

## ABSTRAK

Kompleksitas sistem hukum di Indonesia menimbulkan tantangan signifikan dalam akses informasi hukum yang akurat dan mudah dipahami oleh masyarakat luas. Struktur hierarkis dokumen hukum yang terdiri dari berbagai tingkatan peraturan, ditambah dengan terminologi hukum yang formal, menyebabkan kesulitan bagi masyarakat non-ahli dalam mengakses dan memahami isi peraturan. Di sisi lain, *Large Language Model* (LLM) yang digunakan dalam sistem penjawab pertanyaan sering mengalami halusinasi, menghasilkan informasi yang tidak akurat ketika menghadapi pertanyaan spesifik di luar data pelatihan. Untuk mengatasi tantangan tersebut, penelitian ini mengembangkan sistem *Retrieval-Augmented Generation* (RAG) yang mengintegrasikan kemampuan pengambilan informasi relevan dengan pemahaman mendalam terhadap granularitas dokumen hukum Indonesia. Sistem dirancang untuk mempertahankan struktur hierarkis dokumen hukum (Bab, Pasal, Ayat, Huruf) dalam proses *chunking* dan *indexing*. Metodologi penelitian mencakup pengumpulan 15 dokumen hukum autentik dari portal JDIH yang meliputi Undang-Undang, Peraturan Pemerintah, Peraturan Menteri, dan peraturan tingkat daerah. Arsitektur sistem menggabungkan pendekatan *dense retrieval* melalui WeaviateDB dengan *sparse retrieval* melalui Elasticsearch menggunakan metode *Reciprocal Rank Fusion* (RRF), serta Google Gemini 2.5 Flash sebagai layanan LLM. Evaluasi dilakukan menggunakan *framework* RAGAS pada *dataset* 150 pasangan pertanyaan-jawaban yang mencakup berbagai tipe *query*. Hasil evaluasi menunjukkan sistem berhasil melampaui seluruh target metrik: *faithfulness* mencapai 0.88 yang mengindikasikan sistem sangat jarang menghasilkan halusinasi, *answer relevancy* 0.88 menunjukkan jawaban yang sangat relevan, *context precision* 0.73 membuktikan mekanisme *retrieval* efektif, dan *context recall* 0.76 menunjukkan sistem mampu mengambil informasi lengkap dari *korpus* dokumen. Penelitian ini membuktikan bahwa implementasi sistem RAG untuk domain hukum Indonesia secara teknis *feasible* dan memberikan nilai signifikan dalam mendukung transformasi digital sektor hukum serta akses informasi hukum yang lebih inklusif di Indonesia.

**Kata Kunci:** Retrieval-Augmented Generation, Question Answering, Dokumen Hukum Indonesia, Large Language Model, Natural Language Processing, Sistem Informasi Hukum

## ABSTRACT

The complexity of the Indonesian legal system poses significant challenges in accessing accurate and easily understandable legal information for the general public. The hierarchical structure of legal documents consisting of various levels of regulations, combined with formal legal terminology, creates difficulties for non-expert communities in accessing and understanding the content of regulations. On the other hand, *Large Language Models* (LLMs) used in question-answering systems often experience hallucinations, producing inaccurate information when faced with specific questions outside their training data. To address these challenges, this research develops a *Retrieval-Augmented Generation* (RAG) system that integrates relevant information retrieval capabilities with deep understanding of Indonesian legal document granularity. The system is designed to maintain the hierarchical structure of legal documents (Chapter, Article, Verse, Letter) during the *chunking* and *indexing* processes. The research methodology includes collecting 15 authentic legal documents from the JDIH portal covering Laws, Government Regulations, Ministerial Regulations, and regional-level regulations. The system architecture combines *dense retrieval* approach through WeaviateDB with *sparse retrieval* through Elasticsearch using *Reciprocal Rank Fusion* (RRF) method, along with Google Gemini 2.5 Flash as the LLM service. Evaluation is conducted using the RAGAS *framework* on a *dataset* of 150 question-answer pairs covering various *query* types. The evaluation results show that the system successfully exceeds all metric targets: *faithfulness* reaching 0.88 indicating the system very rarely produces hallucinations, *answer relevancy* 0.88 showing highly relevant answers, *context precision* 0.73 proving effective retrieval mechanism, and *context recall* 0.76 demonstrating the system's ability to retrieve complete information from the document *corpus*. This research proves that the implementation of a RAG system for the Indonesian legal domain is technically *feasible* and provides significant value in supporting digital transformation in the legal sector as well as more inclusive access to legal information in Indonesia.

**Keywords:** Retrieval-Augmented Generation, Question Answering, Indonesian Legal Documents, Large Language Model, Natural Language Processing, Legal Information System