OmniQuery: Enabling Question Answering on Personal Memory by Augmenting Multimodal Album Data

Jiahao Li, Zhuohao (Jerry) Zhang, Jiaju Ma

UCLA UNIVERSITY of WASHINGTON Stanford University

OVERVIEW

- OmniQuery enables natural language interaction with users' past memory.
- Images and videos stored in personal album
  - **episodic memory**.
- Constructs **semantic memory** by processing multiple episodic memories that are semantically, temporally and spatially relate:
  - Personal facts
  - Personal preferences
  - Individual experiences

SYSTEM DESIGN

Augmenting Memory

1- Preprocessing
- Raw media (images and videos) -> structured text

2- Semantic Knowledge Inference
- Use LLMs to detect events happened on different time window:
- Revisit episodic memories to infer high-granularity semantic knowledge

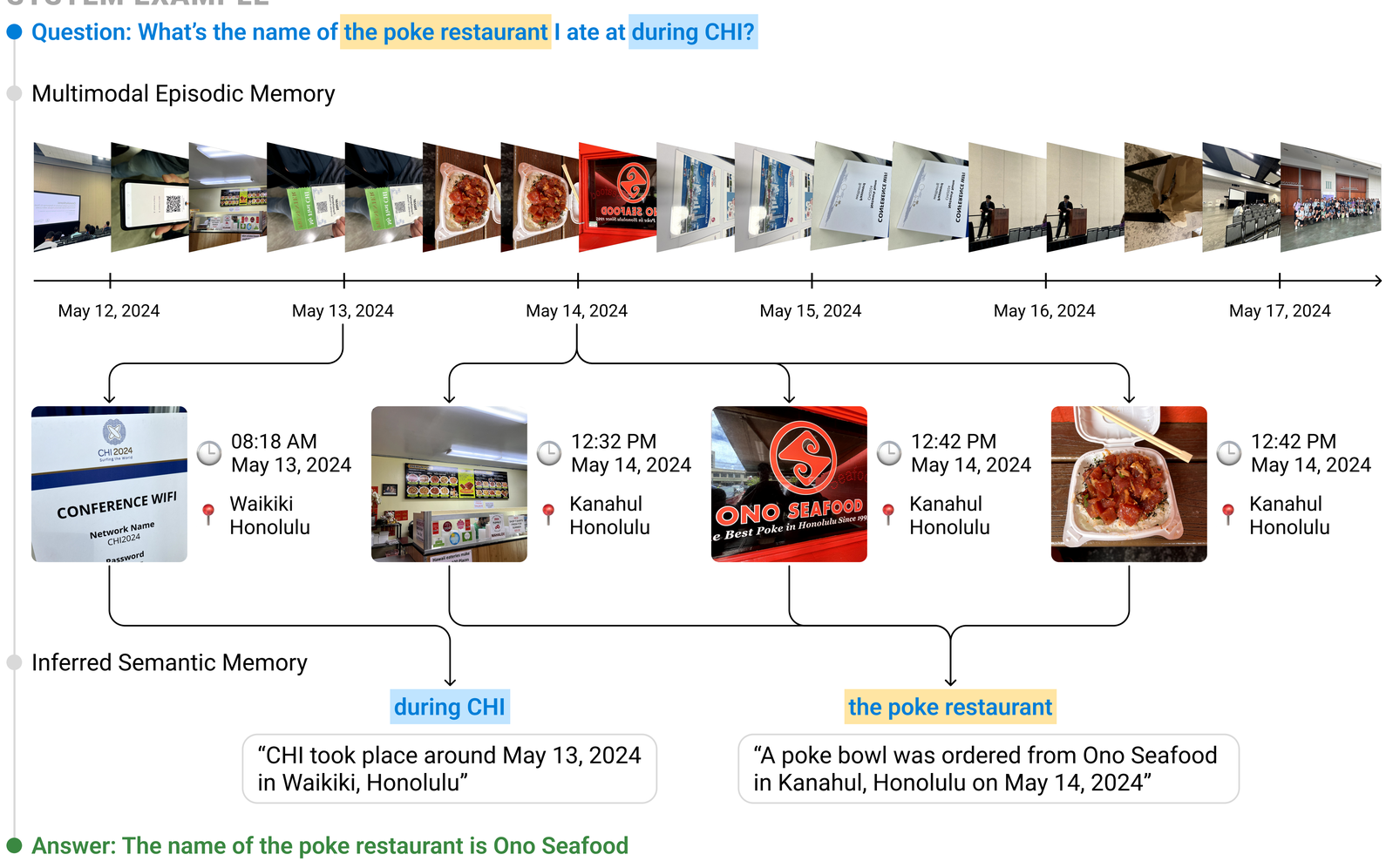
3- Indexing
- Save the memory (episodic + semantic) in vector databases

End-to-end Interactive System

- RAG-based Architecture
  - Query -> Text embeddings
  - Retrieve Top K in vector datases
  - Use LLMs to generate answers
  - Reranking also applies

UIST PITTSBURGH | 2024

SYSTEM EXAMPLE



SYSTEM ARCHITECTURE