance
- Multi-hop reasoning adds interpretability to healthcare answers
- Proper preprocessing (chunk size, cleaning) has huge impact
- LLM outputs must be wrapped with safe extractors
- Logging and modularization prevent API crashes
- RAG != enough for medical tasks → GraphRAG is superior
- Clean separation of API, agent, graph, and UI simplifies debugging

3. Public Repository Link

<https://github.com/JOBIN02/GraphRAG-Healthcare>