

D-RAG

**Densifying Retriever for
Appropriate Generation**

The problem

- Gieni is an LLM agent answering deep reasoning questions over the highly specialized knowledge sector of supply chain management
- Gieni needs to make best use of the sparse information found on company websites

...by analysing our user persona we found that:

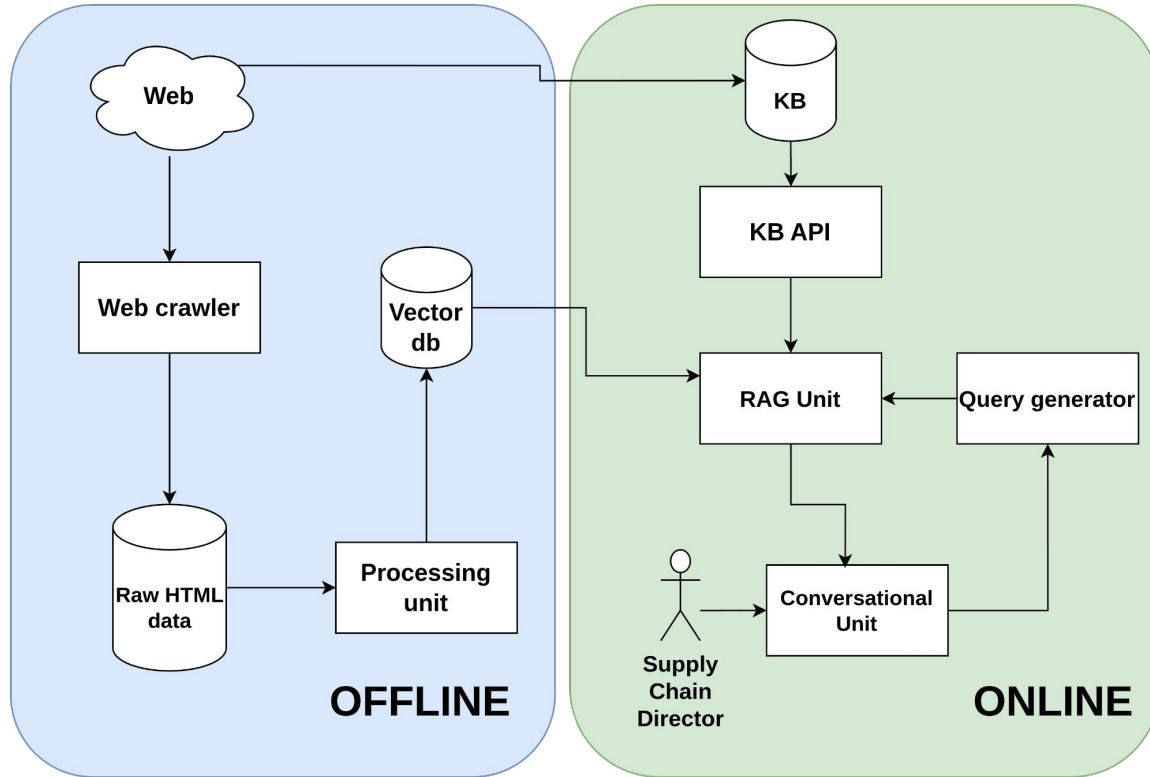
- Spatial awareness is key to provide accurate localized reports
- Feature engineering and query augmentation are required to deepen question answering in the supply chain analysis

The solution

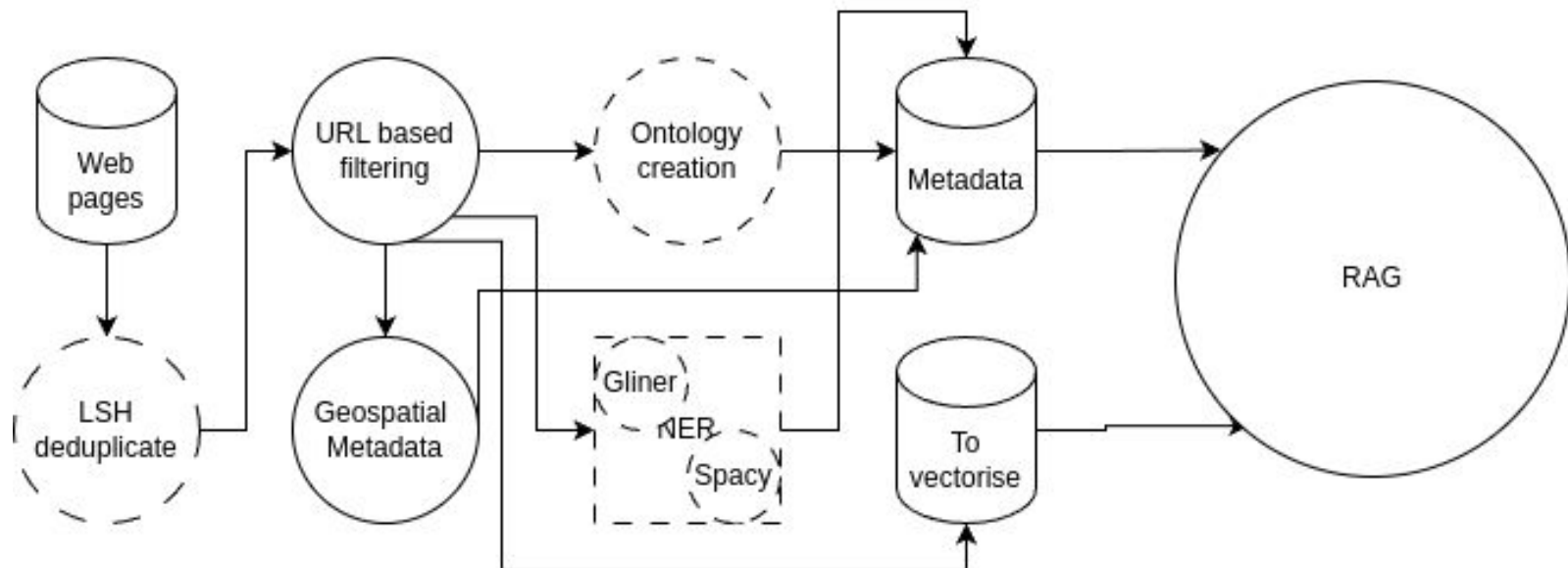
A RAG system designed to:

- Increase the information density, with a lightweight large-scale dataset preprocessing
- Leverage the topological properties of the set
- Incorporate additionally engineered features on geolocation (and potentially much more)
- Rely on KB and anthologies to narrow the user's request by context augmentation
- Run multiquery beams during the conversation (parallel historical data)

The Big picture

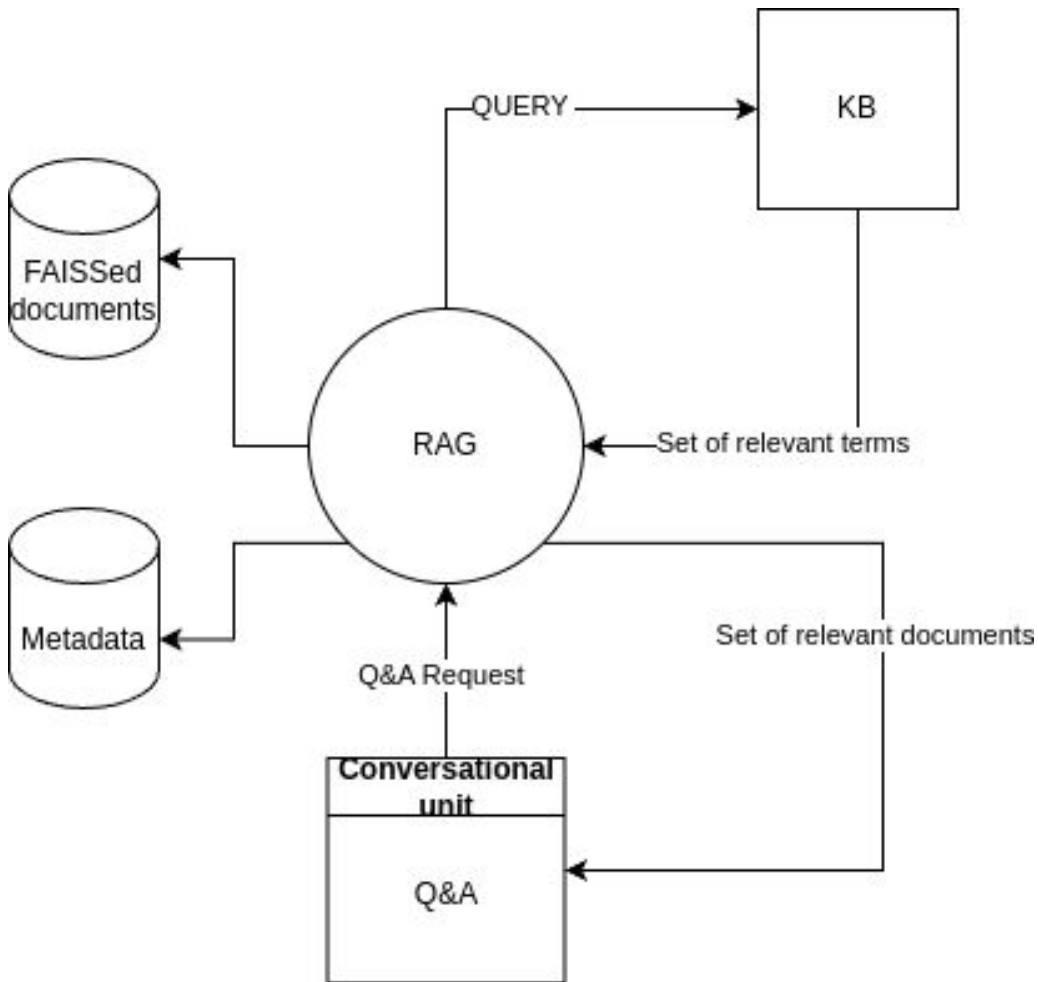


Preprocessing pipeline



RAG

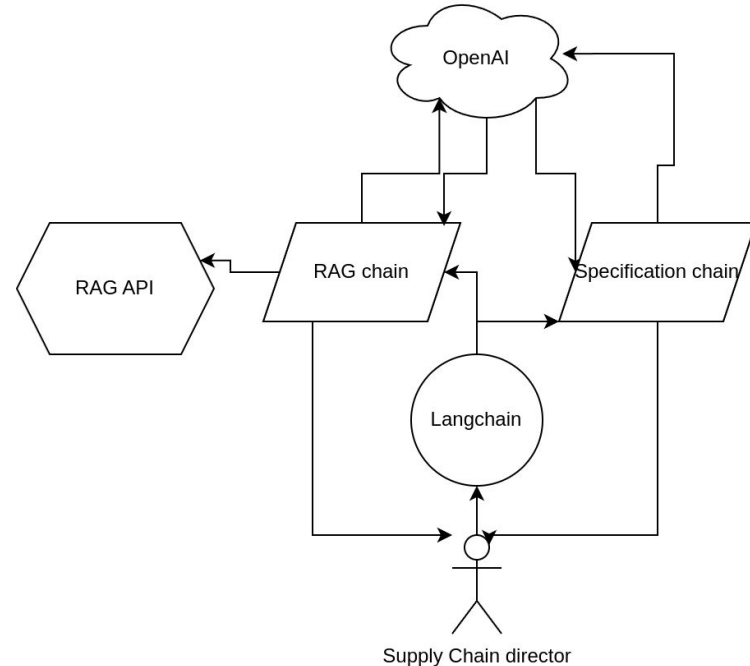
1. KB specifies the request of user
2. RAG queries both databases
3. A set of relevant documents with a combined scoring is retrieved



Conversational unit



- Specification chain is responsible for making additional user questions
- RAG chain is responsible for the retrieval process
- The chat is leveraged with the historical chat context



The end

