UNIVERSITAS
GADJAH MADA
**Rancang Bangun Sistem Manajemen Dokumen dengan Fitur Pencarian Cerdas Berbasis Retrieval-Augmented
Generation Menggunakan Large Language Model pada PT Parama Data Unit**
Muhammad Hanif Sauqi, Dr.Eng. Ir. Ganjar Alfian, S.T., M.Eng.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

INTISARI

RANCANG BANGUN SISTEM MANAJEMEN DOKUMEN DENGAN FITUR PENCARIAN CERDAS BERBASIS RETRIEVAL-AUGMENTED GENERATION MENGGUNAKAN LARGE LANGUAGE MODEL PADA PT PARAMA DATA UNIT

Muhammad Hanif Sauqi
22/495097/SV/20949

Dokumen operasional merupakan aset informasi penting yang mendukung kelancaran kegiatan dan pengambilan keputusan pada suatu perusahaan. Jumlah dokumen operasional pada perusahaan jasa penunjang industri minyak dan gas bumi terus bertambah, sehingga dibutuhkan sistem yang mampu memusatkan pengelolaan dokumen sekaligus mempermudah pencarian kembali informasi yang relevan. Pencarian berbasis kesamaan kata kunci yang umum digunakan pada sistem manajemen dokumen konvensional memiliki keterbatasan, karena dokumen yang menggunakan istilah berbeda dapat tidak ditemukan meskipun isinya relevan. PT Parama Data Unit, sebagai perusahaan yang mengelola dokumen operasional dan teknis proyek pengeboran dalam jumlah besar, membutuhkan sistem manajemen dokumen dengan kemampuan pencarian yang lebih kontekstual. Penelitian ini merancang dan membangun Sistem Manajemen Dokumen yang mengintegrasikan fitur manajemen dokumen konvensional dengan fitur pencarian cerdas (Smart Search) berbasis Retrieval-Augmented Generation (RAG) menggunakan Large Language Model, dengan alur RAG mencakup ekstraksi teks, text chunking, pembentukan embedding menggunakan model BGE-M3, penyimpanan vektor pada basis data PostgreSQL dengan ekstensi pgvector, retrieval berbasis cosine similarity, serta penilaian relevansi oleh model Qwen3 4B Instruct yang dijalankan secara lokal melalui Ollama. Tiga parameter operasional Smart Search, yaitu chunk size, similarity threshold, dan evidence count, ditetapkan melalui pengujian empiris satu per satu menggunakan 20 kueri uji dengan ground truth yang divalidasi bersama ahli domain, menghasilkan konfigurasi akhir berupa chunk size 700 karakter, similarity threshold 0,20, dan evidence count 5 potongan dengan Precision 0,887, Recall 0,976, dan F1-Score 0,923. Pengujian fungsional dengan metode black-box terhadap 22 skenario menghasilkan tingkat keberhasilan 100%, dan User Acceptance Testing dengan lima responden dari PT Parama Data Unit dan mitra operasionalnya menghasilkan skor 84,67% yang dikategorikan Sangat Baik, sehingga sistem dinyatakan layak digunakan untuk mendukung pengelolaan dan pencarian dokumen operasional perusahaan.

Kata kunci: Sistem Manajemen Dokumen, Retrieval-Augmented Generation, Large Language Model, Pencarian Cerdas, Text Chunking

UNIVERSITAS
GADJAH MADA
**Rancang Bangun Sistem Manajemen Dokumen dengan Fitur Pencarian Cerdas Berbasis Retrieval-Augmented
Generation Menggunakan Large Language Model pada PT Parama Data Unit**
Muhammad Hanif Sauqi, Dr.Eng. Ir. Ganjar Alfian, S.T., M.Eng.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

ABSTRACT

DESIGN AND DEVELOPMENT OF A DOCUMENT MANAGEMENT SYSTEM *WITH* RETRIEVAL-AUGMENTED GENERATION-*BASED* SMART SEARCH *USING* LARGE LANGUAGE MODEL*S AT PT PARAMA DATA UNIT*

Muhammad Hanif Sauqi
22/495097/SV/20949

Operational documents are an important information asset that supports the smooth running of activities and decision-making within a company. The volume of operational documents in oil and gas support service companies continues to grow, creating a need for a system capable of centralizing document management while easing the retrieval of relevant information. Keyword-based search, commonly used in conventional document management systems, has limitations, as documents using different terminology may go undiscovered despite being relevant. PT Parama Data Unit, as a company managing a large volume of operational and technical drilling project documents, requires a document management system with more contextual search capability. This study designs and develops a Document Management System that integrates conventional document management features with an intelligent search feature (Smart Search) based on Retrieval-Augmented Generation (RAG) using a Large Language Model, with the implemented RAG pipeline covering text extraction, text chunking, embedding generation using the BGE-M3 model, vector storage in a PostgreSQL database with the pgvector extension, cosine-similarity-based retrieval, and relevance assessment by the Qwen3 4B Instruct model, all running locally through Ollama. Three Smart Search operational parameters—chunk size, similarity threshold, and evidence count—were determined through empirical one-factor-at-a-time testing using 20 test queries with ground truth validated alongside domain experts, yielding a final configuration of chunk size 700 characters, similarity threshold 0.20, and evidence count 5 chunks, with a Precision of 0.887, Recall of 0.976, and F1-Score of 0.923. Functional testing using the black-box method across 22 scenarios achieved a 100% success rate, and User Acceptance Testing with five respondents from PT Parama Data Unit and its operational partner yielded a score of 84.67%, categorized as Very Good, indicating that the system is suitable for supporting the company's operational document management and retrieval activities.

Keywords: Document Management System, Retrieval-Augmented Generation, Large Language Model, Smart Search, Text Chunking