Meeting Minutes – Technical Direction & Research Planning

Prepared by: Anandhu Raveendran (24254189)
Date: 8 August 2025
Project: Language-Based Geographic Reasoning System
Attendees: Whole group

1. Query Handling and Data Retrieval

- All queries must be **clearly checked and verified** to ensure **high-quality output** from the spatial database.
- Queries should be **spatially aware**, i.e., written to fetch **spatial data** using appropriate **spatial SQL**.
- Once the correct data is retrieved, it should be used to **generate map visualizations** for the end user.

2. Codebase Management and Annotation Workflow

- Use **Jupyter Notebook** for:
  - Writing and testing code
  - Keeping a record of the entire codebase
  - Maintaining an **annotated workflow** that supports future **research publication**
- Follow a structured **annotation pipeline** that documents:
  - Code logic
  - Dataset usage
  - Query examples
  - Outputs and interpretations

3. Table and Attribute Identification

- Critical task: **Identify which table and attribute(s)** should be used to answer a given query.
- Potential approaches:
  - Leverage **Large Language Models (LLMs)** for mapping queries to relevant tables/fields.
  - Implement a **classification model** to automatically:
    - Detect the relevant table
    - Identify specific attributes needed to resolve the query
- Current Advantage:
  - Since the dataset is small, we already know the structure and can use it as **training data**.

4. Machine Learning Implementation

- Build a **sentence classification model** to categorize and interpret incoming natural language queries.
- This model will serve as the **first step in the NLP pipeline**, mapping text to schema components.
- Future Goal:
  - **Fine-tune the classification model**, but note that a **large annotated dataset** will be required for effective performance.

5. Optimization and Efficiency

- Emphasize **finding the most efficient approach** at every stage of the project.
- Focus on:
  - Query optimization

    - Code reuse
    - Scalable model training
    - Efficient data handling

6. Presentation Note for Xudong

- For Xudong's upcoming presentation:
    - Consider using **3 classification levels** instead of 4.
    - This may simplify the model architecture and improve interpretability.