Meeting 0424

Meeting Topic: Information Knowledge System Retrieval Demo Meeting

Meeting Time: April 25 (Friday), 11:01 – 11:28 (GMT+08)

Participants: @Zhou Junwei @Chen Wenhao @Jiang Qiyi @Zhu Qicheng @Luke Anderson

Note: This intelligent meeting summary was generated by AI and may contain inaccuracies. Please review with caution before use.

Summary

The meeting discussed the demo requirements and related work for document retrieval, including data collection, architecture design, and future development plans. Key topics included:

- **Development Requirements and Plans for the Information Knowledge System Retrieval Demo**

    - **Confirmed Demo Requirements**: Users input a query to quickly retrieve the most relevant documents from a large local file set. This involves both semantic retrieval and hierarchical classification-based retrieval.

    - **Data Collection and Labeling**: Thousands of files simulating real users need to be collected. Data collection comes first, followed by analysis for labeling strategies, keeping generality and scalability in mind.

    - **Architecture Design Considerations**: Developers are to think about the system architecture over the next two days and align with colleagues. Current tasks should be paused to prioritize this.

    - **Model Training Plan**: Open-source tools will be used initially. Fine-tuning may be considered later, but current data is not real business data, so training now is of limited value.

    - **Version Differences**: The Demo has both a Mac local version and an online version. The main difference lies in backend deployment, while the frontend is largely similar.

    - **Future Development Plan**: The goal for May is to develop a RAG (Retrieval-Augmented Generation) version based on the current demo, enabling direct answers to user queries.

- **Discussion on the Multimodal Information Knowledge System Demo Development**
    - **Office Usage**: The current office is available until the end of April. Regular holiday for May Day. Beijing office relocation is likely.
    - **Demo Delivery Deadline**: The multimodal knowledge system demo is to be delivered to Professor Luo by May 9.
    - **Technical Approach for the Demo**: A two-tier retrieval system is adopted. Documents are tagged with hierarchical labels and stored in a MySQL table. Natural language queries are converted into SQL for initial filtering, followed by semantic search within the subset. Both local and online versions need to be deployed.
    - **Data Collection**: At least 1,000 realistic documents are to be collected. Each participant should list categories and examples, then collectively define the scope. Files will be collected separately, tagged hierarchically, anonymized, and should cover both work and personal life topics.

Note: This content is temporarily unavailable for display outside of Feishu Docs.

To-Do Items

- **Architecture Design Confirmation**: Consider the retrieval-side architecture design in light of current needs such as data collection, labeling strategy, and future RAG features; align thoughts with Jiang Qiyi.

- **Data Collection Plan**: Start from your own files, supplement with relevant content from online searches, and then analyze how to label the data.

- **File Collection & Organization**: Collect various types of files from your personal computer and phone, including life photos, recipe pictures, etc. No need to pre-classify; de-identify the data and organize them based on personal habits. Submit the files to Zhou Junwei (Frank) for later review and analysis.
    Mentioned participants: @Jiang Qiyi @Chen Wenhao @Zhu Qicheng @Luke Anderson

Note: This content is temporarily unavailable for display outside of Feishu Docs.

Related Links

- **Text Record**
    - *Information Knowledge System Retrieval Demo Meeting – April 25, 2025*