Background

Spatiotemporal Index Extraction. Converting unstructured data into structured representations is a critical step toward enabling intelligent systems to reason and query effectively. Knowledge graphs are often considered an ideal target structure; however, their construction remains challenging and labor-intensive. Human memory and reasoning, by contrast, are naturally indexed along spatial and temporal dimensions: we recall events by when and where they occurred, supporting fast retrieval and causal reasoning. Inspired by this, we argue that extracting spatiotemporal indices from unstructured data offers a practical and broadly applicable alternative. Such indices can provide a lightweight but powerful way to organize information, enable efficient retrieval, and serve as a foundation for higher-level reasoning, while avoiding much of the complexity of full knowledge graph construction.

Task Definition

Input: Unstructured text.

Output:

- Identify and extract temporal mentions, and convert them into normalized timeline points or intervals.
- Identify and extract spatial mentions, and convert them into geospatial locations (e.g., map coordinates or regions).

Example:

> **Input text:** "On March 15, 2022, a strong cyclone hit the coastal areas near Broome, Western Australia and later moved inland towards Fitzroy Crossing by March 17."
>
> **Temporal output:**
>
> - March 15, 2022 → 2022-03-15
> - March 17 → 2022-03-17
>
> **Spatial output:**
>
> - Broome, Western Australia → (17.9614° S, 122.2359° E)
> - Fitzroy Crossing, Western Australia → (18.1976° S, 125.5669° E)

Roadmap

1. **LLM-based Prototype:** Leverage large language models (LLMs) to extract spatiotemporal indices from text and build an initial dataset. Deliver an LLM-driven prototype package for this task.

2. **Model Refinement:** Annotate the collected dataset to ensure quality, and use it to fine-tune an LLM or train a specialized extraction model.

3. **Production-Ready System:** Develop and release a mature, stable Python package that automates the full pipeline for spatiotemporal index extraction.