Semantic GraphRAG

Everything You Need to Know
About Retrieval Augmented Generation,
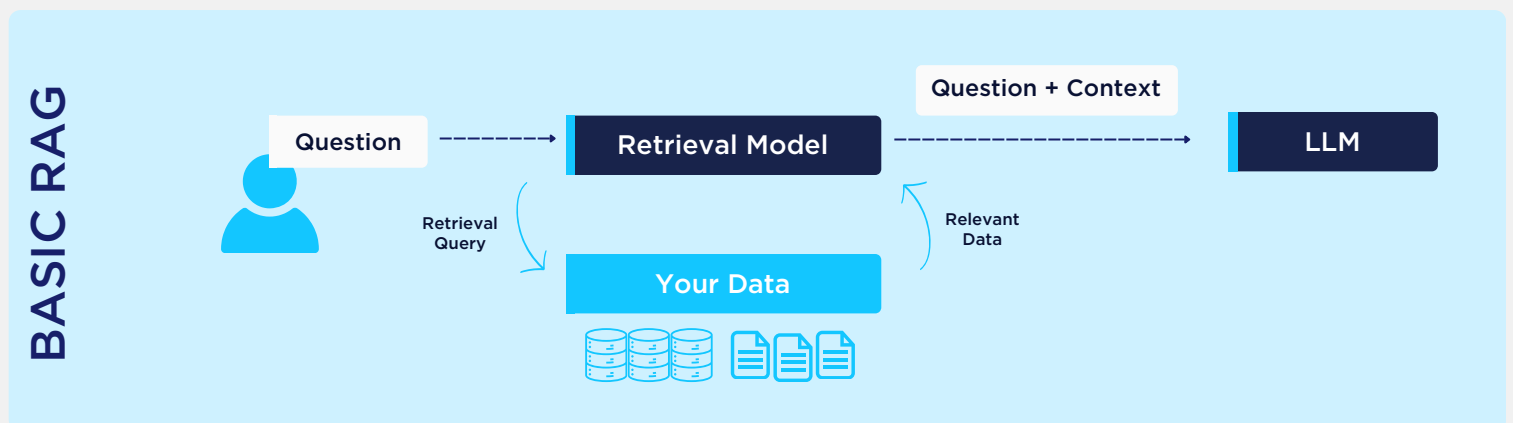Knowledge Graphs and LLMS

fluree

Understanding RAG

The idea behind RAG is quite simple: Adding proprietary information as context to a Generative AI application in hopes that it can accurately answer questions based on that data.
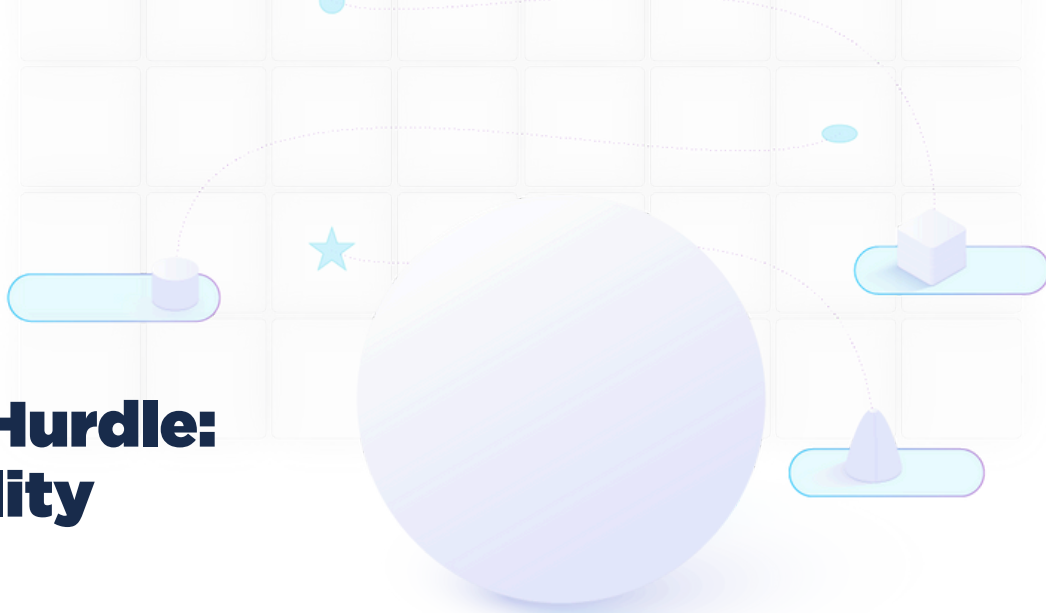
Do you trust LLMs with your data?

Large Language Models (LLMs) are aptly named because they're built on vast amounts of information from the internet, books, articles, and many other publicly available sources. This means they possess a broad and impressive general knowledge base, allowing them to answer a wide range of questions and assist in many scenarios.

However, despite this deep foundation, LLMs are not designed to handle specific, business-oriented questions—such as understanding the history of individual customer interactions or the unique context of an organization's data. LLMs simply don't have direct access to an organization's private or proprietary data; they can't "see" details about a specific customer or analyze custom data unless it's fed to them.

That's where Retrieval-Augmented Generation (RAG) becomes invaluable. RAG bridges this gap by enabling the LLM to retrieve and use up-to-date, relevant information from a specified data source in real-time. This means that instead of relying solely on what it already "knows," the LLM can pull in current data and provide responses tailored to real, business-specific contexts, such as recent customer transactions or interaction histories.

BASIC RAG

Question → Retrieval Model

Retrieval Query

Your Data

Relevant Data

Question + Context → LLM

The Last Hurdle: Data Quality

For businesses, RAG is in the spotlight - it helps answer the question: how can organizations connect their vast information systems to LLMs for Generative AI applications that are actually helpful?

If they can do so successfully, it can become a game-changer for the business - instead of using LLMs for simple tasks, like writing blog posts, successful RAG applications can hook into an organization's unique, live information sources and become helpful for any kind of advanced task.

But this is easier said than done: Data quality is incredibly important for an LLM to effectively answer questions. It's the classic principle of "Garbage In, Garbage Out"—an LLM can only be as accurate and reliable as the data it's fed. Before an organization can leverage an LLM successfully, it must first ensure that its data is well-organized, consistent, and clean.

The problem is that data quality often takes a dive when information needs to be integrated from separate data sources - and GenAI projects are now the single most driving force behind data integration efforts. VentureBeat released a study that while Generative AI grew 17% in 2024, organizations report that their data quality has plummeted.

Generative AI grows 17% in 2024, but data quality plummets

VentureBeat

78% of businesses unprepared for Gen AI due to poor data foundations

MIT Technology Review

Fewer AI projects are reaching deployment, and those that do are showing less ROI.
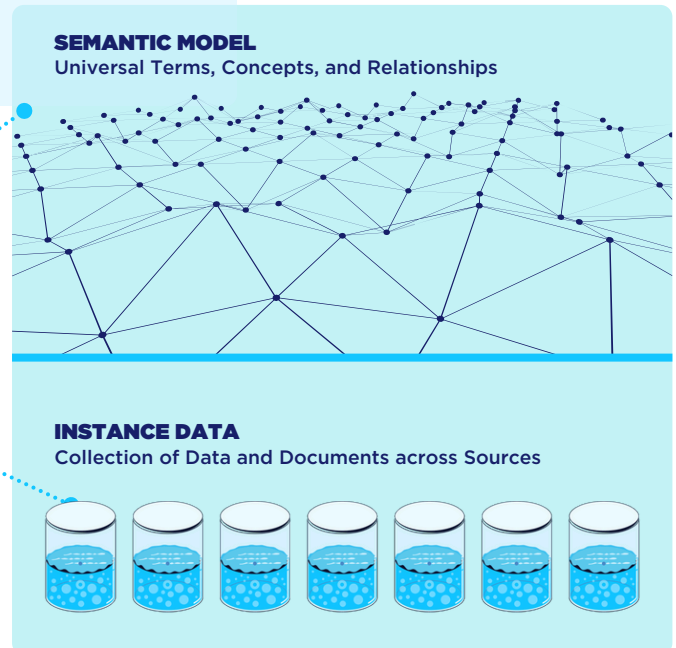
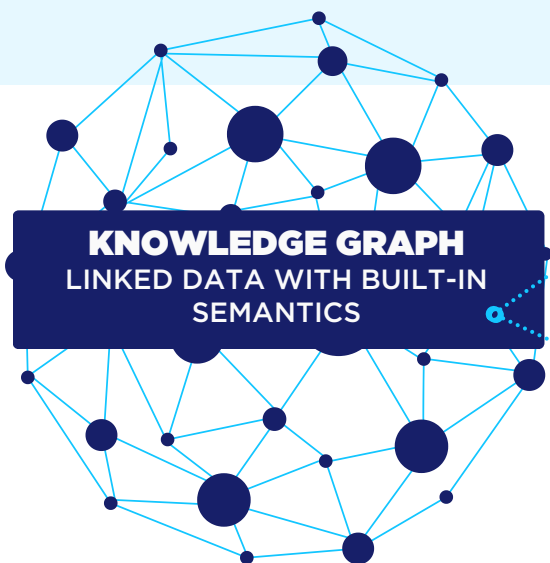**Appen**

Unifying Data with Semantics

The most effective RAG will need to find and use data across multiple different data sources. Today, integrating two or more data sources is a pain, because of proprietary schemas, entity resolution, and other data quality issues associated with integrating multiple heterogeneous data sources.

Every database system, ERP, or content management system has its own schema - a way of managing the concepts and metadata that defines what things are. In addition, each organization has their own unique terminology for terms and concepts. And here's the final layer of ambiguity: every department within a business unit probably has their own way of defining terms and concepts within their own systems.

What a finance department's ERP might call "client", is actually what a marketing department's application schema defines as a "customer."

This ambiguity spells disaster for enterprise-wide AI initiatives that might need to pull together information from multiple sources: if we unleash RAG on top of all our database systems as they are today, we might get plausible answers, but they couldn't be 100% trusted. We might get duplicates. We might not get the full picture. And worse of all - we might get hallucinations that we act on.

This is where knowledge graphs come into play.

KNOWLEDGE GRAPH
LINKED DATA WITH BUILT-IN SEMANTICS

SEMANTIC MODEL
Universal Terms, Concepts, and Relationships

INSTANCE DATA
Collection of Data and Documents across Sources

Semantics:
Your Global Blueprint

Think of semantics as a global blueprint of your terms and concepts. A semantic model helps us define, at a global layer: "this means that."

Once we have this global blueprint in place, we can apply it to any kind of information in our downstream systems. This helps with things like entity resolution, where we can determine that data in one system holds reference to the same exact entity in a different system. Semantic models not only define "things" but also globally define the relationships between those things.

Let's say we had an ontology - which is a semantic model within a particular domain - that defined "client" in the finance system and "customer" in the marketing system are actually the same entity. By using this ontology, our system could automatically recognize that the "client" in the finance ERP and the "customer" in the marketing database refer to the same individual. This is the first layer of power that semantics brings: it helps us align and unify concepts across different systems, eliminating duplicates.

But there's more. Let's also say that our ontology knows that clients (or customers) have interactions with sales agents, and that sales agents work within specific business units. In this case, it might automatically form a new relationship between the client entity in the finance ERP and the sales team stored in the CRM. This relationship didn't exist explicitly before because the two systems weren't connected, but the ontology recognizes the link through shared relationships and concepts.

By leveraging this semantic blueprint, we not only ensure that entities are correctly identified across data sources but also automatically discover relationships that were hidden or fragmented.

Let's Review

Semantic GraphRAG refers to Information Retrieval that leverages knowledge graphs as context to pass along to GenAI applications.

Knowledge graphs are comprised of semantic data models that globally define concepts, terms, and relationships, as well as instance data mapped to the semantic model.

SEMANTIC MODEL
Universal Terms, Concepts, and Relationships
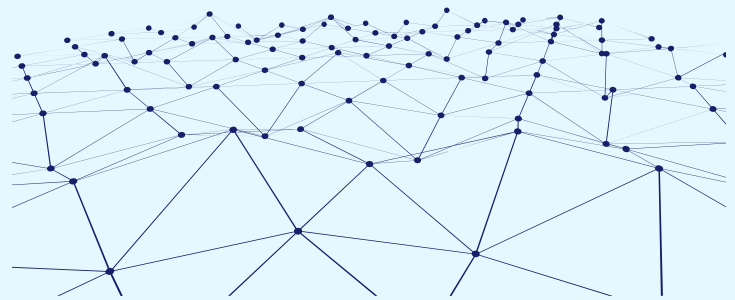
INSTANCE DATA
Collection of Data and Documents across Sources
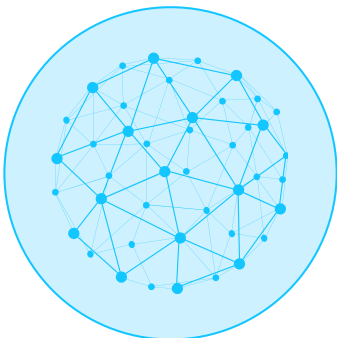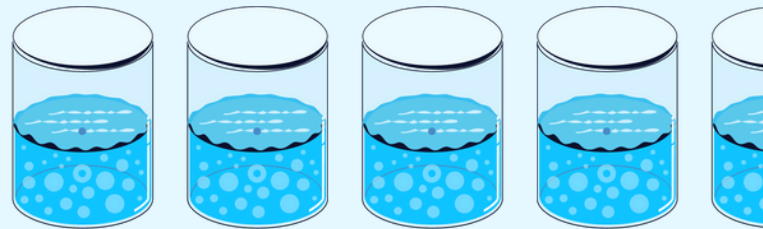
Knowledge Graphs help unify data silos under common domain concepts, lending to higher data quality and increased knowledge discovery.

Organizations that use a semantic approach to their RAG projects will see fewer hallucinations, more contextual and explainable answers, and an overall lower cost of IT overhead associated with connecting AI models with enterprise data.

Building Your Knowledge Graph - With AI

1. Create your semantic model

So, how do we build this semantic model?

Well, if you are in the majority of organizations, you likely don't have an ontology. You might have a few taxonomies or data dictionaries spread out across a few departments. If this is the case, there a few fantastic options to build one:

- Start with an off-the-shelf ontology. An upper ontology like GIST provides a solid foundation for broad, general concepts, while a domain-specific ontology, such as Allotrope for pharmaceutical manufacturing, offers industry-standard definitions and relationships tailored to laboratory analytical processes.

- Use machine learning and Generative AI to build one. With Fluree, you can reverse engineer an ontology from your existing taxonomies, schemas, content and data. Once candidate terms are suggested, we can tweak concepts, definitions, and relationships as necessary.

Using both of these strategies, we can build our blueprint of global terms and concepts that are unique to your business, jargon, and data.

2. Classify and link your data

Once we have our ontology defined, we need to classify instance data against that ontology. Fluree has out of the box algorithms to do just that: drop or load any unstructured content (like PDFS, audio, text, video), structured data (like relational databases), or ERP system (like workday or salesforce). We'll classify the data against our ontology, resolve duplicate entities, and form relationships across the various information
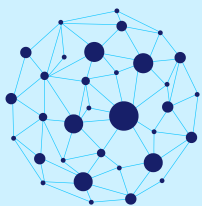
We now have a semantic data model that can be used as master context for any LLM, and instance data that is classified, cleaned, and linked together as a unified data store for information retrieval.

Conclusion

Why use Semantics in RAG? For information retrieval that requires deterministic answers based on available data, semantics gives us the structure to ground LLMs in actual enterprise knowledge.

Whereas traditional RAG may be helpful for simple applications, an ontological approach to connecting enterprise data systems to GenerativeAI is the key to unlocking grounded, explicit information retrieval.

Ready to Learn More?

Production enterprise AI requires accuracy, reduced errors, and compliance with global privacy and copyright laws. Fluree uniquely offers trusted, verifiable data management and built-in programmable policy control.

With a decentralized knowledge graph approach, Fluree enables you to build an Enterprise Corpus of data that can easily link all of your proprietary data anywhere in the world, dynamically connect it with industry and market data, and safely expose it to real-time LLM applications.

Visit us at: http://flur.ee

Contact us at: sales@flur.ee

fluree

Zero-Hallucination Enterprise AI Data Platform

Fluree helps organizations securely integrate and connect data from various silos, making it ready for use with Generative AI to support accurate and explainable decisions.

All of your data, ready for AI

Fluree transforms data silos into a secure, unified data plane for AI.

Organizations that use Fluree's approach see up to 4X increase in GenAI accuracy over traditional approaches when it comes to information retrieval.*

Connect to any data source

Fluree provides out-of-the-box connectors to ingest structured, unstructured, and semi-structured data from any source.

Auto-extract and link knowledge together

Fluree extracts the semantic meaning from your data, classifying and unifying it into a connected source of integrated knowledge

Build secure real-time access to knowledge

Fluree delivers real-time, secure data access to LLMs, with embedded policies that guarantee confidentiality, governance, and compliance

Why Fluree

Platform Agnostic

- Runs Anywhere: On-Prem, Cloud, Snowflake, AWS, Azure
- Connects to Any Source: PDF, Audio, Text, Oracle, SAP, CMS, APIs

Zero Hallucinations

- Context-aware answers grounded only on your enterprise knowledge
- Data lineage and provenance built-in increases trust and explainability

Secure and Private

- Add and manage data security policies across all data and users
- Never leak data to an LLM or unauthorized user

Trusted By

Gartner COOL VENDOR 2024 · DOWJONES · CBC · Medscape · WARNER BROS. DISCOVERY · WebMD Ignite · NSW Health · n p r · Dotdash meredith · AP Associated Press