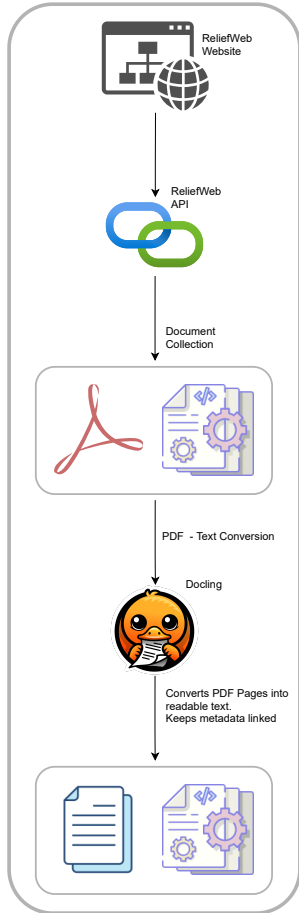
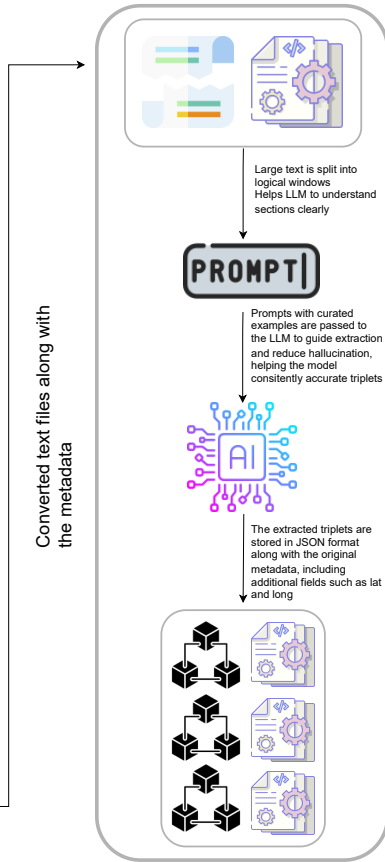


## Data Preparation



- Selected required reports using ReliefWeb filters (country, category, date).
- Used ReliefWeb API to download PDFs + metadata automatically.
- Collected 5 key report types: Analysis, Appeals, New & Press Release, UN Documents, Situation Reports.
- Countries focused: South Sudan and Ukraine.
- Passed downloaded PDFs to Docling for text extraction.
- Extracted text and kept original metadata linked to each document.

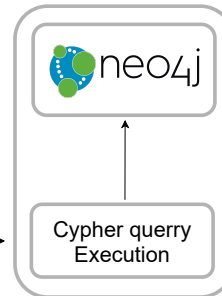
## Data Processing



Converted text files along with the metadata

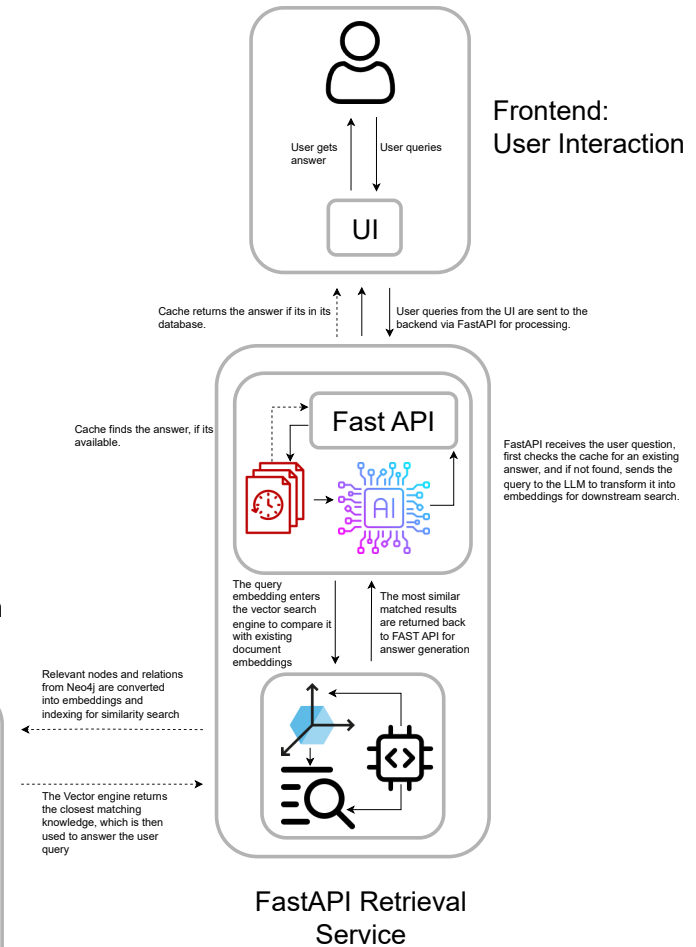
- The extracted text + metadata are split into fixed window sizes.
- Each text window is combined with the relevant metadata.
- Windows are sent to the LLM through API calls with a focused prompt.
- LLM generates subject–relation–object triplets from each window.
- Triplets are saved together with original metadata.
- Final output is stored in structured JSON format.

## Knowledge Graph Construction



The generated JSON triplets are ingested through Cypher scripts into Neo4j, where subjects and objects become nodes, relations become edges, and all metadata is stored as properties inside the Knowledge Graph

## Frontend: User Interaction



Cache finds the answer, if its available.

Cache returns the answer if its in its database.

User queries from the UI are sent to the backend via FastAPI for processing.

FastAPI receives the user question, first checks the cache for an existing answer, and if not found, sends the query to the LLM to transform it into embeddings for downstream search.

## FastAPI Retrieval Service