

ABSTRAK

IMPLEMENTASI *RETRIEVAL-AUGMENTED GENERATION* PADA ASISTEN VIRTUAL UNTUK PENINGKATAN LAYANAN INFORMASI DALAM PENGEMBANGAN WEBSITE SIWUR BERBASIS TALL STACK

Ageng Praba Wijaya
H1D022016

Di era digital, kebutuhan pengguna terhadap akses informasi yang cepat, relevan, dan terpercaya melalui website semakin meningkat, namun metode konvensional seperti *Frequently Asked Questions* (FAQ) statis dan pencarian berbasis kata kunci seringkali belum mampu menjawab pertanyaan kompleks secara efisien; untuk menjawab tantangan tersebut, penelitian ini mengusulkan implementasi asisten virtual berbasis *Retrieval-Augmented Generation* (RAG) pada website Siwur yang dikembangkan menggunakan *TALL Stack* (*Tailwind CSS*, *Alpine.js*, *Laravel*, *Livewire*), di mana RAG dipilih karena mampu menggabungkan kekuatan pencarian berbasis basis pengetahuan terkurasi dengan kemampuan generatif *Large Language Model* (LLM) sehingga jawaban yang dihasilkan lebih relevan, kontekstual, dan dapat dipertanggungjawabkan; sistem kecerdasan buatan dibangun sebagai layanan *backend* terpisah menggunakan *FastAPI* (Python), dengan *Google Gemini Flash 2.0* sebagai LLM, *ChromaDB* sebagai *vector store*, dan *Sentence-Transformers* untuk *embedding*, sementara metode penelitian yang digunakan adalah *Rapid Application Development* (RAD) yang menekankan iterasi cepat dan keterlibatan pengguna; evaluasi kinerja dilakukan secara kuantitatif menggunakan kerangka kerja *RAGAS* (*Retrieval-Augmented Generation Assessment*) dan secara kualitatif melalui *User Acceptance Test* (UAT), di mana hasil evaluasi *RAGAS* menunjukkan nilai *Faithfulness* sebesar 0.7401 dan *Answer Relevancy* sebesar 0.8053 yang mengindikasikan bahwa sistem mampu menghasilkan jawaban yang faktual dan relevan, sedangkan nilai *Context Precision* sebesar 0.4279 dan *Context Recall* sebesar 0.4624 menunjukkan kemampuan sistem dalam mengambil konteks yang cukup relevan, serta pengujian UAT menghasilkan skor rata-rata 83% dengan kategori “Sangat Baik”, yang menegaskan bahwa integrasi RAG pada asisten virtual Siwur dapat meningkatkan kualitas layanan informasi, mengurangi halusinasi jawaban, dan dapat diterima dengan baik oleh pengguna untuk mendukung operasional toko.

Kata Kunci : *Retrieval-Augmented Generation*, *RAGAS*, Asisten Virtual, *TALL Stack*, *Large Language Model* (LLM)

ABSTRACT

IMPLEMENTATION OF RETRIEVAL-AUGMENTED GENERATION ON VIRTUAL ASSISTANT TO ENHANCE INFORMATION SERVICES IN THE DEVELOPMENT OF THE SIWUR WEBSITE BASED ON THE TALL STACK

Ageng Praba Wijaya
H1D022016

In the digital era, users' demand for fast, relevant, and trustworthy access to information through websites has increased significantly; however, conventional methods such as static Frequently Asked Questions (FAQ) and keyword-based search often fail to efficiently address complex queries. To address this challenge, this study proposes the implementation of a Retrieval-Augmented Generation (RAG)-based virtual assistant on the Siwur website, developed using the TALL Stack (Tailwind CSS, Alpine.js, Laravel, and Livewire). RAG is chosen for its ability to combine curated knowledge-base retrieval with the generative capabilities of a Large Language Model (LLM), enabling more relevant, contextual, and accountable responses. The artificial intelligence system is built as a separate backend service using FastAPI (Python), with Google Gemini Flash 2.0 as the LLM, ChromaDB as the vector store, and Sentence-Transformers for embeddings, while the research methodology adopts Rapid Application Development (RAD), emphasizing rapid iteration and user involvement. Performance evaluation is conducted quantitatively using the RAGAS (Retrieval-Augmented Generation Assessment) framework and qualitatively through a User Acceptance Test (UAT). The RAGAS results show a Faithfulness score of 0.7401 and an Answer Relevancy score of 0.8053, indicating that the system is capable of producing factual and relevant answers, while Context Precision 0.4279 and Context Recall 0.4624 reflect the system's ability to retrieve sufficiently relevant context. Furthermore, the UAT yields an average score of 83% in the "Very Good" category, confirming that the integration of RAG into the Siwur virtual assistant enhances the quality of information services, reduces answer hallucination, and is well accepted by users in supporting store operations.

Keywords: Retrieval-Augmented Generation, RAGAS, Virtual Assistant, TALL Stack, Large Language Model (LLM)