CITS5553 Meeting Minutes

Date: 2025-08-07

Week: W3, Thursday

Type (client/internal/mentor): client

Meeting Name/Purpose: Weekly Client Update Meeting

Time: 0930-0950

Location / Platform: Teams

Attendees:

- ☑ Sirui Li (Client)
- ☐ Wei Liu (Unit Coordinator)
- ☐ Jichunyang Li (Project Facilitator)
- ☑ Franco Meng (Student)
- ☑ RuiZhe Wang (Student)
- ☑ Aswathy Mini Sasikumar (Student)
- ☑ Nirma Rajapaksha Senadherage (Student)
- ☐ Cedrus Dang (Student) - Absent: Driving Test
- ☑ Laine Mulvay (Student)

Agenda

1. Review and discuss Francos Local Project Proposal Notes
2. Clarify technical direction and dataset usage
3. Discuss expectations and timeline for proposal submission

1. Project Proposal notes (Cedrus' Diagram and Franco's Notes)

- Franco presented the Team 9 draft proposal and Cedrus' flowchart.
- Cedrus' diagram illustrated a **multi-agent system** where each agent (e.g., for query analysis, table identification, record selection) plays a specialized role.
- Siuri confirmed that agents can be implemented using:
    - Fine-tuned LLMs
    - Prompt engineering
- The system should be **sequential**, with each agent passing information to the next.

2. Dataset Clarification (Spider)

- The project will use the **Spider dataset**, preferably **version 1** for ease of use.
- Spider 2 is also acceptable if the team chooses.
- The dataset includes **natural language questions** with **difficulty labels** (easy, medium, hard).
    - Siuri recommends focusing on **easy and medium** difficulty questions during development.
- Discussed two dataset split strategies:
    - **Example split**: same database can appear in both train/test (not recommended).
    - **Database split**: each database appears in only one of train or test — **this is the preferred split**.
- No strong preference from Siuri on whether to use:
    - Local data files
    - Snowflake for database access
- There was a discussion on whether the system should identify the correct **database** as well — Siuri said yes.
- Select **easy and medium difficulty questions** from the dataset for initial development. The focus is on building the pipeline rather than accuracy.

3. Proposal Draft and Submission Timeline

- The proposal must be **sent to Siuri by Tuesday** so she can provide feedback.
- Final submission is due **Friday**.
- Proposal should include:
    - Overview of the agents and their roles

- Whether schema data is used and how
- Which GPT models or APIs the team wants to use (so Siuri can provide keys)
- Dataset choice (Spider v1/v2) and selected question difficulty

Next Steps Needed

1. **Specify which GPT models and API keys** are required; include this information in the proposal.
2. **Confirm the dataset version** (Spider v1 or v2) and preferred data access method (local files or Snowflake).
3. **Decide on and allocate tasks** for drafting the proposal.
4. **Send the draft proposal to Siuri by Tuesday** for feedback.
5. **Incorporate Siuri's feedback and submit the final proposal by Friday**.
6. **Finalize the multi-agent system flow** by clearly defining the sequential pipeline between agents.
7. **Decide whether to include schema visualization** in the UI for added explainability (to be considered later).

Next Meeting Name:

- **Date & Time:** 12-2pm Thursday 7th August (same day)
- **Location / Platform:** CITS5553 Lab
- **Next Meeting's Minutes to be prepared by:** ?

This Meetings Minutes prepared by: Laine Mulvay