UNIVERSITAS
GADJAH MADA
**Aplikasi Chatbot Berbasis Web Menggunakan RAG (Retrieval-Augmented Generation) untuk Eksplorasi
Dokumen Akademik**
Rasyid Kusnady, Dr.Eng. Ir. Ganjar Alfian, S.T., M.Eng.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

INTISARI

APLIKASI *CHATBOT* BERBASIS WEB MENGGUNAKAN RAG (*RETRIEVAL-AUGMENTED GENERATION*) UNTUK EKSPLORASI DOKUMEN AKADEMIK

Rasyid Kusnady
22/505791/SV/21936

Pertumbuhan publikasi ilmiah yang masif menuntut efisiensi dalam peninjauan literatur. Oleh karena itu, teknologi *Artificial Intelligence* (AI) berperan krusial dalam membantu proses peninjauan literatur. Namun, penggunaan *Large Language Models* (LLM) secara langsung terkendala oleh fenomena halusinasi informasi. Penelitian ini mengembangkan sistem *chatbot* berbasis web menggunakan arsitektur *Retrieval-Augmented Generation* (RAG) dengan *generator* Qwen3-8B untuk eksplorasi dokumen ilmiah. Evaluasi dilakukan pada arsitektur *single-hop* dan *iterative multi-hop* dengan variasi ukuran *chunk*, nilai *top-k*, serta tiga model *embedding,* yaitu BGE-M3, Octen-Embedding-0.6B, dan Qwen3-Embeddings-0.6B menggunakan dataset QASPER. Hasil eksperimen menunjukkan konfigurasi optimal ditemukan pada *single-hop* RAG dengan *chunk* 1.000 karakter, *top-k* 10, dan *embedding* BGE-M3 yang menghasilkan *F1-Score* 0,5927 dalam waktu 1,63 detik, melampaui performa *iterative multi-hop* RAG dengan *chunk* 750 karakter, *top-k* 10, dan Octen-Embedding-0.6B dengan *F1-Score* 0,5864 dan waktu komputasi 4,01 detik. Sistem diimplementasikan dengan Next.js, FastAPI, dan MongoDB. Hasil pengujian menunjukkan keberhasilan pada 16 skenario *Black Box Testing* dan tingkat penerimaan pengguna sebesar 90% melalui *User Acceptance Testing* (UAT). Sistem ini diharapkan dapat membantu peneliti dalam analisis dokumen ilmiah secara lebih akurat dan efisien.

Kata kunci: *Retrieval-Augmented Generation*, *Chatbot*, LLM, BERTScore, QASPER.

UNIVERSITAS
GADJAH MADA
**Aplikasi Chatbot Berbasis Web Menggunakan RAG (Retrieval-Augmented Generation) untuk Eksplorasi
Dokumen Akademik**
Rasyid Kusnady, Dr.Eng. Ir. Ganjar Alfian, S.T., M.Eng.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

ABSTRACT

WEB-BASED CHATBOT APPLICATION USING RETRIEVAL-AUGMENTED GENERATION (RAG) FOR ACADEMIC DOCUMENT EXPLORATION

Rasyid Kusnady
22/505791/SV/21936

The massive growth of scientific publications demands efficiency in literature reviews. Therefore, Artificial Intelligence (AI) plays a crucial role in streamlining the process. However, the direct use of Large Language Models (LLM) is hindered by the phenomenon of information hallucination. This research develops a web-based chatbot system using Retrieval-Augmented Generation (RAG) architecture with the Qwen3-8B generator for scientific document exploration. Evaluation was conducted on single-hop and iterative multi-hop architectures with variations in chunk size, top-k values, and three embedding models: BGE-M3, Octen-Embedding-0.6B, and Qwen3-Embeddings-0.6B, utilizing the QASPER dataset. Experimental results show that the optimal configuration is achieved by single-hop RAG with a 1,000 character chunk size, top-k of 10, and BGE-M3 embedding, yielding an F1-Score of 0.5927 in 1.63 seconds, outperforming iterative multi-hop RAG with a 750 character chunk size, top-k of 10, and Octen-Embedding-0.6B, which achieved an F1-Score of 0.5864 in 4.01 seconds. The system is implemented using Next.js, FastAPI, and MongoDB. Testing results demonstrate success across 16 Black Box Testing scenarios and a 90% user acceptance rate through User Acceptance Testing (UAT). This system is expected to assist researchers in performing more accurate and efficient analysis of scientific documents.

Keyword: Retrieval-Augmented Generation, Chatbot, LLM, BERTScore, QASPER.