

Towards Agentic Graph RAG

Combining the Power of Structured Outputs from LLMs with Graph Databases

Prashanth Rao

Al Engineer, Kùzu Inc. 💽

kuzudb.com

The Al Conference
San Francisco | 17 Sep 2025

RAG bridges knowledge gaps



- Intertwined with the use of agents
- Most agent workflows are multi-step (observe, reflect, act)
- RAG is:
 - The foundation of memory (supports the agent's memory system)
 - An extension of the model's knowledge with external facts
 - An important source of context at each stage
- There are many flavours of RAG (not just chat bots or Q/A)

For autonomous retrieval systems, context is king





Knowledge is the *skeleton*Agents are the *muscles* that decide how to move.

Knowledge graphs can be the bones (backbone?) of such systems.

How to make messy, unstructured data queryable?



"How many patients in Massachusetts are allergic to shellfish?"

Ms. Marisol Rodriguez, who was born on December 1, 1995, and resides at her home at 547 Effertz Extension Unit 28, East Longmeadow, Massachusetts, 00000, is never married. She identifies Spanish as her primary language. She can be reached at her home phone number, which is 555-808-5474.

In the past, Marisol visited the ENCOMPASS HEALTH REHAB HOSPITAL OF WESTERN MASS on December 13, 2013, from 9:15:24 until 9:30:24 in Central European Standard Time for a postnatal visit. During this visit, Dr. Cletus Paucek, a primary performer, was the practitioner managing her care. You may reach out to Dr. Paucek at his work email, which is Cletus494.Paucek7550example.com.

Of note, Marisol has a confirmed low criticality allergy towards shellfish, first noted on June 2, 1997, at 22:15:24 in Central Eastern Standard Time. Reactions to an exposure to shellfish in her case have varied. In some instances, she experiences a moderate eruption of skin. In severe cases, she has experienced anaphylaxis. More often, her symptoms are mild and include itching and coughing. Another noted allergy intolerance is currently active, yet the clinical details are not specified here.

```
l,
  "prefix": "Ms."
},
  "age": null,
  "genden": null,
  "birthDate": "1995-12-01",
  "address": {
    "line": "547 Effertz Extension Unit 28"
    "city": "East Longmeadow",
    "state": "Massachusetts",
    "postalCode": "00000",
    "country": "US"
```

"phone": "555-808-5474".

"maritalStatus": "NeverMarried",

"primaryLanguage": "Spanish"

"email": null,

"family": "Rodríquez",

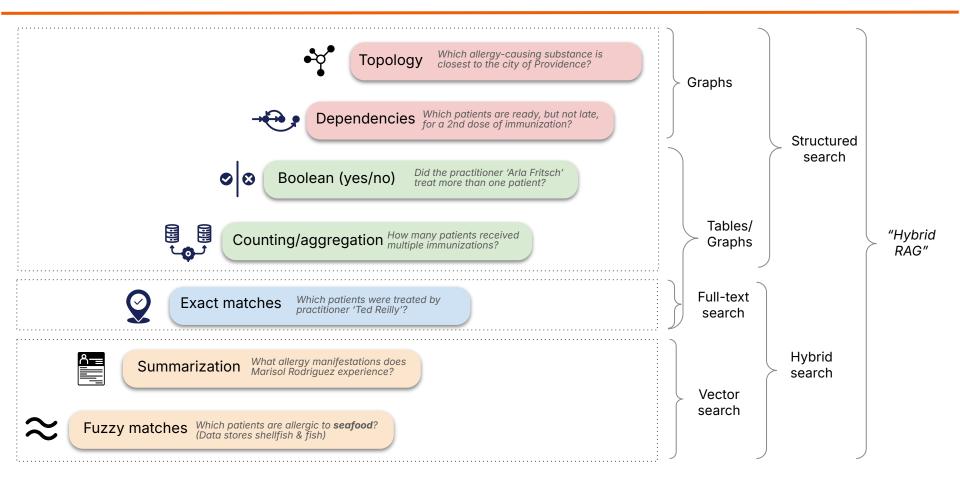
"name": {

"qiven": [

"Marisol"

The best retrieval tool depends on the question





Taking knowledge graphs to production



Part 1: Graph construction 🚜

- Schema definition
- Structured outputs: Extract nodes, relationships and properties
- Entity deduplication: Identify and handle duplicate entities

Part 2: Retrieval and RAG

- Graph RAG: Use the knowledge graph as the retrieval layer (structured facts)
- Vector-based RAG: Use semantic similarity between the search query and the data to retrieve chunks
- Hybrid RAG: Combining various retrieval results via reranking



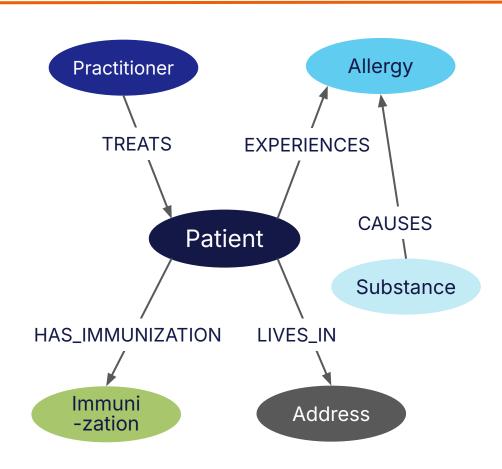
Part 1: Information Extraction

Information extraction pipeline



Use case: Patient questions on unstructured patient notes

- Does using Graph RAG improve our ability to answer questions?
- The data lacks structure to start with
- Knowledge graphs help impose some degree of structure, baking in domain knowledge
- Paths are a first-class citizen



From unstructured → structured data



Frameworks like BAML, DSPy (and many others) allow us to generate **high-quality** structured outputs from LLMs

```
from pydantic import BaseModel
class PersonNameAndTitle(BaseModel):
    family: str | None = Field(default=None, description="surname")
   given: list[str] | None = Field(
        default=None,
        description="Given names (first and middle names)".
   prefix: str | None = Field(
        default=None, description="title of the person: Dr.
                                                            import dspy
                                                            class PatientInfo(dspv.Signature):
class Address(BaseModel):
   line: str | None = Field(default=None, alias="street")
                                                                Extract patient information from the given note.
   city: str | None
                                                                - Do not infer any information not explicitly mentioned.
   state: str | None
                                                                - If you are unsure about any field, leave it as None.
   postalCode: str | None
   country: Literal["US"] | None
class Patient(BaseModel):
                                                                note: PatientNote = dspv.InputField()
   record id: int | None = Field(default=None)
                                                                patient: Patient = dspv.OutputField()
   name: PersonNameAndTitle | None
   age: int | None
   gender: Literal["male", "female"] | None
   birthDate: str | None = Field(default=None, description="ISO-8601 format for date")
   phone: str | None = Field(default=None, description="Phone number of the patient")
   email: str | None = Field(default=None, description="Email address of the patient")
   maritalStatus: Literal["Married", "Divorced", "Widowed", "NeverMarried"] | None
   address: Address | None = Field(
        default=None, description="Residence address of the patient"
```

```
"name": {
 "family": "Rodríquez",
 "given": [
   "Marisol"
  "prefix": "Ms."
"age": null.
"gender": null,
"birthDate": "1995-12-01",
"address": {
  "line": "547 Effertz Extension Unit 28",
  "citv": "East Longmeadow",
  "state": "Massachusetts",
  "postalCode": "00000",
  "country": "US"
"phone": "555-808-5474",
"email": null,
"maritalStatus": "NeverMarried",
"primaryLanguage": "Spanish"
```

Nested JSON is a graph in disguise!



```
"allergy": {
                                        "substance":
                                             "category": "food",
                                             "name": "shellfish",
                                             "manifestation": [
                                               "moderate eruption of skin",
                                               "anaphylaxis",
 "name": {
                                               "itching",
"family": "Rodríguez",
                                               "coughing"
 "qiven": [
   "Marisol"
  "prefix": "Ms."
"age": null,
"gender": null,
"birthDate": "1995-12-01",
"address": {
  "line": "547 Effertz Extension Unit 28".
  "city": "East Longmeadow",
  "state": "Massachusetts",
  "postalCode": "00000",
                                       "practitioner": {
  "country": "US"
                                          "name": {
                                            "family": "Paucek",
"phone": "555-808-5474",
                                            "qiven": [
"email": null,
                                              "Cletus"
"maritalStatus": "NeverMarried",
"primaryLanguage": "Spanish"
                                            "prefix": "Dr."
                                         "address": null,
                                         "phone": null,
                                         "email": "Cletus494.Paucek755@example.com"
```

How accurate is LM-based extraction in practice?





Hugging Face

kishanbodybrain/test-fhir

- Synthetic dataset available on Hugging Face Datasets
- 2,726 human-annotated records in FHIR format
- Easy to evaluate!

Exact match accuracy

Model	BAML	DSPy
google/gemma-3-27b-it	91%	94%
openai/gpt-4.1-mini	95%	94%
google/gemini-2.5-flash	95%	96%

openai/gpt-4.1	1	1
google/gemini-2.5-pro	1	1

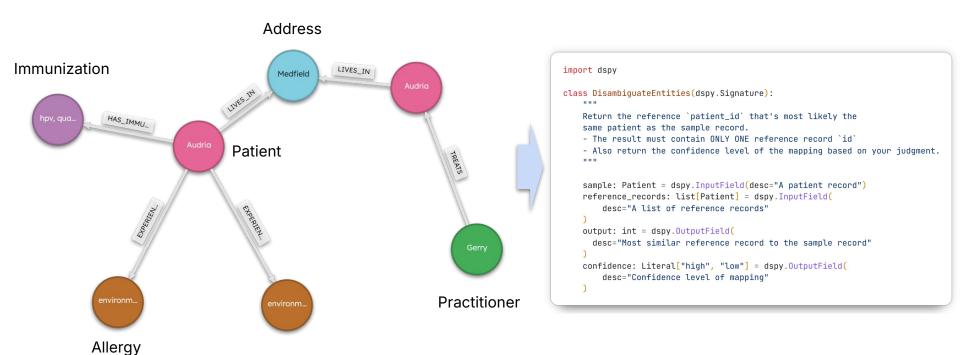
General-purpose models continue to improve (as does constrained generation)

github.com/prrao87/structured-outputs

Handling duplicates upon extraction



"Audria Schinner, lives in Medfield, Massachusetts, treated by Gerry Block"



blog.kuzudb.com: A gentle introduction to DSPy for graph data enrichment



Part 2: Agentic RAG

Graph RAG is inherently agentic



Graph-based retrievers eventually become a suite of tools/techniques

Route the query to the appropriate tool
E.g. text2cypher or templated
Cypher with parameters

Locate semantically relevant entry points

Fuzzy searches (based on embedding) can help narrow down on the right entry points to begin traversing paths



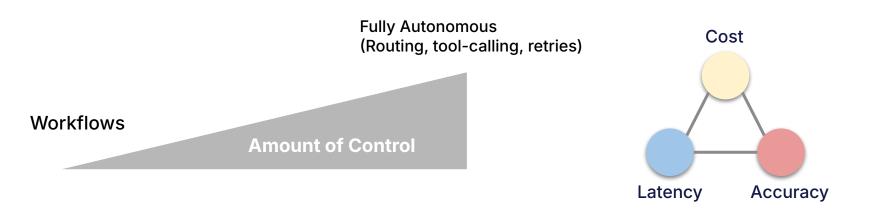
Error handling and recovery
E.g. incorrect Cypher syntax or query
terms generate by the LLM

Reranking results using graphs

E.g. prioritized results based on centrality

Should everything be fully agentic?



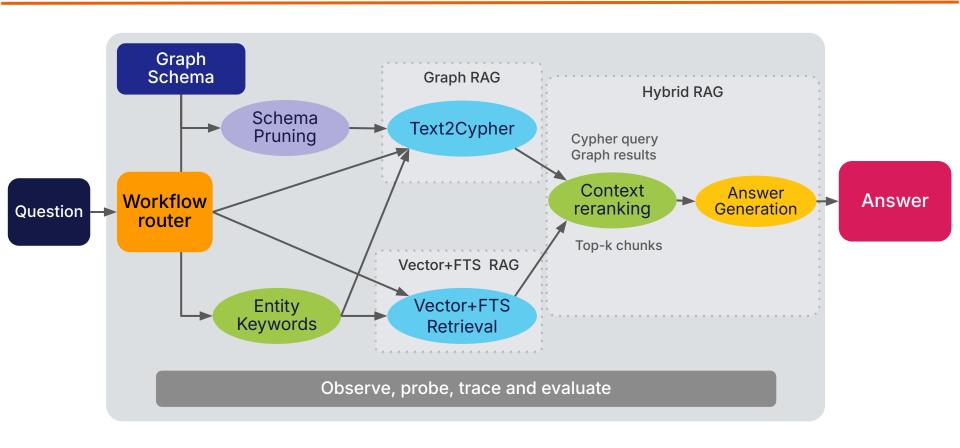


Sliding scale of autonomy:

- Start with a workflow, increase the level of autonomy gradually
- Add orchestration, adaptive logic and more elaborate routing logic down the line
- Evals, evals, more evals (In DSPy, they help build self-improving systems)

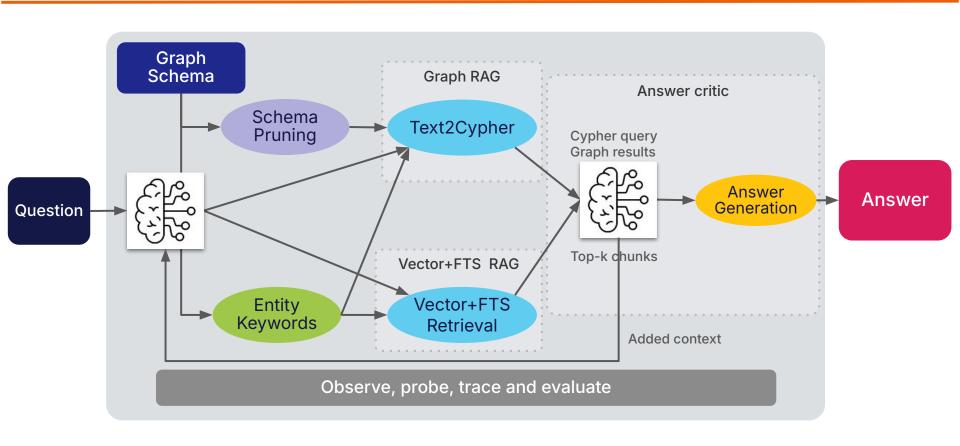
Combining retrievers for RAG (1)





Combining retrievers for RAG (2)





How graphs can help with retrieval: Example



Is the patient ID 45 allergic to the substance 'shellfish'? If so, Yes/No + what city and state do they live in, and what is the full name of Follow-up the practitioner who treated them? How many substances cause allergies in the category 'food'? Counting + Exact match What are the unique categories of substances patients are Counting +

Yes, patient ID 45 is allergic to shellfish. They live in East Longmeadow. Massachusetts, and were treated by a practitioner named Cletus Paucek.

There are 13 distinct substances that cause allergies in the 'food' category.

The unique categories of substances patients are allergic to are food, other,

Did the practitioner 'Arla Fritsch' treat more than one patient?

Yes, the practitioner 'Arla Fritsch' treated more than one patient. The results show that they treated 4 patients.

The full names of the patients treated by the practitioner named Josef Klein

Gabrielle Claudie Medhurst, Shaunte Adeline Batz, Jovan Rolanda Ondricka,

are: Lili Abbie Brekke, Marinda Lindsay Veum, Gary Everette Abshire,

What are the full names of the patients treated by the practitioner named Josef Klein? **Exact Match**

allergic to?

Follow-up

Exact match +

Yes/No

Counting

Counting

Counting

Jerrell Weston Graham, and Della Legros. There are 184 patients born between 1990 and 2000.

How many patients were born in between the years 1990 and Counting 2000?

How many patients have been immunized after January 1, 2022?

The number of patients who have been immunized after January 1, 2022 is 65.

Which practitioner treated the most patients? Return their full name and how many patients they treated.

The practitioner who treated the most patients is Ted Reilly, with a total of 19 patients.

How many patients are immunized for influenza?

Based on the context, 204 patients are immunized for influenza.

medication, and environment.

There is one patient named Rodrigo Ernie Rosenbaum who received multiple immunizations. On June 12, 2015, he received an injectable, preservative-free seasonal Influenza vaccine and a meningococcal MCV4P vaccine.

How many patients with the last name 'Rosenbaum' received Counting multiple immunizations?

How good are LLMs at Text2Cypher in practice?





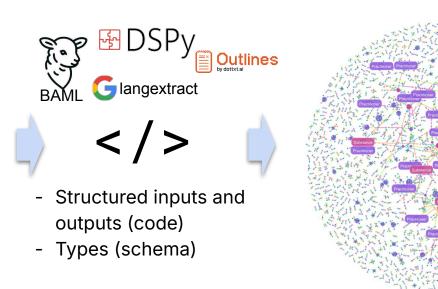
github.com/kuzudb/text2cypher

blog.kuzudb.com: Improving Text2Cypher for Graph RAG via schema pruning

Summary



[f"record id":1,"note":"Clinical Note:\n\n- Patient: Mr. Rudolf Nolan Corkery\n -- Gender: Male\n -- Contact Information: Home phone number is 555-124-9334\n - Address: Resides at 983 Gutmann Burg, Boston, Massachusetts, 02111, US\n - Marital Status: Married\n -Sibling Status: No multiple birth record\n ---- Primary Language: Communicates in English (United States) \n\n- Medical Encounter:\n--Status: Unknown\n -- Classification: Ambulatory\n -- Connection to Patient: Refers to patient Mr. Rudolf Nolan Corkery\n\n"\. {"record_id":2,"note":"This health record pertains to Mr. Hobert Armand Bashirian. Born on November 12, 1988, Mr. Bashirian lives at 76 mindler Well. Easthampton, Massachusetts, 01027, U.S. and speaks English. He is widowed and does not have any siblings. His medical history reveals a condition that began on February 3, 2014, at 80:38 AM (+01:80 timezone), \n\nMr. Bashirian also experienced a case of being a victim of intimate partner abuse. This started on January 12, 2008, at 01:89 AM (+01:00 timezone) and is currently active. \n\nRecently, on finalized on the same date and time.\n\n"},{"record_id":3,"mote":"A clinical note for Ms. Sonia Maria Bañuelos: \n\nOn December 24, 2013, at 16:59:45 GMT+1, Ms. Bañuelos received an injectable, preservative-free, seasonal influenza vaccine. This immunization procedure was completed smoothly. \n\nRegarding her personal information, Ms. Bañuelos, identifies as female and is not a multiple birth individual, meaning she doesn't have any twins or triplets. Her official residence is listed as 130 Rempel Vale, Boston, Massachusetts, 02111, United States. Moreover, it's noted that her primary language is Spanish. Hence, all communication with her should ideally be in Spanish for ensuring comprehension.\n\n\n"},{"record_id":4,"note":"1. Patient Information:\n···- Name: Mr. Asa Sherman Nolan\n··-- Contact Details Home Phone - 555-216-4138 \n -- Address: 698 Kunde Spur, Dedham, Massachusetts (Postal Code: 02026), United States\n -- Date of Birth: --- Time of Death: August 6, 1972 at 4:50 AM (GMT+1)\n --- Is a Single Birth: No\n --- Preferred Language: English (United States)\n\n2. Clinical Observations:\n -- Final Observation: Body height measured to be 174 cm. \n -\n3. Patient Survey Responses:\n\n - Has not feared partner or ex-partner in the past year.\n -- Feels physically and emotionally safe in current living conditions.\n-Not classified as a refugee. \n --- Has not been convicted and stayed over 2 nights in jail or juvenile correctional facility in the past year. \n -- Currently experiencing a little bit of stress.\n -- Communicates with close people (friends, family) 3 to 5 times a week.\n No missed medical appointments, meetings, work or daily needs due to lack of transportation.\n -- Opted not to answer if there has been a time in the past year where they couldn't provide for family when needed.\n -- Estimated total income before taxes for family in the past year was around 140,090 annually.\n -- Primary Insurance: Medicare\n -- Employment Status: Full-time work \n -- Education Level: High school diploma or GED \n -- Address: 698 Kunde Spur\n -- Does not worry about losing current housing.\n -- Has stable housing, not homeless.\n -- Residence houses 3 people.\n -- Preferred Language: English\n -- Has not been discharged from the United States armed forces.\n...- No income from seasonal or migrant farm work in the past 2 years for patient or family.\n...- Race: White\n...- Does not consider self as Hispanic/Latino.\n\n"}.{"record id":5,"note":"Patient Information:\n- Name: Mrs. Shanon Wolff, previously known as Mrs. Shanon Gerlach\n- Contact Number: Home contact, 555-968-5677\n- Gender: Female\n- Birth date: January 22, 1992\n- Address: 237 Koeloin Lane, Lyon, Massachusetts, 81982, United States\n- Multiple Birth: No\n- Preferred Language of Communication: English (United States) \n\nClinical Observations;\n1. Triglycerides Measurement;\n - Status; Final\n - Test Conducted; Triglycerides\n - Effective Time April 5, 2023, 14:14:00 (GMT+2)\n - Time of Issue: April 5, 2023, 14:14:00,790 (GMT+2)\n - Result: 124,76 mg/dl\n\n2, MCHC Measurementin -- Status: Unknownin -- Test Conducted: MCHC [Mass/volume] by Automated countin -- Result: 33.887 g/dlining. Social History Assessment:\n - Status: Final\n - Test Conducted: Tobacco smoking status\n - Effective Time: April 5, 2023, 14:14:00 (GMT+2) \n -- Time of Issue: April 5, 2023, 14:14:00.790 (GMT+2)\n\n"\. ("record id":6. "note": "Clinical Note:\n\n- Patient: Mr. Vince Maria Bergnaum\n- Gender: Male\n- Date of Birth: September 30, 1978\n- Allergy Status: Active\n- Alarm Verification: Confirmed\n- Type of Allergy: Environmental\n- Recorded Date of Allergy: July 11, 1999 at 89:83:29 GMT+82:88\n- Relation between resources: The indicated information describes an environmental allergy for Mr. Vince Maria Bergnaum, which has been confirmed and is active as recorded on July 11. 1999 at 09:03:29 GMT+02:00.\n\n"}.{"record id":7."note":"Clinical Note:\n\nMr. Sheldon Alfonso Reynolds, from West Wareham, Massachusetts postal code 02576, USA, living at 541 Ortiz Rue \n\nTwo observations reported, Unknown status, \n\nFirst, dated 24th Dec 2015, 02:19:44 +81:90, pertains to body weight. Second, associated with COZ total, represented as moles/volume, detected in blood.\n\nTwo procedures mentioned. First, marked as completed. Status of second procedure, unknown.\n\nNote: All events sequenced relates to Mr. Reynolds.\n\n"}. {"record id":8,"note":"This clinical note pertains to a patient named Ms. Gail Scarlet Hahn. Ms. Hahn is a never-married female, born on October 13, 2004. Her pronounced language of communication is English (specifically United States English). \n\nMs. Hahn currently lives at 183 Murazik Flat Apt 19, Cohasset, Massachusetts, in the United States. Her home contact number is 555-482-1978. Please note; Cohasset is in the Eastern Standard Time zone. Therefore, any scheduled calls or appointments must consider this time difference.\n\nFor this clinical process, the associated practitioner is Dr. Jon Keeling. The best mode of communicating with Dr. Keeling is through his work email, Jon665 Keeling57@example.com. \n\nAs we move forward in care, please recall the importance of maintaining clear, precise communication amongst all parties. This will foster an efficient and effective treatment process for Ms. Hahn. We are committed to delivering the utmost care and attention she deserves as we navigate through her health journey together.\n\n"},{"record_id":9,"note":"Clinical Note:\n\nPresenting before us today is Mr. Vicente Francisco Ornelas. He is our distinguished patient with whom we have an ongoing medical relationship. Mr. Ornelas has been patient and cooperative throughout the course of our medical consultations.\n\nIn connection to patient care, the medical



We showed how to go from messy, unstructured data → "insights"

- We want to be able to answer difficult questions and identify patterns
- We want to be able to *interpret* and *explain* the results



Why Kuzu?



- Embedded graph database that supports the property graph model
 - Runs everywhere, without servers, or with a server on top
 - Brings a fast, portable graph query engine to wherever your data sits
- Seamlessly switch between ephemeral and persistent modes
- Don't sweat over performance & scalability



Thank you! Learn more at kuzudb.com



Go above and beyond your traditional RAG architectures with Kuzu.

Give Kuzu a ★ on GitHub – and try it for yourself!



github.com/kuzudb/kuzu



@kuzudb

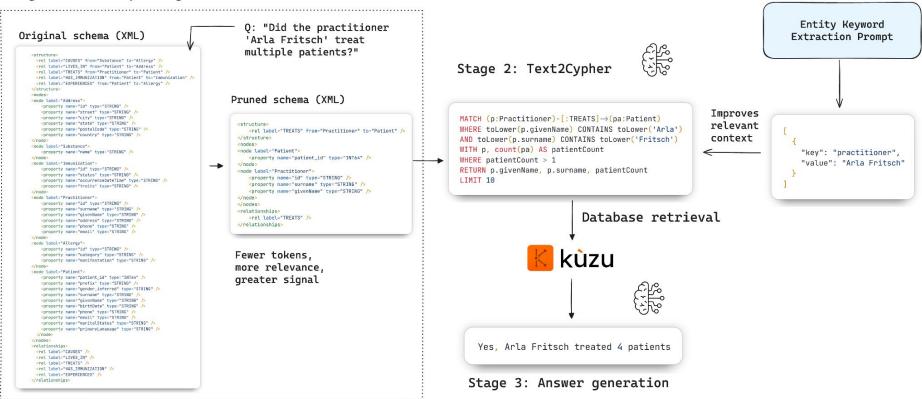


Read more about our work at <u>blog.kuzudb.com</u>

Appendix 1: Schema pruning to improve Text2Cypher



Stage 1: Schema pruning



Cypher vs. SQL: Differences in verbosity



"Are there any comments created by person that reply to a post also created by the same person?"

Cypher

```
MATCH (c:Comment)-[r1:commentHasCreator]→(p:Person)
←[r2:postHasCreator]-(post:Post)
←[r3:replyOfPost]-(c)

RETURN * LIMIT 10;
```

SQL

```
SELECT
  C.*,
  r1.comment id AS r1 comment id, r1.person id AS r1 person id,
  p.*,
               AS r2 post id.
                                 r2.person_id AS r2_person_id,
 r2.post_id
 post.*,
 r3.comment_id AS r3_comment_id, r3.post_id AS r3_post_id
FROM Comment AS c
JOIN commentHasCreator AS r1
  ON r1.comment id = c.id
JOIN Person AS p
  ON p.id = r1.person id
JOIN postHasCreator AS r2
  ON r2.person_id = p.id
JOIN Post AS post
  ON post.id = r2.post_id
JOIN replyOfPost AS r3
 ON r3.post_id = post.id
AND r3.comment_id = c.id
LIMIT 10:
```