2024 15th International Conference on Information and Communication Technology Convergence (ICTC) | 979-8-3503-6463-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICTC62082.2024.10826816

Design and Implementation of Counseling Chatbot using Knowledge Graph-Based RAG

Jongmyeong Lee
dept. Autonomous Intelligent System
Korea Electronics Technology Institute
Gyeonggi-do, Republic of Korea
bsljm2002@keti.re.kr

Hyoseon Kye
dept. Autonomous Intelligent System
Korea Electronics Technology Institute
Gyeonggi-do, Republic of Korea
hs.kye@keti.re.kr

Jeehyeong Kim
dept. Autonomous Intelligent System
Korea Electronics Technology Institute
Gyeonggi-do, Republic of Korea
jkim8@keti.re.kr

***Abstract*—The physical constraints of offline counseling and the burden of face-to-face counseling have resulted in a demand for online counseling platforms. Therefore, we propose a psychological counseling chatbot system based on the Retrieval-augmented Generation (RAG) framework using a knowledge graph to provide real-time and non-face-to-face specialized psychological counseling. The proposed system can support systematic data management and retrieve information based on the user's conversation records and domain knowledge to generate customized answers based on accurate expertise.**

***Index Terms*—knowledge graph, retrieval-augmented generation, large language models**

I. Introduction

Following the COVID-19 pandemic, the proportion of South Korean citizens classified as being at high risk for depression has increased more than fivefold, making psychological counseling and mental health care essential [1]. Traditional methods such as hospital visits and offline counseling centers often face challenges related to accessibility, including physical location, cost, and time constraints, preventing individuals from receiving counseling at their preferred times. Consequently, there has been a growing demand for online platforms. This study proposes an LLM-based psychological counseling chatbot system utilizing knowledge graph-based Retrieval-augmented Generation (RAG) to provide real-time, non-face-to-face professional psychological counseling. To address the limitations of existing LLM models, such as difficulties in providing up-to-date information, challenges in highly specialized domains, and issues related to hallucinations arising from probability-based text generation, we applied knowledge graph-based RAG [2], [3].

II. Proposed System

We propose a psychological counseling chatbot system based on the RAG framework using a knowledge graph to enhance the performance of LLM. The proposed system consists of domain knowledge generation and storage, an RAG framework, and a chatbot interface that provides question-and-answer to users, which are discussed in detail as follows.

A. Domain Knowledge Generation and Management

We generate N sets of domain expertise in psychology using ChatGPT 4o, one of the LLMs. The prompt is entered so that ChatGPT 4o considers itself a psychologist and repeatedly generates psychological expertise knowledge data. The generated knowledge data is stored in the knowledge graph, a DB that improves knowledge retriever, which is one of the RAG functions. Part of the knowledge graph on psychology expertise data is shown in Fig. 2.

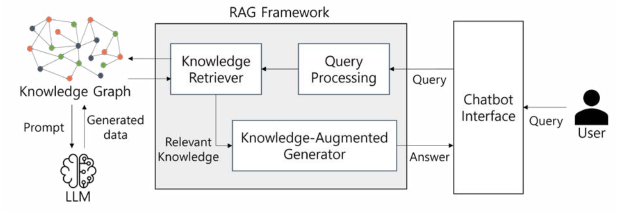


Fig. 1. Architecture of the proposed system

In the RAG framework, vector DB is mainly used as a database. The vector DB is a database optimized for storing and searching embedded vectors. It has the advantage of scalability to quickly search for data that is semantically similar to a user's question in the DB and to effectively manage and process large-scale vector data. However, the vector DB has several limitations. In the case of high-dimensional vector data, the vector DB can increase processing costs and slow response time. It is also difficult to secure the accuracy of the search of vector DB because vector DB returns data based on the similarity between the embedding vector and the query. This limitation makes it impossible for the vector DB to consider the minute semantic difference of terms and information used in the specialized field. Since psychological counseling requires high expertise, we use a knowledge graph as a database for the RAG framework instead of a vector DB. A knowledge graph is a knowledge structure in which information is expressed as edges and nodes. It is the most advantageous database for accumulating and delivering knowledge by storing highly correlated information in the form of a graph. Due to these advantages, knowledge graphs are used in various fields such as data integration, search engine optimization, recommendation systems, and question-and-answer systems.

To express the generated knowledge data in text format as a knowledge graph structure, we implement the knowledge

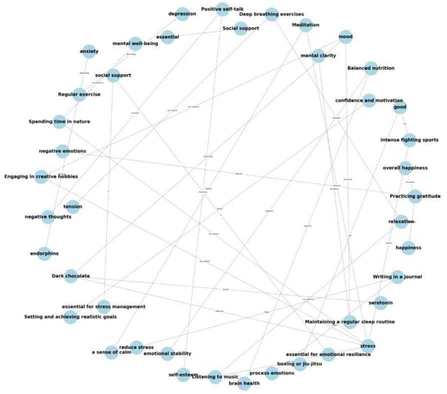


Fig. 2. Part of the Knowledge graph

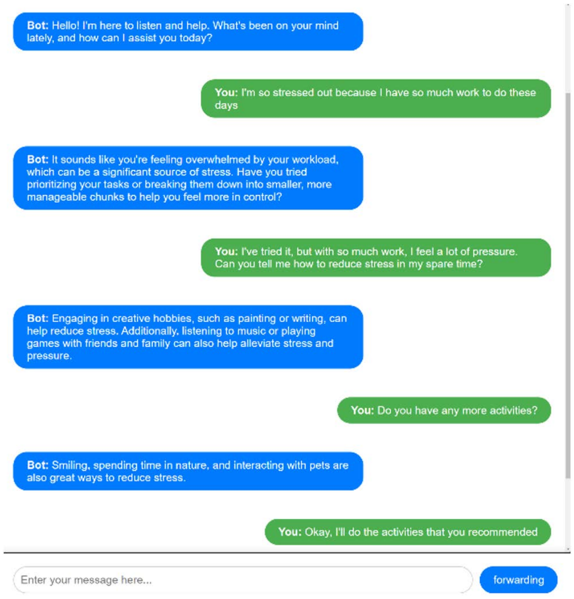
graph based on the Resource Description Framework (RDF) schema which is useful for natural language processing. The RDF schema is a language standard developed by the World Wide Web Consortium (W3C) for expressing resource information on the web. It consists of a triple composed of a subject, a predicate, and an object. The triple represents the subject and the object in a node, and the predicate in a directional edge. We use LlaMA3 70B Instruct [4] to transform the expertise knowledge data into the RDF schema form and store it in the knowledge graph.

B. Knowledge Graph-Based RAG Framework

Our proposed RAG framework can provide professional and customized counseling by storing expertise knowledge and user-specific object conversation data. The RAG framework consists of query processing, a knowledge retriever, and a knowledge-augmented generator. When a user's question (query) is input in the chatbot interface, the query processing part simplifies the query and converts it into a triple form. The converted query is stored as object data in the knowledge graph while the knowledge graph returns related knowledge data when information matching the query is searched and found. The object conversation data and related knowledge data are summarized and used as a prompt of knowledge-augmented generator based on the LLaMA3 70B Instruct model to generate an answer and provide it to the user through the chatbot interface. When related knowledge data cannot be found in the knowledge graph, the knowledge-augmented generator generates an answer by itself. Therefore, we can generate customized answers with expertise. We provide the chatbot interface through FastAPI and implement a real-time chat function through a web socket.

III. System Implementation

We implement the proposed psychological counseling chatbot system based on the RAG framework using a knowledge graph. Fig. 3 shows the demonstration of psychological counseling in the chatbot interface of the proposed system. We generated 65 types of domain knowledge data suitable for psychological counseling and stored them in the knowledge graph. When the user accesses the chatbot system at first,



Bot: Hello! I'm here to listen and help. What's been on your mind lately, and how can I assist you today?

You: I'm so stressed out because I have so much work to do these days

Bot: It sounds like you're feeling overwhelmed by your workload, which can be a significant source of stress. Have you tried prioritizing your tasks or breaking them down into smaller, more manageable chunks to help you feel more in control?

You: I've tried it, but with so much work, I feel a lot of pressure. Can you tell me how to reduce stress in my spare time?

Bot: Engaging in creative hobbies, such as painting or writing, can help reduce stress. Additionally, listening to music or playing games with friends and family can also help alleviate stress and pressure.

You: Do you have any more activities?

Bot: Smiling, spending time in nature, and interacting with pets are also great ways to reduce stress.

You: Okay, I'll do the activities that you recommended

Fig. 3. Demonstration of psychological counseling chatbot

the chatbot gives a welcome comment. After that, the chatbot naturally asks the user what kind of concerns the user has, beginning the consultation. In Fig. 3, the user asked how to relieve the stress because the user had a lot of work. We can confirm that our system properly provides a way to help relieve stress based on expertise knowledge data.

IV. Conclusion

We propose and implement a psychological counseling chatbot system based on the RAG framework using a knowledge graph. The proposed RAG framework stores domain knowledge and the user's conversation history, enabling customized psychological counseling based on expertise. The users also can receive real-time and non-face-to-face counseling when they want without physical constraints. For future work, we will focus on further expanding domain knowledge to improve the performance of chatbots and operating the system autonomously.

Acknowledgment

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (2022-0-00043, RS-2022-II220043,Adaptive Personality for Intelligent Agents)

References

[1] H. Lee, D. Choi, and J. J. Lee, "Depression, anxiety, and stress in Korean general population during the COVID-19 pandemic," Epidemiology and Health, vol. 44, Jan 2022.

[2] P. Lewis, et al. "Retrieval-augmented generation for knowledge-intensive nlp tasks," Advances in Neural Information Processing Systems 33, 2020, pp. 9459-9474.

[3] S. Ma, et al. "Think-on-Graph 2.0: Deep and Interpretable Large Language Model Reasoning with Knowledge Graph-Guided Retrieval," 2024, arXiv preprint arXiv:2407.10805. [Online]. Available: https://arxiv.org/abs/2407.10805

[4] A. Dubey, et al. "The llama 3 herd of models," 2024, arXiv preprint arXiv:2407.21783.