

neo4j

Going Meta

S3 - Episode 3

Ontologies, LLMs, Property
Graphs.. Do we really have a
"Jaguar problem"?

A few weeks ago on LinkedIn...

Niklas Emegård  · 1st
Full Stack Enterprise Architect | Military Grade Leadership | AI & Knowled...
1mo · Edited · 0

After seeing yet another Graph RAG demo using Neo4j with no ontology, I decided to show what real semantic Graph RAG looks like.

The Problem with Most Graph RAG Demos:
Everyone's building Graph RAG with LPG databases (Neo4j, TigerGraph, Arrango etc.) and calling it "knowledge graphs." But here's the thing:
Without formal ontologies, you don't have a knowledge graph—you just have a graph database.

The difference?
 LPG: Nodes and edges are just strings. No semantics. No reasoning. No standards.
 RDF/SPARQL: Formal ontologies (RDFS/OWL) that define domain knowledge. Machine-readable semantics. W3C standards. Built-in reasoning.

So I Built a Real Semantic Graph RAG
Using:
- Microsoft Agent Framework - AI orchestration
- Formal ontologies - RDFS/OWL knowledge representation
- Ontotext GraphDB - RDF triple store
- SPARQL - semantic querying
- GPT-5 - ontology-aware extraction

It's all on github, a simple template as boilerplate for your project:

The "Jaguar problem":
What does "Yesterday I was hit by a Jaguar" really mean? It is impossible to know without concept awareness. To demonstrate why ontologies matter, I

The "Jaguar problem":

What does "Yesterday I was hit by a Jaguar" really mean? It is impossible to know without concept awareness. To demonstrate why ontologies matter, I created a corpus with mixed content:

-  Wildlife jaguars (*Panthera onca*)
-  Jaguar cars (E-Type, XK-E)
-  Fender Jaguar guitars

   1,505 · 195 Comments

https://www.linkedin.com/posts/niklasemegard_graphrag-knowledgegraphs-semanticweb-activity-7390898141820092416-0ojW/

Inaccuracy #1

LLMs cannot deal with lexical ambiguity without help



The "Jaguar problem":

What does "Yesterday I was hit by a Jaguar" really mean? It's impossible to know without concept awareness. To demonstrate why ontologies matter, I created a corpus with mixed content:

- 🐆 Wildlife jaguars (*Panthera onca*)
- 🚗 Jaguar cars (E-Type, XK-E)
- 🎸 Fender Jaguar guitars

I fed this to GPT-5 along with a jaguar conservation ontology.

The result? The LLM automatically extracted ONLY wildlife-related entities—filtering out cars and guitars—because it understood the semantic domain from the ontology.

Inaccuracy #2

Feeding an Ontology in the context window + prompting guarantees schema alignment

Solution: By providing GPT-5 with the **jaguar conservation ontology**, the LLM:

4. Generates RDF Turtle aligned with the ontology

Source: https://github.com/nemegrod/graph_RAG/blob/main/docs/agent_design.md

Generated triples misaligned with the Ontology

Ontology

```
ont:occursIn a owl:ObjectProperty ;  
    rdfs:domain ont:Animal ;  
    rdfs:range ont:Location ;  
    rdfs:comment "Indicates a state where an animal has been  
    observed or is known to occur".
```

Extracted
Triples

```
:HuasabasSahuaripaPop a ont:JaguarPopulation .  
:HuasabasSahuaripaPop ont:occursIn :Sonora .  
. :Sonora a ont:State .
```

Consequence?

:HuasabasSahuaripaPop a ont:Animal
-> wrong results on animal count

Source: line 327 in https://github.com/nemegrod/graph_RAG/blob/main/data/jaguars.ttl

Inaccuracy #2

Feeding an Ontology in the context window + prompting guarantees schema alignment



Solution: By providing GPT-5 with the jaguar conservation ontology, the LLM:

4. Generates RDF Turtle aligned with the ontology

Source: https://github.com/nemegrod/graph_RAG/blob/main/docs/agent_design.md

Inaccuracy #3

The extracted triples don't need post-processing

2. Ontology-Aware Knowledge Extraction

- Zero Post-Processing: Extracts clean, structured data without manual cleanup

Source: https://github.com/nemegrod/graph_RAG/blob/main/docs/agent_design.md

Inaccuracy #3

The extracted triples don't need post-processing

- Inconsistent use of rdfs:label

Some extracted entities have labels, others don't. This cannot be inferred/completed.
- Duplicated triples

:ElJefe ont:facesThreat :BorderBarrierThreat .
- Incorrect interpretation of the ontology

:FelizardoValenciaStudents a ont:Person;
rdfs:label "Felizardo Valencia Middle School students".

Source: https://github.com/nemegrod/graph_RAG/blob/main/data/jaguars.ttl

Inaccuracy #3

The extracted triples don't need post-processing



2. Ontology-Aware Knowledge Extraction

- Zero Post-Processing: Extracts clean, structured data without manual cleanup

Source: https://github.com/nemegrod/graph_RAG/blob/main/docs/agent_design.md

Inaccuracy #4

LLMs can run OWL reasoning (formal semantics)

The Magic: Concept Understanding Through Formal Semantics

This is where **ontology-driven extraction** happens. We provide GPT-5 with:

1. The **ontology** (semantic structure)
2. The **corpus** (unstructured text)
3. Instructions to extract entities and relationships that **align with the ontology**

Source: https://github.com/nemegrod/graph_RAG/blob/main/text2knowledge.ipynb

If you want to run
OWL reasoning en
Deterministic reas
inference engine.
This is Dangerous!!

Ora Lassila  • 1st
Principal Technologist at Amazon Neptune (RDF, knowledge g...
7mo ...

Best comment I heard at the conference (from Jesús Barrasa):
"Why do you need an LLM, the problem is completely
deterministic?" and I realized I wanted to ask that after almost
every presentation.

Like  16 | Reply · 2 replies

David Wood  • 2nd
Entrepreneurial technology leader and multi-disciplinar...
7mo ...

Ora Lassila nailed it!

Like | Reply

Chimezie Ogbuji  • 2nd
Writer, informatics engineer, musician, and entrepreneur
7mo ...

Ora Lassila Brilliant question. Concisely packages what I've
been trying to articulate

Like | Reply

Inaccuracy #4

LLMs can run OWL reasoning (formal semantics)



The Magic: Concept Understanding Through Formal Semantics

This is where **ontology-driven extraction** happens. We provide CPT-5 with:

1. The **ontology** (semantic structure)
2. The **corpus** (unstructured text)
3. Instructions to extract entities and relationships that **align with the ontology**

Source: https://github.com/nemegrod/graph_RAG/blob/main/text2knowledge.ipynb

Inaccuracy #5

LLMs perform better when we prompt them with an ontology in RDF/OWL/etc for entity extraction vs when we do it with NL

Things, not strings... blah blah... well, turns out that LLMs ❤️ strings 🤗
Let's prove it.

Inaccuracy #5

LLMs perform better when we prompt them with an ontology in RDF/OWL/etc for entity extraction vs when we do it with NL



And an extra one for free:
The extraction includes hallucinations!!! 😱

```
:HatoLaAuroraPop ont:isDependentOn :HuasabasSahuaripaPop .  
  
:HatoLaAuroraPop ont:occursIn :LlanosRegion .  
  
:HuasabasSahuaripaPop ont:occursIn :Sonora .
```

Source: line 131 https://github.com/nemegrod/graph_RAG/blob/main/data/jaguars.ttl

Inaccuracy #6

This CANNOT be done with LPG databases
because...



Takeaways?



See you next month!