Extract Define Canonicalize Presentation

Advisor: Dr. Fabio Miranda
Team: Vamsi Dath Meka, Gustavo Moreira

Friday / October 10 2025

Presented by

Vamsi Dath Meka

Selection of Files

Step 1: Data Processing

The data that is to be input into the Extract, Define, Canonicalize Pipeline should be refined

Extracting Triplets

Step 2: EDC Pipiline

Two types of triplets can be obtained

1. IOE triplets (Open Information Extraction)
2. Canon Triplets (pruned by definition)

Graph Construction

Step 3: Using Visualization tools for Graph Exploration

Refining nodes and relations

Step 4: Schema Formation

The data that is to be input into the Extract, Define, Canonicalize Pipleline should be refined

How does EDC Work?

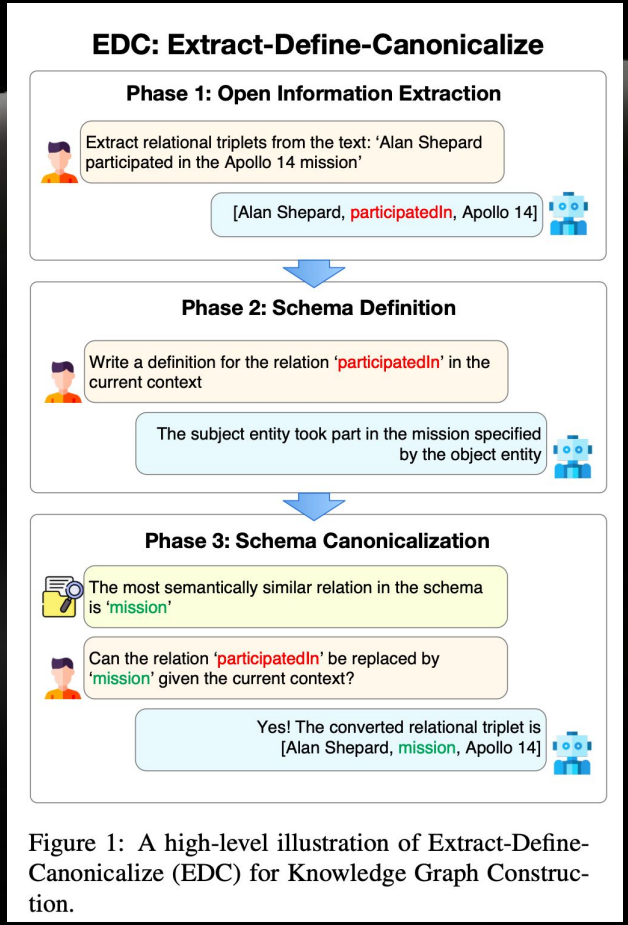
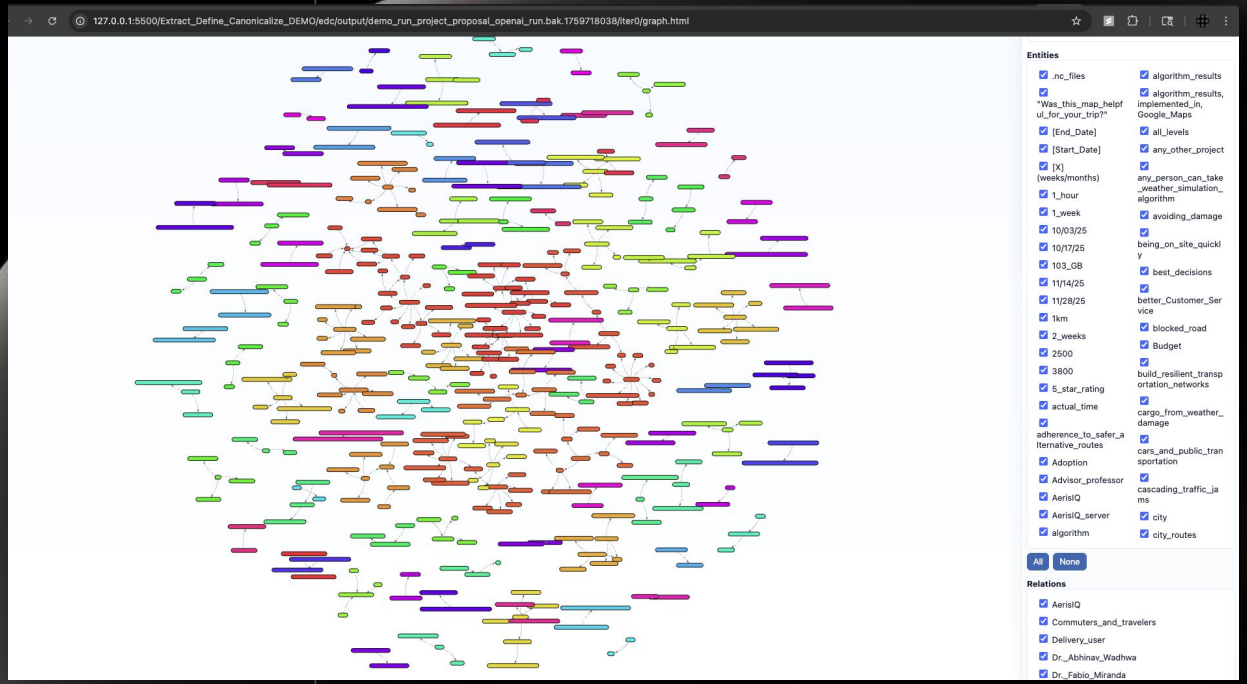
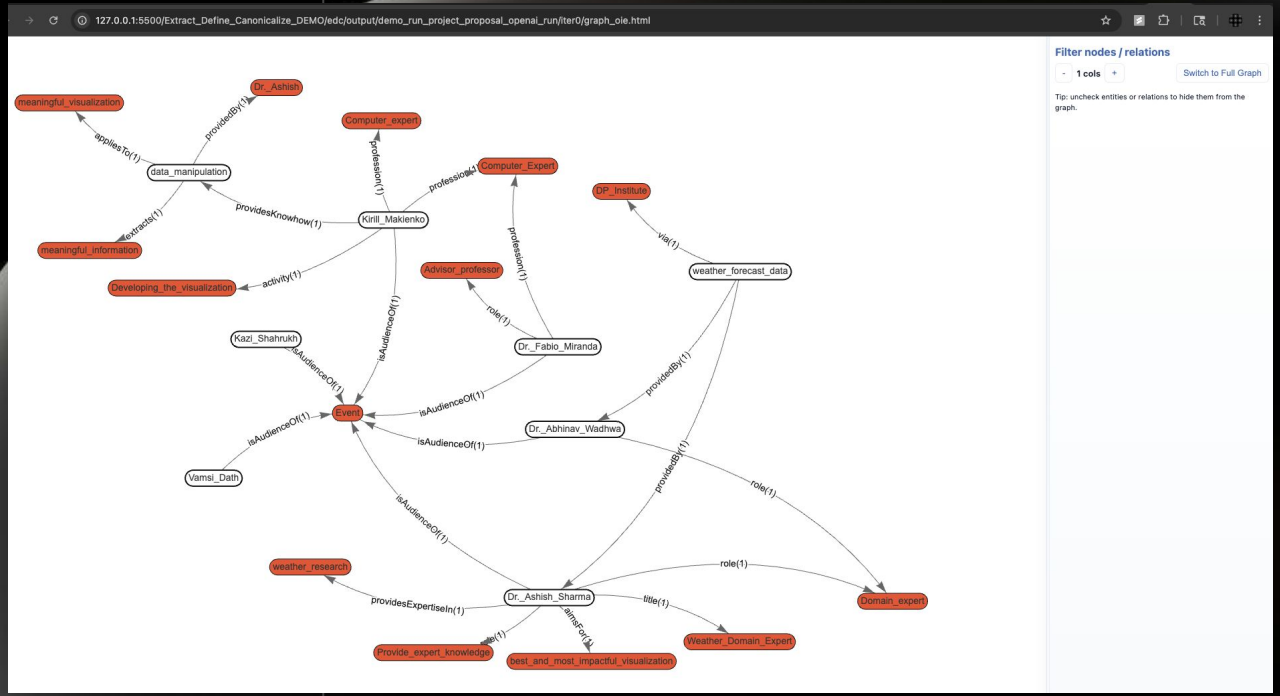


Figure 1: A high-level illustration of Extract-Define-Canonicalize (EDC) for Knowledge Graph Construction.

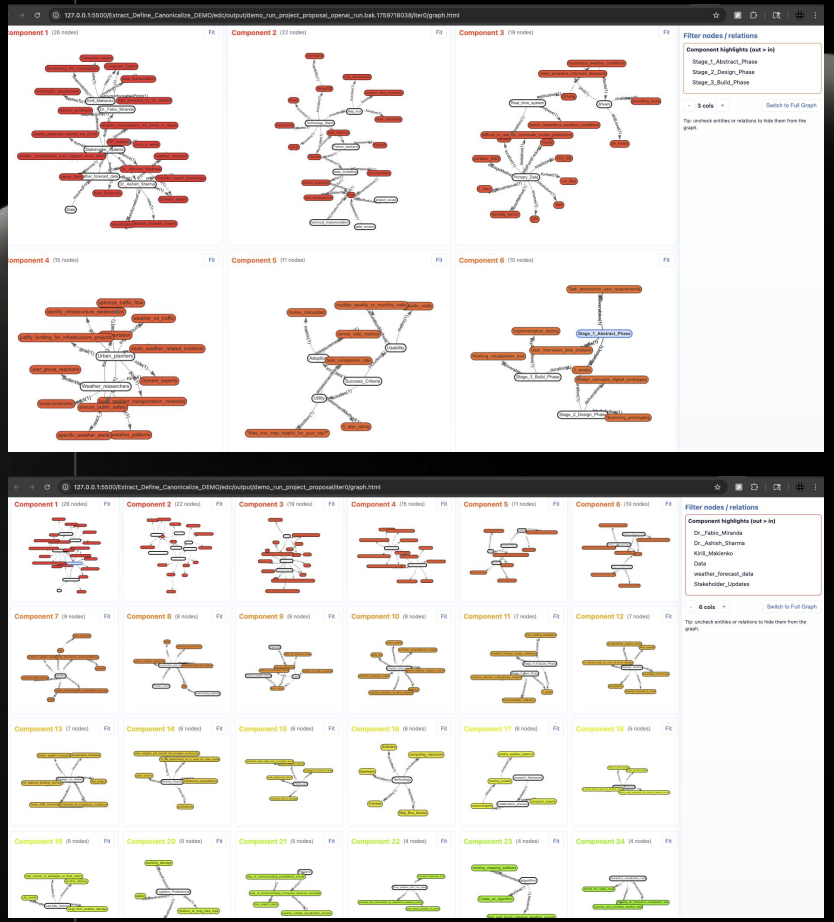
Overall Graph



Components of Graph

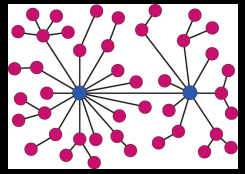


Exploration (Graph)



Future Work

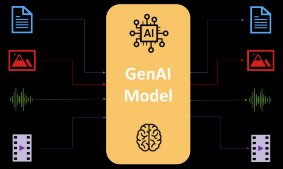
Connecting various Components, nodes and relations with Human induced relations



Using existing Tools like neo4j for this process, or explore alternatives

Refine the Schema

Add functionality for processing Multi-Modal Data



References

Extract, Define, Canonicalize: An LLM-based Framework for Knowledge Graph Construction
Link: https://arxiv.org/abs/2404.03868

GitHub: github.com/clear-nus/edc
Link: https://github.com/clear-nus/edc

Directed Graph in NetworkX
Link: https://networkx.org/documentation/stable/tutorial.html#directed-graphs

GitHub: Implementation
Link: https://github.com/Vamsi-Dath/extract_define_canonicalize_pipeline

Thank you

Any Questions?

Vamsi Dath Meka
vmeka@uic.edu
+1 312-776-6210