6-11-2024

Leveraging Large Language Models for Effective Organizational Navigation

Haresh Chandrasekar

Srishti Gupta

Chun-Tzu Liu

Chun-Hua Tsai

Leveraging Large Language Models for Effective Organizational Navigation

HARESH CHANDRASEKAR, University of Nebraska at Omaha, USA
SRISHTI GUPTA, University of Nebraska at Omaha, USA
CHUN-TZU LIU, University of Nebraska at Omaha, USA
CHUN-HUA TSAI, University of Nebraska at Omaha, USA

The advent of the internet has significantly enhanced accessibility to information, facilitating the engagement of diverse communities with online resources. Despite the abundance of information available, navigating the structures of large organizations and effectively digesting essential personalized information remains a challenge. Consequently, individuals may be deterred from extracting valuable insights from already available resources. This paper addresses this issue by integrating a university's official website into an AI chatbot powered by large language models (LLMs). We demonstrate use cases to provide information tailored to general information-seeking and personalized information needs for college major selection. We present a novel approach for individuals to gain insights into large organizations via interactive conversation. Based on our system demonstration, we further delve into the role of generative AI in synthesizing vast organizational datasets into user-friendly formats accessible to the public and its implications for E-government and open government research.

CCS Concepts: • **Computing methodologies** → **Natural language processing**; • **Information systems** → **Web searching and information discovery**.

Additional Key Words and Phrases: Information Seeking, Chatbot, University

ACM Reference Format:
Haresh Chandrasekar, Srishti Gupta, Chun-Tzu Liu, and Chun-Hua Tsai. 2024. Leveraging Large Language Models for Effective Organizational Navigation. In *25th Annual International Conference on Digital Government Research (DGO 2024), June 11–14, 2024, Taipei, Taiwan.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3657054.3657272

INTRODUCTION

Online information-seeking practices are increasing in prevalence. Having social media platforms, official websites, or online forums to publish and share organizational information with citizens, users, or customers is not a novel idea. These abundant information sources rely on users' capability and literacy to navigate, search, localize, and digest this semi-structured information for their needs, such as finding online personal health information or accessing information from government portals [3, 5]. As the volume of information grows exponentially, the importance of compiling and organizing data has become increasingly prominent. Developments in generative AI, such as large language models (LLMs), offer solutions to address this problem [2].

This paper aims to present a prototype using an AI chatbot to navigate an organization's official website. We utilize compiled organizational data and embed the data into a large language model using the retrieval-augmented generation (RAG) method [1]. Specifically, we downloaded and archived a university website as the candidate dataset. Given the complexity of the university website structure, such as the volume of information, format, and information organization, incoming students often miss out on opportunities they could pursue if they delve deeper. This can be daunting for those who are unsure of what they seek, which is a common challenge for online information seeking [3]. Based on our prototype, we present use cases illustrating how such an AI tool could assist incoming college students in understanding the college's programs and help them make personalized decisions when selecting their majors.

Use the following context as your learned knowledge, inside <context></context> XML tags.

<context>
We offer programs of study in the following areas: Biomedical Informatics & Bioinformatics Computer Science Computer Science Education Cybersecurity Data Science (M.S.) IT Innovation Management Information Systems Executive Master of Science in Information Technology Computing & Information Science (Ph.D.)
</context>

When answer to user:
- Please provide a short and concise answer.
- If you don't know, just say that you don't know.
- If you don't know when you are not sure, ask for clarification.
Avoid mentioning that you obtained the information from the context. And answer according to the language of the user's question.

User:
Can you tell me the program offer in this college?

Fig. 1. The system utilizes a prompt template for user interactions. In this template, the green-highlighted context section is dynamically inserted based on the indexed context from the archived dataset. User input questions are highlighted in yellow. The remaining text serves as a fixed component across all prompts.

METHOD

Cyotek WebCopy [1], a website scanning tool, was used to download all web page contents from the College of Information Science & Technology, University of Nebraska at Omaha (UNO)[2]. We archived the target websites locally which enabled us to obtain a comprehensive snapshot as the dataset used in this demonstration. After archiving the dataset, we conducted a data-cleaning process. We utilized BeautifulSoup[3], a Python library, to capture the HTML body and remove other markup languages and codes that are not intended for human reading. All filtered contents were stored in a single text file. The final dataset contains 3,185,038 words.

To create a chatbot app, we utilized Dify[4], a development framework for building generative AI applications. We adopted the Llama 2 70B model, an open-source LLM released by Meta[5], as the inference model. Cohere[6] was used as the embedding model to

1 https://www.cyotek.com/cyotek-webcopy

2 https://www.unomaha.edu/college-of-information-science-and-technology/index.php

3 https://beautiful-soup-4.readthedocs.io/en/latest/

4 https://dify.ai/

5 https://llama.meta.com/

6 https://cohere.com/

index the college dataset in a local vector database. The dataset was indexed in chunks (N=500), resulting in 6,943 paragraphs, which will serve as the context embedded in the prompt sent to the language model. These embeddings could assist the LLMs in answering organization-specific questions, such as "*What programs does this college offer?*" Such questions might not be answerable by general-purpose LLM tools such as ChatGPT, but our prototype will be able to respond by inserting a feasible context into the prompt. A sample prompt template with embedded context can be found in Figure 1.

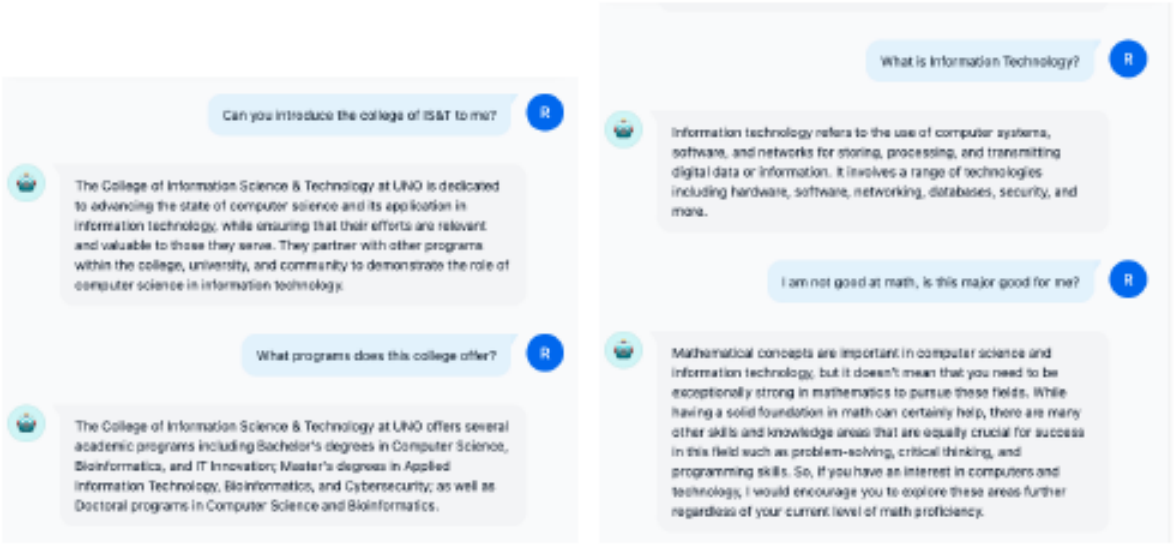

Fig. 2. Sample conversations generated using the proposed app: General information-seeking about the college (left) and personalized major selection by inquiring about the student's background (right). These responses indicate that the proposed chatbot could effectively understand and respond to users' inquiries.

RESULTS

We demonstrate two use cases to test the effectiveness of our proposed prototype. The first use case involves querying general information about the college. We submitted prompts to inquire about an introduction to the college and followed up with a question about the programs offered. The sample conversation is depicted in Figure 2. We observed that the chatbot could successfully understand and respond to user requests. The response was based on the given context, the college's archived dataset, making the response specific to the organization's website for the users. In the conversation, the user did not need to specify the

university name, and the app could correctly understand the user's intent to inquire about college information in the specific college context.

Our second use case aims to demonstrate whether the app can address information needs based on tailored, personalized circumstances. We submitted a prompt to inquire about the Information Technology program. Following the response, we further asked whether the student, if not proficient in math subjects, would find this major suitable. We observed that the app could accurately respond to the prompt and provide an encouraging message to the user, urging them to consider the importance of math in selecting this major. This use case highlights the superiority of using the app to navigate the degree program, as it might be challenging to locate or comprehend such responses by simply searching online. It relies on users' literacy to digest this information or would require a mediator to assist them in understanding it. Our attempt indicated the feasibility of using language models for such information-seeking needs.

CONCLUSIONS

We demonstrated that the LLM-powered AI chatbot effectively addressed diverse user queries. Specifically, the AI chatbot exhibited the capability to provide accurate responses to inquiries related to university websites. Moreover, it possesses the ability to personalize responses upon request. In our use cases, the chatbot can address inquiries about the organization, providing insights accessible to users by simply responding to their inquiries to assist in the decision-making process regarding college majors. Our prototype and use cases demonstrate the use of LLM to improve the traditional barriers of information seeking, behavior, and organization that might hinder users from seeking information online due to their information literacy limits.

Our exploration will contribute positively to the Digital Government Research (DGO) community by showcasing how generative AI techniques can navigate organizational datasets, such as governmental social media, open data portals, and official websites. Our work aligns with efforts to utilize LLMs in open government data and highlights the benefits and public values that can be generated from them [4]. For instance, the high usability and

capability to assist users in personalized ways represent an active research area aimed at increasing accessibility in E-government for citizens [5]. We are also aware of the issues and limitations of using LLMs in the presented use cases, such as mismatching, data quality, and the effects of hallucination [6], which we believe could benefit from presentation and discussion from the DGO research community.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. 2153509.

REFERENCES

[1] Yunfan Gao, Yun Xiong, Xinyu Gao, Kangxiang Jia, Jinliu Pan, Yuxi Bi, Yi Dai, Jiawei Sun, and Haofen Wang. 2023. Retrieval-augmented generation for large language models: A survey. *arXiv preprint arXiv:2312.10997* (2023).

[2] Thashmee Karunaratne and Adenike Adesina. 2023. Is it the new Google: Impact of ChatGPT on Students' Information Search Habits. In *the 22nd European Conference on e-Learning ECEL 2023, Hosted by the University of South Africa, 26-27 October 2023.* Academic Conferences International (ACI).

[3] Susan A LaValley, Marc T Kiviniemi, and Elizabeth A Gage-Bouchard. 2017. Where people look for online health information. *Health Information & Libraries Journal* 34, 2 (2017), 146–155.

[4] Euripidis Loukis, Stuti Saxena, Nina Rizun, Maria Ioanna Maratsi, Mohsan Ali, and Charalampos Alexopoulos. 2023. ChatGPT Application vis-a-vis Open Government Data (OGD): Capabilities, Public Values, Issues and a Research Agenda. In *International Conference on Electronic Government*. Springer, 95–110.

[5] Ria Lyzara, Betty Purwandari, Muhammad Fadhil Zulfikar, Harry Budi Santoso, and Iis Solichah. 2019. E-government usability evaluation: Insights from a systematic literature review. In *Proceedings of the 2nd International Conference on Software Engineering and Information Management*. 249–253.

[6] Ariana Martino, Michael Iannelli, and Coleen Truong. 2023. Knowledge injection to counter large language model (LLM) hallucination. In *European Semantic Web Conference*. Springer, 182–185.