PENGEMBANGAN LAYANAN INFORMASI PUBLIK PROGRAM STUDI PENDIDIKAN TEKNIK INFORMATIKA UNDIKSHA MELALUI ASISTEN VIRTUAL BERBASIS LLM DAN RAG

OLEH

I Putu Agus Wendika Ferdiana, NIM 2215051017

Jurusan Teknik Informatika

ABSTRAK

Penelitian ini dilatarbelakangi oleh kesenjangan informasi, tingginya repetisi pertanyaan administratif, dan rendahnya minat baca mahasiswa terhadap teks panjang di Program Studi Pendidikan Teknik Informatika (PTI) Universitas Pendidikan Ganesha. Tujuan penelitian ini adalah mengembangkan asisten virtual layanan informasi publik yang mengintegrasikan Large Language Model (LLM) dan Retrieval-Augmented Generation (RAG) menggunakan model GPT-4o mini, Gemini 2.5 Flash, serta Llama 3.1 8B sebagai LLM pembanding. Penelitian ini menggunakan metode Research and Development (R&D) dengan kerangka kerja ADDIE. Hasil pengembangan menunjukkan bahwa arsitektur backend Flask dan vector database FAISS dapat beroperasi secara efisien dengan latensi manajemen dokumen kurang dari 3 detik dan respons kueri 28,40–33,89 detik. Evaluasi RAGAS menunjukkan bahwa GPT-4o mini unggul dibandingkan Llama 3.1 8B pada sebagian besar metrik, yaitu Faithfulness (0,8869 berbanding 0,7430), Context Precision (0,8093 berbanding 0,7915), Answer Correctness (0,6784 berbanding 0,5478), dan Answer Relevancy (0,6970 berbanding 0,6093). Namun, Llama 3.1 8B memperoleh nilai lebih tinggi pada Context Recall (0,9944 berbanding 0,9819), yang menunjukkan kemampuannya dalam menemukan informasi yang dibutuhkan dari konteks. Selain itu, perbandingan arsitektur menunjukkan bahwa Modular RAG lebih stabil dibandingkan Native RAG pada metrik Faithfulness (0,8869 berbanding 0,8198), Context Recall (0,9819 berbanding 0,9750), Context Precision (0,8093 berbanding 0,8075), dan Answer Relevancy (0,6970 berbanding 0,6668), meskipun Native RAG unggul pada Answer Correctness (0,7036 berbanding 0,6784). Kelayakan sistem teruji dengan nilai sempurna 1,00 (Sangat Valid) dari Ahli Isi dan Ahli Teknis/Usability, serta skor System Usability Scale (SUS) sebesar 81,11. Secara keseluruhan, sistem asisten virtual yang dikembangkan terbukti layak, memiliki tingkat kebergunaan sangat baik, dan efektif mendukung efisiensi layanan informasi akademik.

Kata Kunci: Asisten virtual, *Large Language Model*, *Retrieval Augmented Generation*, RAGAS, layanan informasi publik, PTI Undiksha.

DEVELOPMENT OF PUBLIC INFORMATION SERVICES FOR THE INFORMATICS ENGINEERING EDUCATION STUDY PROGRAM OF UNDIKSHA THROUGH AN LLM AND RAG-BASED VIRTUAL ASSISTANT

BY

I Putu Agus Wendika Ferdiana, NIM 2215051017

Department of Informatics Engineering

ABSTRACT

This research is motivated by the information gap, the high repetition of administrative questions, and the low reading interest of students in long textual information within the Informatics Engineering Education (PTI) Study Program at Ganesha University of Education. The objective of this research is to develop a public information service virtual assistant that integrates Large Language Models (LLMs) and Retrieval-Augmented Generation (RAG) using GPT-4o mini, Gemini 2.5 Flash, and Llama 3.1 8B as a comparative LLM. This research applies the Research and Development (R&D) method using the ADDIE framework. The development results show that the Flask backend architecture and FAISS vector database operate efficiently, with document management latency of less than 3 seconds and query response times ranging from 28.40 to 33.89 seconds. The RAGAS evaluation shows that GPT-4o mini outperforms Llama 3.1 8B in most metrics, namely Faithfulness (0.8869 vs 0.7430), Context Precision (0.8093 vs 0.7915), Answer Correctness (0.6784 vs 0.5478), and Answer Relevancy (0.6970 vs 0.6093). However, Llama 3.1 8B achieves a higher score in Context Recall (0.9944 vs 0.9819), indicating its ability to retrieve the required information from the provided context. Furthermore, the architectural comparison shows that Modular RAG is more stable than Native RAG in Faithfulness (0.8869 vs 0.8198), Context Recall (0.9819 vs 0.9750), Context Precision (0.8093 vs 0.8075), and Answer Relevancy (0.6970 vs 0.6668), although Native RAG performs better in Answer Correctness (0.7036 vs 0.6784). The system feasibility was validated with a perfect score of 1.00 (Highly Valid) from Content Experts and Technical/Usability Experts, along with a System Usability Scale (SUS) score of 81.11. Overall, the developed virtual assistant system is proven to be feasible, highly usable, and effective in supporting the efficiency of academic information services.

Keywords: *Virtual assistant, Large Language Model, Retrieval Augmented Generation, RAGAS, public information services, PTI Undiksha.*