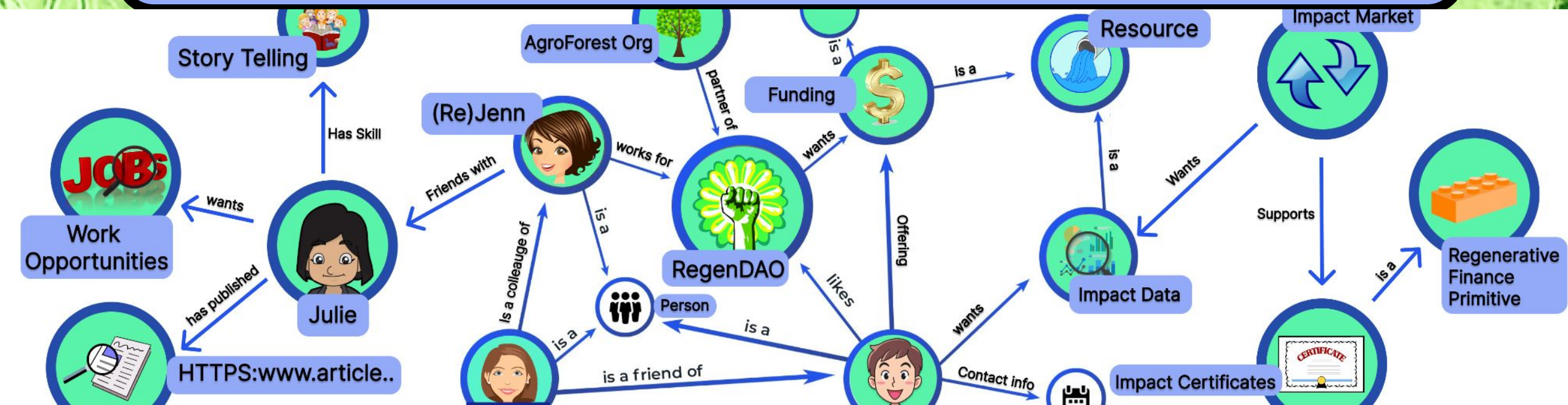


Natural Language Knowledge Graph Question Answering

Darren Zal





Darren Zal
Software Engineer
Vancouver Island, Turtle Island (North America)

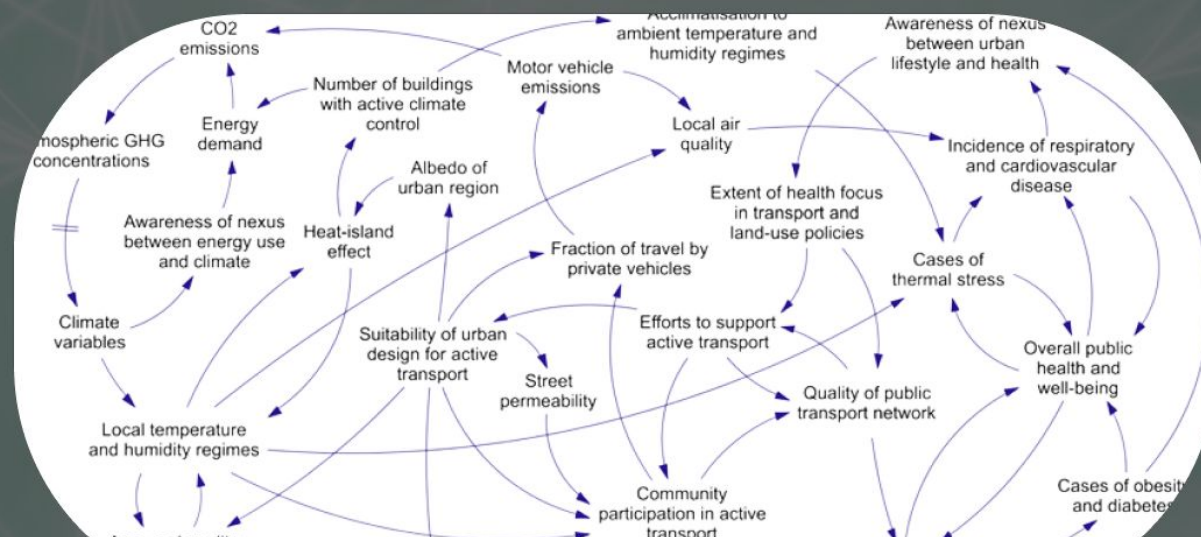
Background: Relational databases, Web3, and regenerative finance (ReFi)
Motivation: Build tools for collective sensemaking and wise decision making

Humans face a converging cascade of interconnected crises:
ecological, economic, geopolitical, educational, psychological..

Global challenges are interconnected & multi-stakeholder

Interdependencies
are too complex
for individuals
to disentangle

Lack of whole
system map for
bottoms up
collaboration at
global scale



Data is siloed, not
Findable,
Accessible,
Interoperable,
Re-usable

Inconsistent and
imprecise taxonomies
impedes the
synthesis of data

DEMOCRATIZING KNOWLEDGE

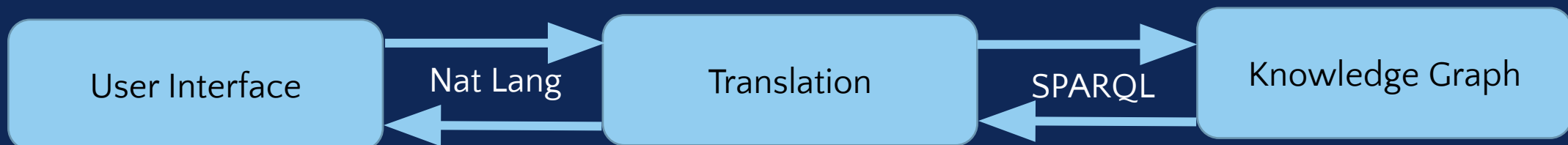
Problem: SPARQL is the powerful standard language of knowledge graphs, but it requires technical skills. A significant portion of people are left out.

Solution: An interface that translates between natural language and SPARQL.

- Utilize AI tools to ensure accurate, context-aware translations, making knowledge graphs as easy to query as asking a question.

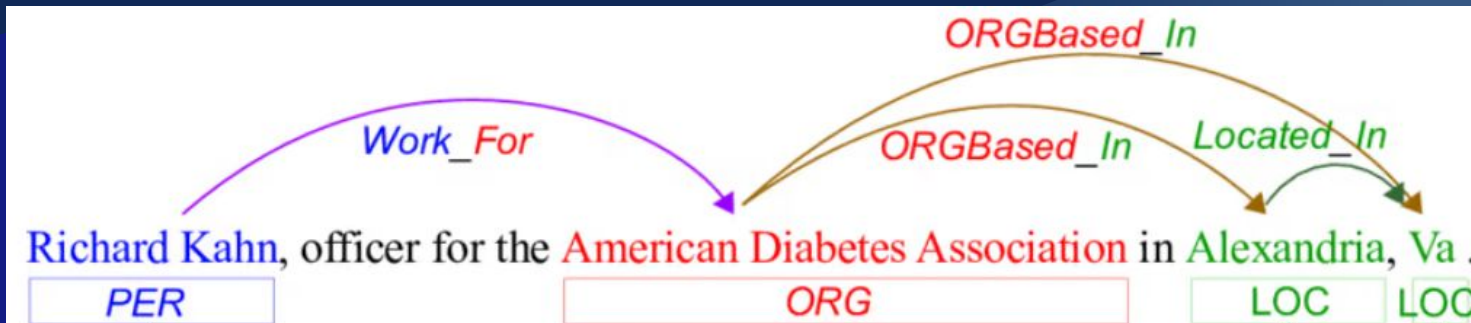
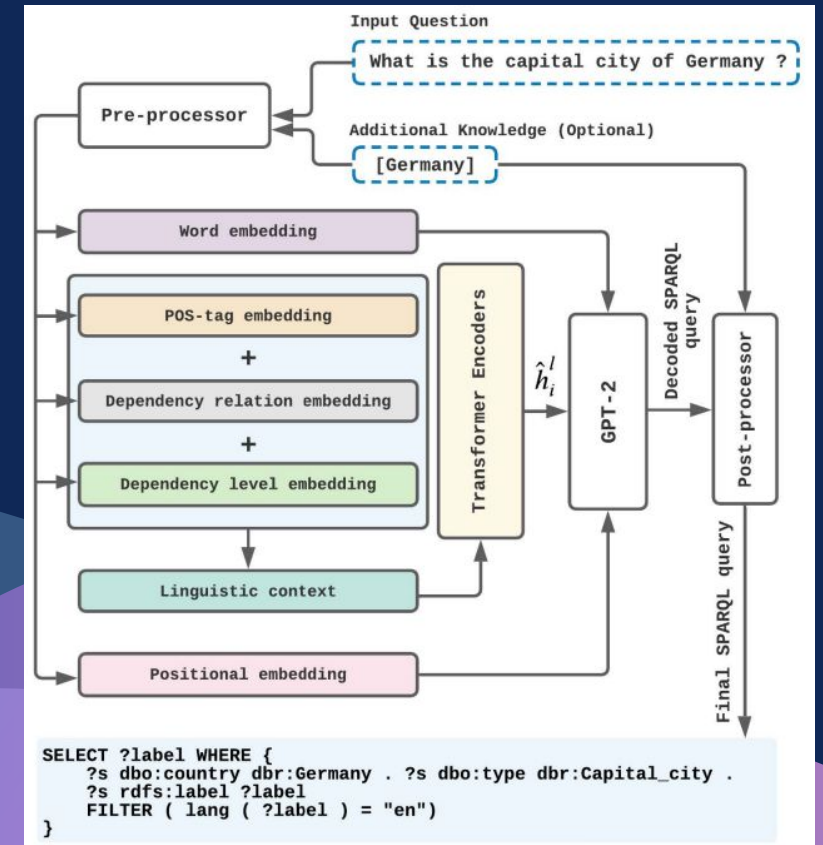
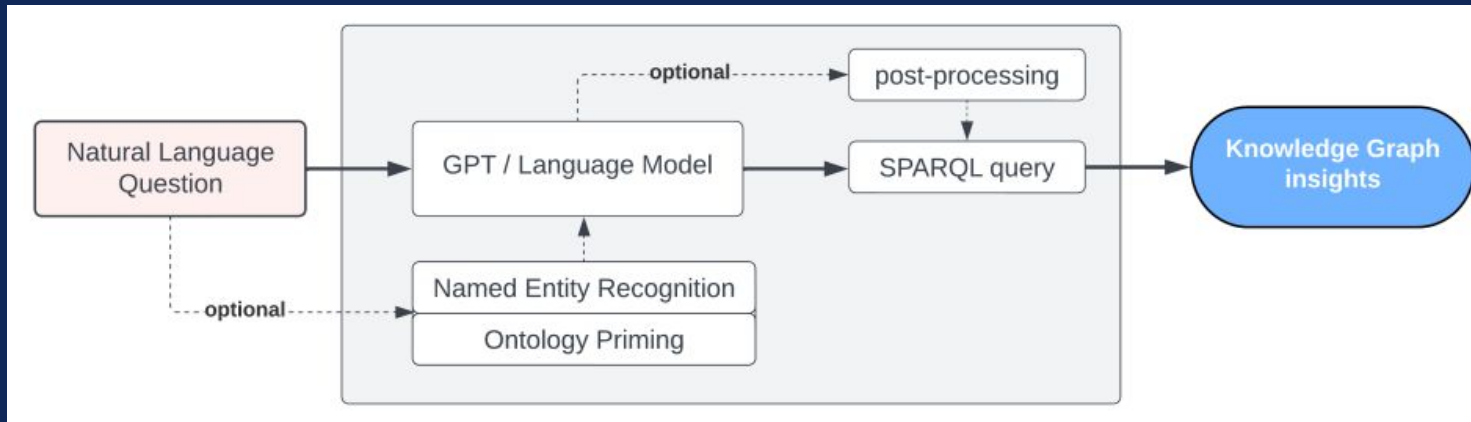
Benefits:

- **Democratization of Data**: Make knowledge graphs accessible to everyone, regardless of technical ability.
 - **Enhanced Decision Making**: Enable a wider audience to derive insights directly.
 - **Innovation & Collaboration**: Unlock creative solutions and foster cross-domain collaborations
 - **User Engagement**: Drive a richer more inclusive data ecosystem.



STATE OF THE ART

- LLMs like GPT-4 can already do a decent job of translating natural language queries to SPARQL
- Large datasets of natural language <> SPARQL exist (LC-QuAD 2.0, VQuAnDA and QALD-9)
- Research and code exists on the topic
 - SGPT: A Generative Approach for SPARQL Query Generation From Natural Language Questions
 - Tree-KGQA: An Unsupervised Approach for Question Answering Over Knowledge Graphs



How good is GPT-4?

Prompt: can you generate my a SPARQL query to get the youtube channel of GS1, using **wikidata**?

```
GPT-4: SELECT ?youtubeChannel WHERE {  
  ?gs1 wdt:P2397 ?youtubeChannel.  
  ?gs1 rdfs:label ?label.  
  FILTER(CONTAINS(LCASE(?label), "gs1"))  
  FILTER(LANG(?label) = "en")  
}
```

..no results.

Prompt: I see GS1 has ID Q731100, does that help?

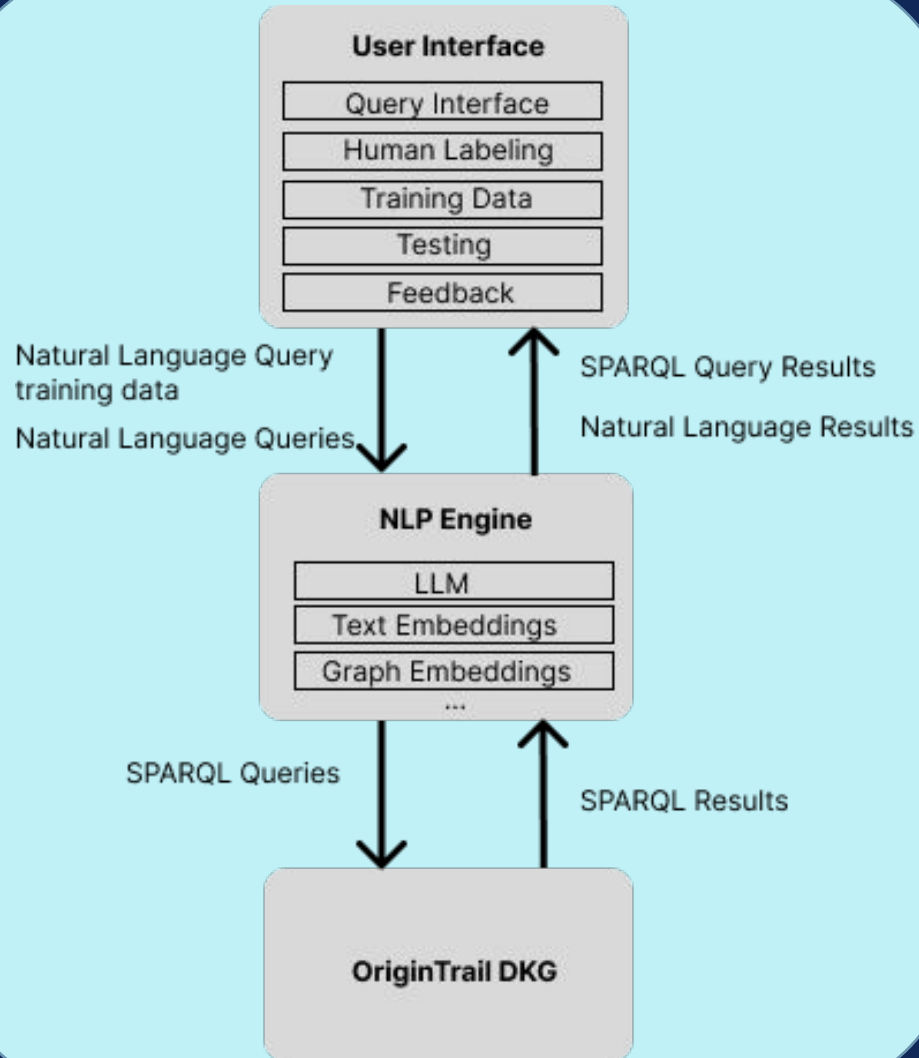
```
GPT-4: SELECT ?youtubeChannel WHERE {  
  wd:Q731100 wdt:P2397 ?youtubeChannel.  
}
```

Wikidata Query Service yields: UCdCkxrh_TGWeBNzQ_QqZx6g

Prompt: what is the wikidata ID for the organization GS1?

GPT-4: As of my last knowledge update in January 2022, the Wikidata ID for the organization GS1 is Q726696.
Wrong!

Architecture



User Interface: where users can ask questions in natural language



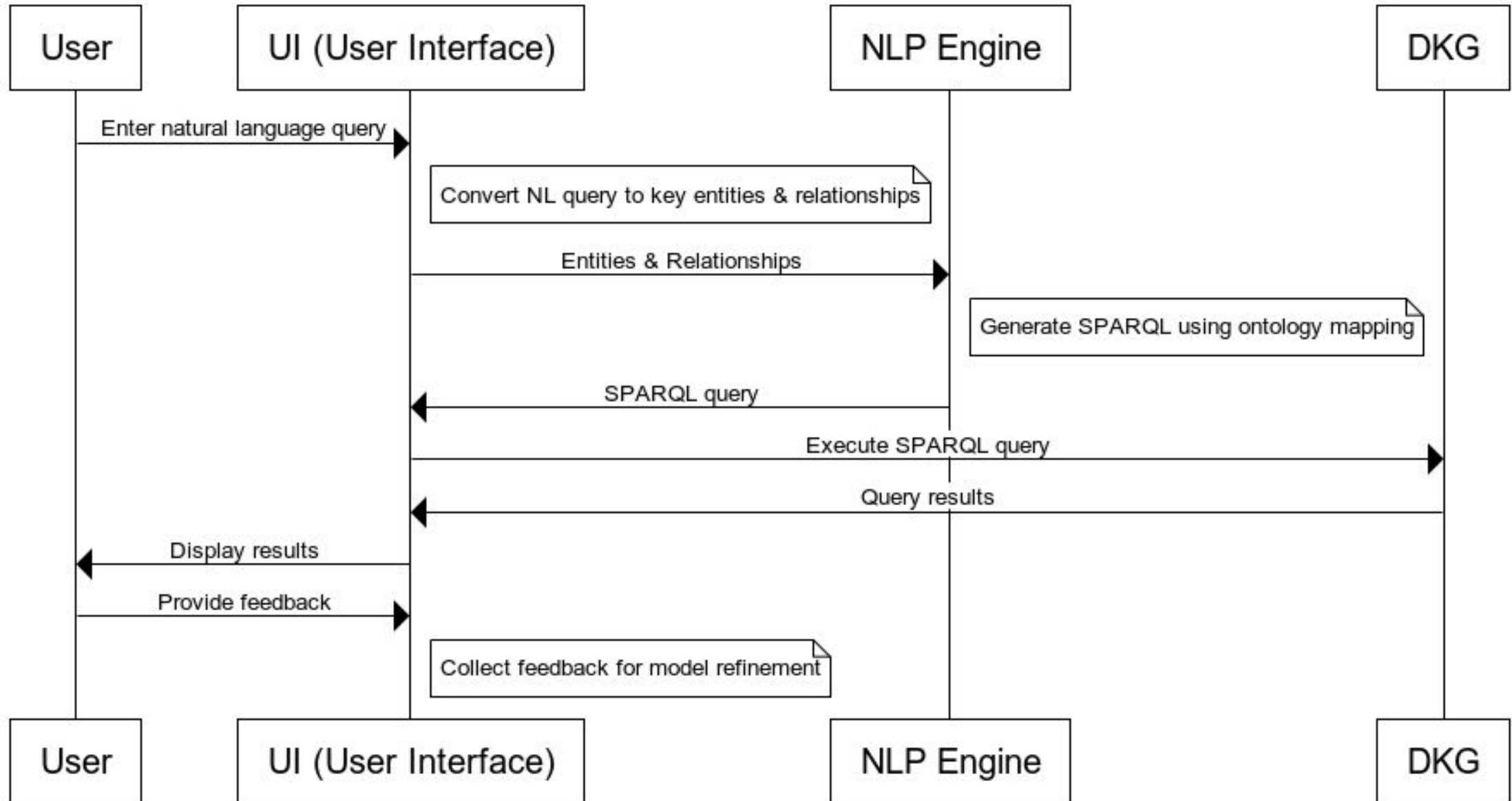
Training Data: Examples of natural language <> SPARQL, which can be used to fine tune a pre-trained LLM.

NLP Engine: Natural Language Processing Engine. Where the magic happens to convert a natural language prompt into a SPARQL query, and convert the results into natural language.

Process

7

Natural Language to SPARQL Conversion



ROADMAP



Step 1 ●

Research and
specification



Step 2 ●

Acquire a dataset
([ReFiDAO](#))



Step 3 ●

Software development

- UI
- Fine tuning a LLM
- Text and Graph Embeddings
- Named Entity Recognition
- Ontology Resolution



Step 4 ●

Finalization and
Launch



Step 5 ●

Ongoing
Stewardship and
Evolution