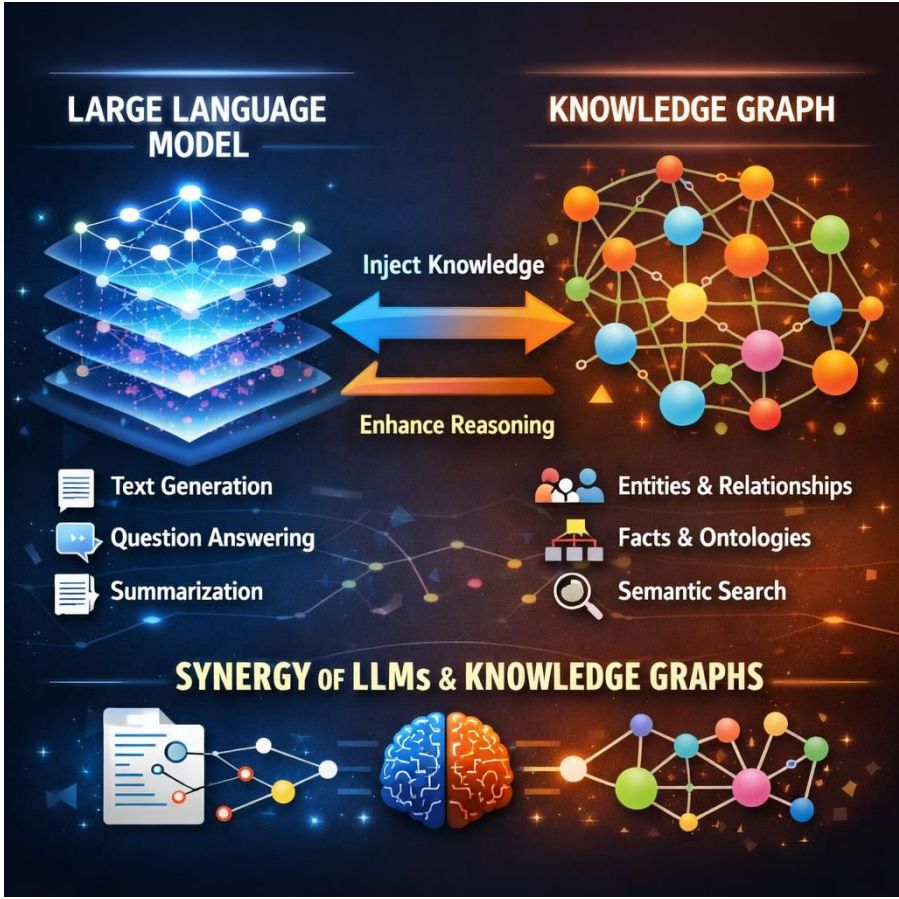
Large Language Models & Knowledge Graphs

Petr Křemen

Ontologies and Semantic Web
Winter 2025

Agenda today

- **Motivation**: mutual benefits + core algorithms
- **Explainability**: graph/ontology-based “why”
- **Learning**: LLM-assisted KG/ontology construction
- **GraphRAG + Talk to Your Graph (GraphDB) demo**
- **Takeaways** + architecture recipes



https://artlist.io/

Depict the interplay between large language models and knowledge graphs.

Why to combine LLMs & KGs?

Let’s start by asking ChatGPT 4o …

In OWL 2 DL, explain why a subproperty of a symmetric property must necessarily be symmetric.

5. Intuitive summary

- Symmetry is a **downward-inheritable constraint** in OWL.
- Subproperties cannot “escape” global relational properties.
- Therefore, symmetry propagates from a property to all its subproperties.

This is why OWL 2 DL reasoners will **entail** symmetry for subproperties, even if you do not explicitly assert it.

Provide a reference of an existing mathematical proof of this claim.

Recommended citation

> W3C (2012). *OWL 2 Web Ontology Language: Direct Semantics*. https://www.w3.org/TR/owl2-direct-semantics/ ↗

or

> Baader et al. (2017). *The Description Logic Handbook*, Cambridge University Press.

What about ontologies & knowledge graphs?

- Linked Open Data - almost two decades, more than hundred billions of triples
  - large
  - varying quality
  - explainable - rich annotations, SPARQL
  - large communities to produce it
- SNOMED-CT - over decades, more than 350k concepts
  - middle-sized
  - high quality (manually curated)
  - explainable - rich annotations, ECL
  - large communities to maintain it

So, where we stand …

LLMs

- constructed automatically from un/structured content
- fluent, extensive inference
- weak on truth and explainability
- considerable resource consumption

KGs

- constructed manually or automatically from structured content
- precise, well-explainable
- weak on authoring and generative reasoning
- considerable manual resources needed

What LLMs get from ontologies/KGs

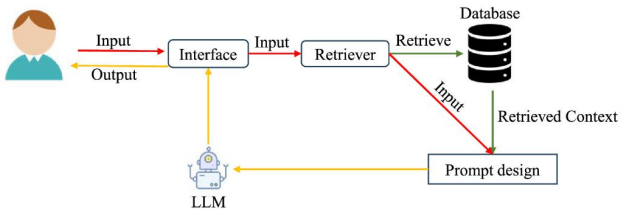
- Disambiguation
- Provenance + governance
- Explainability via schemata (OWL, RDFS, SHACL)

What KGs get from LLMs

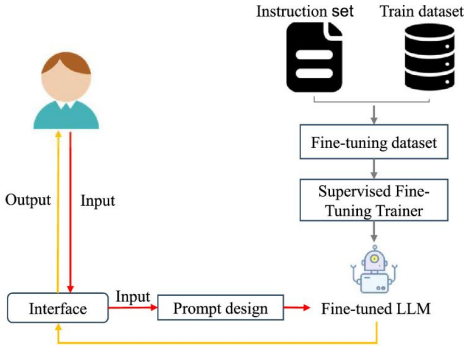
- Extraction at scale (entities/relations/events)
- Schema suggestions (classes/properties)
- Mapping candidates (alignment hints)
- Natural language interface (NLQ → SPARQL / Cypher)

Interaction patterns

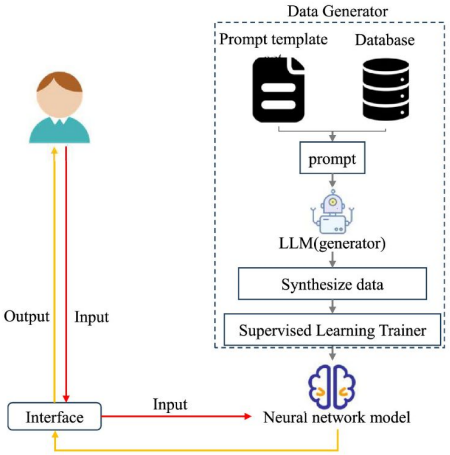
- schema taken from [4]



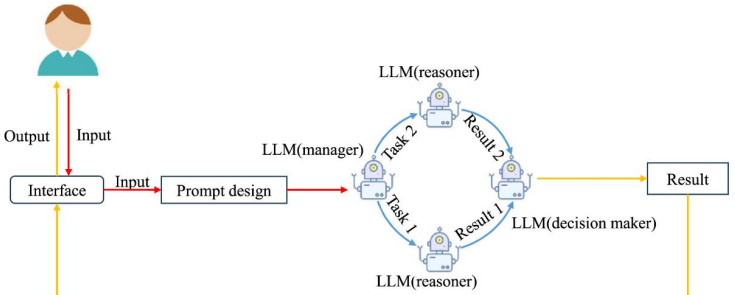
(a) Prompt-based methods



(b) Fine-tuning based methods



(c) Single LLM agent based methods



(d) Multiple LLM agents based methods

Integration patterns

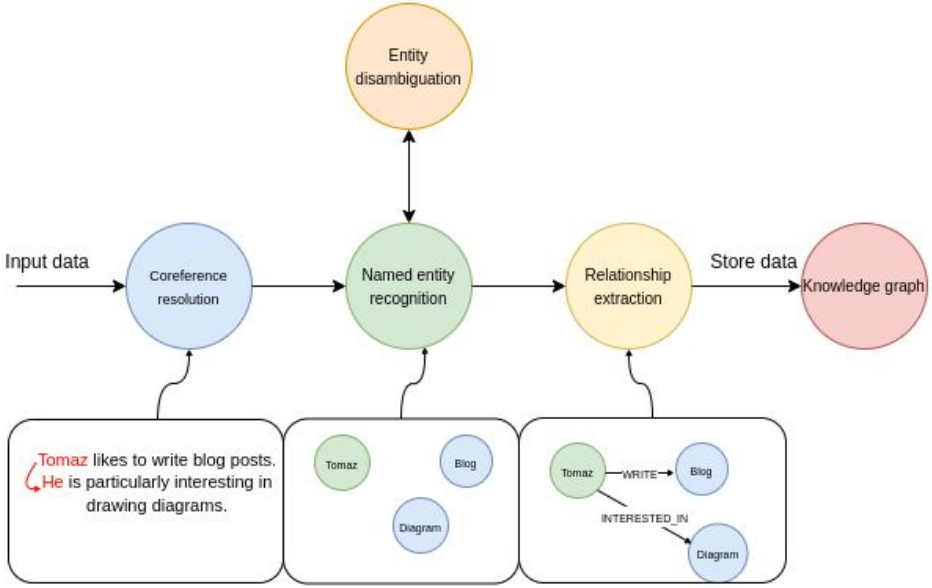
- **Knowledge Creation**
  - LLM-assisted KG / ontology population
- **Explainable Query Answering**
  - KG-grounded LLM answering with justifications (GraphRAG)

Knowledge Graph Creation

- Ontologies & KGs are expensive to build
- LLMs can propose structure at scale but are not authoritative
- Constraints prevent semantic corruption

example tools:

- llmgraph
  https://github.com/dylanhogg/llmgraph
- langchain
  langchain.com



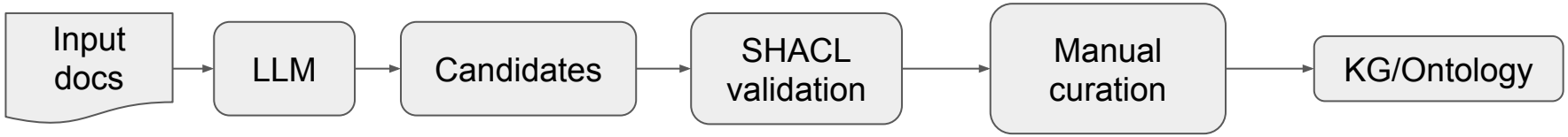
taken from
https://blog.langchain.com/constructing-knowledge-graphs-from-text-using-openai-functions/

Knowledge Graph Creation - Variants

- **Knowledge Graph population (schema fixed)**
  a. candidate triples
  b. entity recognition, entity linking

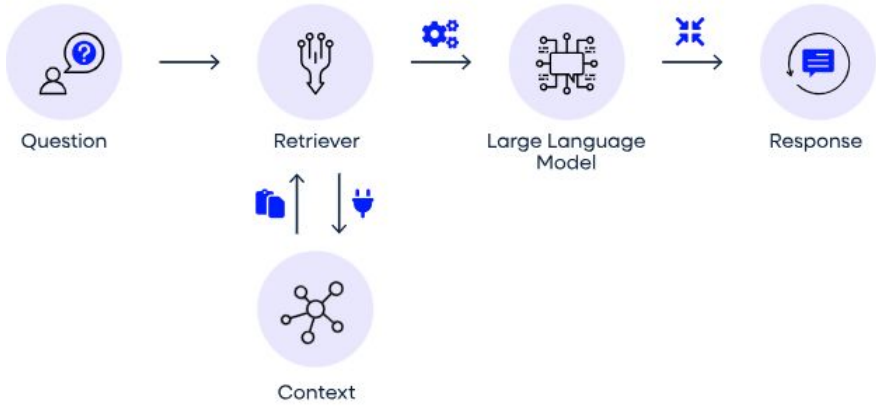
- **Schema generation (Ontology / SHACL)**
  a. candidate OWL classes/properties/axioms
  b. candidate SHACL shapes

Knowledge Graph Creation - Flow



Explainable Query Answering

- augmenting the prompt with SPARQL results to
  a. improve the response quality
  b. reduce token usage
- We don’t ask the LLM to know the world — we ask it to explain what the graph already knows.



taken from https://graphwise.ai/use-cases/graph-rag/

Explainable Query Answering

- Query-first
  a. LLM transforms user query to SPARQL
  b. SPARQL retrieves results
  c. LLM is asked for *communicating them to the user*

- Context-first (GraphRAG)
  a. LLM transforms user query to SPARQL
  b. Graph retrieves results
  c. LLM is asked for *providing results* partially using graph results

Explainable Query Answering - some tools

- Talk to Your Graph
  https://graphdb.ontotext.com/documentation/11.2/talk-to-graph.html

- Microsoft GraphRAG
  https://microsoft.github.io/graphrag/?utm_source=chatgpt.com

How to make Ontology models AI-ready?

- Ontologies have a **descriptive** function
  - they should be able to describe data.

- ***Good*** ontologies have an **explanatory** function (see UFO lecture)
  - they should be able to explain the domain

Talk To Your Graph

- https://graphdb.ontotext.com/documentation/11.2/talk-to-graph.html
- ChatBot (GraphRAG) in GraphDB
- For a response provides
  - token usage
  - explanations in the form of SPARQL queries
- Support for
  - OpenAI Assistants API
  - OpenAI Completions API
  - Gemini

Talk To Your Graph
Demo

Reference

1. James O'Donnell and Casey Crownhart, MIT Technology review. 2025 https://www.technologyreview.com/2025/05/20/1116327/ai-energy-usage-climate-footprint-big-tech/
2. Jiayi Li, Daniel Garijo , María Poveda-Villalón. Large Language Models for Ontology Engineering: A Systematic Literature Review. In *Semantic Web*, 2025 https://www.semantic-web-journal.net/system/files/swj3864.pdf
3. Hamed Babaei Giglou, Jennifer D'Souza, Sören Auer. LLMs4OL: Large Language Models for Ontology Learning, 2023. https://arxiv.org/abs/2307.16648
4. Yancheng Ling, Zhenlin Qin, Zhenliang M. A review of knowledge graph construction using large language models in transportation: Problems, methods, and challenges. In *Transportation Research Part C: Emerging Technologies*. 2026 https://www.sciencedirect.com/science/article/pii/S0968090X25004322