Enabling Personal Communication for Voice-Based Health Assistants in Geriatric Care

Murilo Bellatini, May 2023

Thesis Exposé

This exposé outlines my master's thesis research proposal, supervised by Prof. Dr. Florian Matthes and Research Associate Phillip Schneider at Technical University of Munich (TUM) in collaboration with ALMA PHIL.

1 Motivation & Relevance

This thesis aims to enhance digital assistants by creating a dynamic graph-based user model that gathers personal information for tailored responses. Many assistants lack a personal touch and appear robotic, which is particularly critical in health assistants for elderly, aligning with our partner company's interests. Focus metrics are user satisfaction and engagement.

2 Related Work

The related work can be categorized into two pillars:

1. **Personal Knowledge Graphs (PKG)** [2] are crucial for tailored and engaging conversational agents. They store entities (e.g., family members, hobbies) and their relationships to the user, allowing for continuous extension and curation, see Fig. 2. Previous work [5] has addressed challenges in entity disambiguation and linking.
2. **Retrieval Augmented Generation (RAG)** [3] combines Language Models (LMs) with retrieval techniques to customize responses. Retrieval techniques retrieve relevant information, enhancing the generation process with contextual information. **Prompt Templates** [4] generate personalized responses using placeholders.

3 Solution Approach

Our solution approach, outlined in Appendix Figure 1, includes:

1. **Chat Interface**: Providing a user-friendly and accessible mobile interface. Developed using Python for agility and simplicity.
2. **Retrieval and Recommendation Module**: Seamlessly combining rule-based systems with existing Language Models (LMs) like OpenAI ChatGPT [1] to create a flexible Dialogue Manager, personalized recommendations for intent handling and coherent text generation.
3. **Knowledge Management**: Building PKG through Entity Extraction, Linking, and Disambiguation techniques from GraphWOZ [5]. Incorporating grounding and follow-up questions to address ambiguity and information gaps. Storing memory in document database, see Fig. 1.

4 Research Questions

This research addresses the following key questions:

(a) Which concepts and entities should be included in the data model for effective personalization and user engagement?
(b) What information extraction techniques can be employed to populate the PKG with pertinent information?
(c) How can the extracted knowledge be integrated into the conversational framework to enable personalized responses?
(d) What evaluation methods can be used to assess the performance and effectiveness of the developed system?

5 Expected Outcome

This research aims to develop a system with:

1. **Dynamically constructed PKG** that maps user personal data. Challenges in entity disambiguation, retrieval, and sensitive topic handling will be addressed.
2. **Recommendation modules** that contextualize conversation aiming to enhance user engagement and satisfaction.

6 Evaluation

The evaluation encompasses the following methods:

1. **PKG Reconstruction with Simulations**: Simulate user conversations using LMs like GPT3, with LMs acting as fake users referencing a PKG. Reconstruct the PKG using these simulations and compare it to the reference PKG, assessing accuracy and quality with classification metrics.
2. **Real User Engagement**: Gather qualitative ratings and quantitative metrics from testers, including app usage time, to evaluate user engagement.

7 Research Phases

The research progresses through the following stages:

1. **PKG Data Model Definition**: Defining the PKG data model to facilitate contextual conversations.
2. **Dynamic Population of PKG**: Employing information extraction techniques to dynamically populate the PKG with personal information shared during conversations.
3. **Recommendation Module Development**: Integrating context from the PKG into conversations using techniques like RAG (Retrieval Augmented Generation).

8 Conclusion

This research aims to develop a digital health assistant that automatically creates a PKG as a user model, retrieves relevant entities, and facilitates contextually appropriate conversations, with the goal of enhancing user engagement.

References

[1] OpenAI Chat API Documentation. https://platform.openai.com/docs/guides/chat, 2023. Accessed on 30 May 2023.

[2] Krisztian Balog and Tom Kenter. Personal knowledge graphs: A research agenda. In *Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval*, ICTIR '19, page 217–220, New York, NY, USA, 2019. Association for Computing Machinery.

[3] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. Retrieval-augmented generation for knowledge-intensive nlp tasks, 2021.

[4] Python Langchain. Python langchain: Documentation. https://python.langchain.com/en/latest/.

[5] Nicholas Thomas Walker, Stefan Ultes, and Pierre Lison. Graphwoz: Dialogue management with conversational knowledge graphs, 2022.

9 Appendix

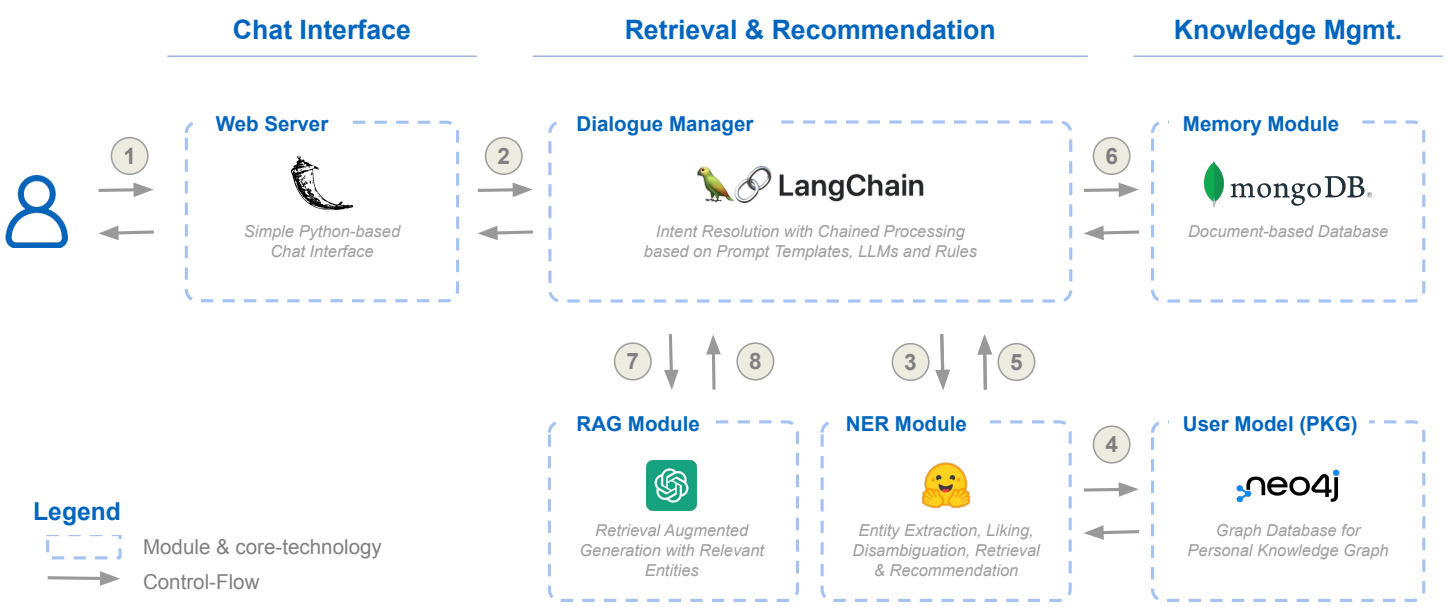

Figure 1: **Proposed System Overview**: The system operates as follows: (1) Users interact with the chat bot, initiating conversations with open-ended questions like "How is your day going?". (2) User inputs are processed by the **Dialogue Manager**, implemented in LangChain, which seamlessly integrates multiple LLMs, including OpenAI ChatGPT, and rule-based systems, offering flexibility and control. (3) The **Named Entity Recognition (NER)** module identifies and extracts entities from user input. (4) Extracted entities are compared against the user's **Personal Knowledge Graph (PKG)**, effectively managing duplicates and related entities. (5) Confirmed entities, as well as those in doubt, are sent back to the Dialogue Manager, each accompanied by a flag indicating its status for subsequent resolution or clarification. (6) Relevant **memory** from previous interactions is retrieved, initially time-capped for the last n turns, with potential expansion to include entity indexed search or vector-based search. (7) Prompt Templates are utilized to generate instructions for ChatGPT, incorporating the relevant entities. 8. **Retrieval Augmented Generation (RAG)** produces tailored responses that are delivered to the user.

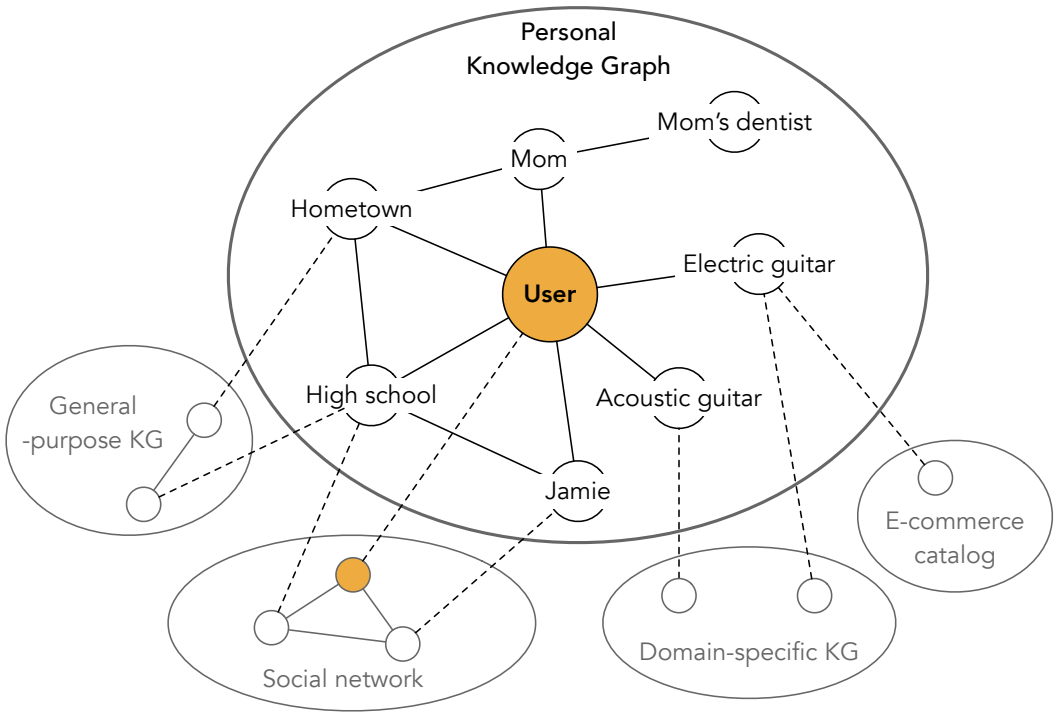

Figure 2: **Personal Knowledge Graph Example** from Google Research [2]