

PERANCANGAN SISTEM KNOWLEDGE MANAGEMENT BERBASIS LLM DAN RAG MENGGUNAKAN MODEL LLAMA3 DENGAN VECTOR DATABASE MONGODB PADA PT PRIMA INTEGRASI NETWORK

Teguh Prasetyo¹, Tri Ismardiko², Gerry Firmansyah³, Diah Aryani⁴

^{2,3,etc} Teknik Informatika, Fakultas Ilmu Komputer, Universitas Esa Unggul, Jakarta, Indonesia

Teguhpraz100@gmail.com

Abstract

In the context of rapidly evolving digital revolution, organizational knowledge management has become a determinant factor for enhancing operational productivity and competitive advantage of companies. PT Prima Integrasi Network experiences complex challenges in optimizing information management systematically and effectively, thus requiring advanced technological solutions in the form of artificial intelligence-based knowledge management systems that possess adaptive and responsive capabilities to organizational dynamic needs. This research aims to develop and implement a knowledge management system architecture that utilizes Large Language Model technology by integrating the Retrieval-Augmented Generation approach through the implementation of the LLaMA3 model combined with MongoDB vector database. The methodological approach applied in this research employs the Design Science Research framework that encompasses a series of systematic stages including comprehensive requirement analysis, system architecture design, technology implementation, and system performance evaluation. The constructed system enables the execution of information search and management based on semantic analysis through the implementation of vector search techniques with cosine similarity algorithms. The system evaluation process is conducted through comprehensive testing that measures crucial aspects including search result relevance levels, system response speed, and user satisfaction levels toward system functionality. Research findings demonstrate that the integration of LLaMA3 technology with MongoDB vector database within the RAG framework is capable of optimizing information search accuracy and significantly improving organizational knowledge management efficiency at PT Prima Integrasi Network. This system implementation shows potential as an innovative solution in modern organizational knowledge management transformation, with flexibility characteristics that enable further development according to specific company needs and development dynamics. The contribution of this research lies in the development of a knowledge management system model that integrates cutting-edge artificial intelligence technology with vector databases to optimize organizational information search and management processes. The developed system provides a robust technological foundation for digital transformation in corporate knowledge management with approaches that can be adapted and developed according to organizational need evolution.

Abstrak

Dalam konteks revolusi digital yang berkembang pesat, pengelolaan pengetahuan organisasi telah menjadi faktor determinan bagi peningkatan produktivitas operasional dan keunggulan kompetitif perusahaan. PT Prima Integrasi Network mengalami permasalahan kompleks dalam mengoptimalkan pengelolaan informasi secara sistematis dan efektif, sehingga memerlukan solusi teknologi canggih berupa sistem manajemen pengetahuan berbasis kecerdasan buatan yang memiliki kemampuan adaptif dan responsif terhadap dinamika kebutuhan organisasi. Penelitian ini bertujuan untuk mengembangkan dan mengimplementasikan arsitektur sistem manajemen pengetahuan yang memanfaatkan teknologi Model Bahasa Besar dengan mengintegrasikan pendekatan Pembangkitan Terintegrasi Pencarian melalui implementasi model LLaMA3 yang dikombinasikan dengan basis data vektor MongoDB. Pendekatan metodologis yang diterapkan dalam penelitian ini menggunakan kerangka kerja

Article History

Submitted: 24 July 2025

Accepted: 03 August 2025

Published: 04 August 2025

Key Words

Knowledge Management;
Large Language Model;
Retrieval-Augmented
Generation; LLaMA3;
MongoDB; Vector
Database

Sejarah Artikel

Submitted: 24 July 2025

Accepted: 03 August 2025

Published: 04 August 2025

Kata Kunci

Knowledge Management;
Large Language Model;
Retrieval-Augmented
Generation; LLaMA3;
MongoDB; Vector
Database

Riset Sains Perancangan yang mencakup serangkaian tahapan sistematis meliputi analisis kebutuhan komprehensif, perancangan arsitektur sistem, implementasi teknologi, dan evaluasi kinerja sistem. Sistem yang dikonstruksi memungkinkan pelaksanaan pencarian dan pengelolaan informasi berdasarkan analisis semantik melalui implementasi teknik pencarian vektor dengan algoritma kemiripan kosinus. Proses evaluasi sistem dilakukan melalui pengujian komprehensif yang mengukur aspek-aspek krusial meliputi tingkat relevansi hasil pencarian, kecepatan respons sistem, dan tingkat kepuasan pengguna terhadap fungsionalitas sistem. Temuan penelitian mendemonstrasikan bahwa integrasi teknologi LLaMA3 dengan basis data vektor MongoDB dalam kerangka kerja Pembangunan Terintegrasi Pencarian mampu mengoptimalkan akurasi pencarian informasi dan meningkatkan efisiensi pengelolaan pengetahuan organisasi di PT Prima Integrasi Network secara signifikan. Implementasi sistem ini menunjukkan potensi sebagai solusi inovatif dalam transformasi pengelolaan pengetahuan organisasi modern, dengan karakteristik fleksibilitas yang memungkinkan pengembangan lebih lanjut sesuai dengan kebutuhan spesifik dan dinamika perkembangan perusahaan. Kontribusi penelitian ini terletak pada pengembangan model sistem manajemen pengetahuan yang mengintegrasikan teknologi kecerdasan buatan terdepan dengan basis data vektor untuk mengoptimalkan proses pencarian dan pengelolaan informasi organisasi. Sistem yang dikembangkan memberikan fondasi teknologi yang kuat untuk transformasi digital dalam pengelolaan pengetahuan perusahaan dengan pendekatan yang dapat diadaptasi dan dikembangkan sesuai evolusi kebutuhan organisasi

PENDAHULUAN

Revolusi teknologi kecerdasan buatan telah mengubah paradigma fundamental dalam pengelolaan informasi dan pengetahuan organisasi di era digital kontemporer. Transformasi ini bukan lagi merupakan konsep futuristik melainkan telah menjadi realitas yang menggerakkan inovasi lintas sektor industri. PT Prima Integrasi Network, sebagai entitas bisnis yang beroperasi dalam domain sistem integrasi dan solusi teknologi informasi, mengakui bahwa kecerdasan buatan tidak sekadar berfungsi sebagai instrumen pendukung operasional, tetapi telah berevolusi menjadi fondasi strategis dalam membangun ekosistem layanan yang lebih cerdas, responsif, dan adaptif terhadap dinamika kebutuhan pengguna [1]. Dalam konteks persaingan bisnis yang semakin intensif, kapabilitas untuk menyajikan solusi inovatif dan efektif menjadi diferensiasi kompetitif yang menentukan, sehingga PT Prima Integrasi Network berkomitmen untuk memposisikan diri di garis terdepan perkembangan teknologi kecerdasan buatan ini. Komitmen strategis tersebut diwujudkan melalui investasi aktif PT Prima Integrasi Network dalam riset dan pengembangan di bidang kecerdasan buatan, dengan fokus utama pada teknologi Retrieval-Augmented Generation (RAG). RAG merepresentasikan pendekatan inovatif yang mengintegrasikan keunggulan Large Language Models (LLMs) dalam memahami dan menghasilkan bahasa alami dengan kemampuan untuk secara dinamis mengambil dan memanfaatkan informasi relevan dari sumber pengetahuan eksternal [2]. Sintesis teknologi ini memungkinkan sistem RAG untuk menyediakan respons yang tidak hanya akurat dan informatif, tetapi juga kontekstual sesuai dengan pertanyaan yang diajukan, menciptakan nilai tambah yang signifikan dalam berbagai aplikasi bisnis. Potensi strategis RAG yang diidentifikasi oleh PT Prima Integrasi Network mencakup peningkatan kualitas layanan pelanggan melalui sistem cerdas yang mampu memberikan jawaban instan dan akurat, mempermudah akses informasi internal perusahaan melalui sistem pencarian enterprise yang intuitif, serta mengotomatiskan pembuatan konten seperti laporan dan materi pemasaran yang pada akhirnya memungkinkan efisiensi waktu dan optimalisasi sumber daya [3].

Dalam upaya pengembangan sistem RAG yang handal dan aman, PT Prima Integrasi Network mengadopsi strategi implementasi menggunakan model LLM sumber terbuka LLaMA3. Keputusan strategis ini didasari oleh pertimbangan fundamental mengenai

pentingnya kontrol penuh atas data dan fleksibilitas dalam penyesuaian model sesuai kebutuhan spesifik pelanggan [4]. Selain itu, vector database MongoDB dipilih sebagai infrastruktur penyimpanan dan pengelolaan data pengetahuan yang digunakan oleh sistem RAG, memungkinkan pencarian informasi berdasarkan kemiripan semantik dan menjamin kinerja yang optimal. Kombinasi teknologi LLaMA3 dan MongoDB memungkinkan PT Prima Integrasi Network untuk menciptakan solusi RAG yang tidak hanya cerdas dan efektif, tetapi juga aman, terukur, dan mudah diintegrasikan dengan sistem yang ada, menciptakan proposisi nilai unik yang akan ditawarkan kepada pelanggan. Berdasarkan analisis kebutuhan dan potensi teknologi yang telah diidentifikasi, penelitian ini bertujuan untuk merancang arsitektur sistem Knowledge Management berbasis LLM dan RAG menggunakan model LLaMA3 dengan vector database MongoDB yang optimal untuk PT Prima Integrasi Network, dengan mempertimbangkan kebutuhan spesifik perusahaan dalam pengelolaan pengetahuan dan keamanan data. Penelitian ini juga mengimplementasikan sistem Knowledge Management tersebut di PT Prima Integrasi Network menggunakan model LLaMA3 yang dijalankan secara on-premise dan vector database MongoDB, dengan fokus pada integrasi yang efektif antara LLM, RAG, dan MongoDB, serta implementasi mekanisme keamanan data yang kuat. Selanjutnya, penelitian ini menguji dan mengevaluasi kinerja sistem Knowledge Management yang telah dirancang dan diimplementasikan, termasuk akurasi, efisiensi, waktu respons, kegunaan, dan tingkat keamanan data untuk memastikan sistem memenuhi kebutuhan PT Prima Integrasi Network. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi signifikan dalam pengembangan solusi kecerdasan buatan yang akan ditawarkan kepada pelanggan PT Prima Integrasi Network, serta memperkuat posisi perusahaan sebagai penyedia solusi inovatif dan terdepan di pasar, menjadikan penelitian ini bukan hanya sekadar eksplorasi akademis, tetapi juga langkah strategis dalam mewujudkan visi PT Prima Integrasi Network sebagai pemimpin dalam industri solusi teknologi informasi berbasis kecerdasan buatan [5].

METODE PENELITIAN

Penelitian ini mengimplementasikan metodologi penelitian dan pengembangan (Research and Development) dengan mengadopsi kerangka kerja Design Science Research (DSR) sebagai pendekatan utama. Pemilihan metodologi ini didasarkan pada karakteristik penelitian yang bertujuan untuk merancang dan mengembangkan artefak berupa sistem manajemen pengetahuan yang berbasis Large Language Model (LLM) dan Retrieval-Augmented Generation (RAG) dengan memanfaatkan model LLaMA3 dan basis data vektor MongoDB di lingkungan PT Prima Integrasi Network. Pendekatan DSR dipandang tepat karena memungkinkan pengembangan solusi teknologi yang fungsional dan praktis untuk mengatasi permasalahan nyata yang dihadapi oleh organisasi, khususnya dalam pengelolaan pengetahuan internal perusahaan [6].

Strategi pengumpulan data dalam penelitian ini menerapkan pendekatan campuran dengan menggabungkan data primer dan sekunder untuk memperoleh pemahaman yang komprehensif tentang kebutuhan sistem. Data primer diperoleh melalui tiga metode utama, yaitu wawancara mendalam dengan manajer teknologi informasi dan representasi pengguna di PT Prima Integrasi Network untuk memahami proses manajemen pengetahuan yang berlangsung saat ini, identifikasi kebutuhan pengguna, dan memperoleh umpan balik terhadap sistem yang dirancang. Selain itu, kuesioner atau survei disebarkan kepada karyawan PT Prima Integrasi Network setelah implementasi sistem untuk mengukur tingkat kepuasan pengguna terhadap fitur-fitur sistem, kemudahan penggunaan, dan relevansi informasi yang disediakan. Data primer juga diperoleh melalui pengujian sistem yang dilakukan untuk mengevaluasi kinerja sistem yang telah dirancang, memastikan fungsionalitasnya berjalan sesuai harapan, termasuk akurasi jawaban yang diberikan oleh LLM dan RAG, kecepatan pencarian informasi,

dan kemampuan MongoDB sebagai basis data vektor.

Objek penelitian ini mencakup empat komponen utama yang saling berkaitan dalam pengembangan sistem manajemen pengetahuan. Komponen pertama adalah proses manajemen pengetahuan di PT Prima Integrasi Network, yang meliputi kajian terhadap cara pengetahuan saat ini dikelola, disimpan, dan dibagikan dalam organisasi, termasuk identifikasi sumber-sumber pengetahuan, metode dokumentasi, dan mekanisme berbagi informasi yang ada. Komponen kedua adalah kebutuhan pengguna yang diidentifikasi melalui analisis kebutuhan informasi dan manajemen pengetahuan dari karyawan PT Prima Integrasi Network untuk memastikan sistem yang dikembangkan mampu memenuhi kebutuhan pengguna secara optimal dan tepat guna. Komponen ketiga adalah sistem manajemen pengetahuan yang dirancang, meliputi arsitektur sistem secara keseluruhan, fitur-fitur yang disediakan seperti pencarian informasi, tanya jawab, forum diskusi, dan antarmuka pengguna. Komponen keempat adalah evaluasi efektivitas model LLaMA3 yang diintegrasikan dengan teknik RAG dalam memberikan jawaban yang relevan dan akurat berdasarkan basis pengetahuan perusahaan, serta analisis kemampuan MongoDB sebagai basis data vektor dalam menyimpan dan mengambil informasi vektor dari dokumen perusahaan secara efisien.

HASIL DAN PEMBAHASAN

Implementasi Sistem Manajemen Pengetahuan Berbasis Model Bahasa Besar dan Generasi Berbantu Pengambilan

Penelitian ini berhasil mengembangkan sistem manajemen pengetahuan yang memanfaatkan teknologi model bahasa besar LLaMA3 terintegrasi dengan metodologi Generasi Berbantu Pengambilan (RAG) serta basis data vektor MongoDB. Arsitektur sistem yang dikembangkan menunjukkan integrasi yang koheren antara komponen-komponen utama, meliputi model bahasa besar LLaMA3 sebagai inti pemrosesan bahasa natural, kerangka kerja LlamaIndex untuk menghubungkan model dengan data eksternal, basis data vektor MongoDB untuk penyimpanan dan pencarian berbasis kesamaan semantik, serta antarmuka pengguna berbasis Streamlit untuk interaksi sistem. Implementasi ini menghasilkan alur kerja yang efisien dimana masukan pertanyaan pengguna melalui antarmuka Streamlit diteruskan ke sistem backend, dilanjutkan dengan pencarian informasi relevan oleh LlamaIndex pada basis data vektor MongoDB, kemudian hasil pencarian diproses oleh model LLaMA3 untuk menghasilkan jawaban kontekstual yang ditampilkan kembali kepada pengguna [7].

Evaluasi Kinerja Sistem Pencarian dan Akurasi Respons

Pengujian sistem menggunakan metodologi kotak hitam menunjukkan kinerja yang memuaskan dalam berbagai aspek operasional. Evaluasi terhadap lima pertanyaan uji coba menghasilkan tingkat relevansi jawaban dengan rata-rata skor 4,6 dari skala maksimal 5, mendemonstrasikan kemampuan sistem dalam memberikan respons yang akurat dan kontekstual. Analisis durasi pemrosesan menunjukkan rentang waktu eksekusi antara 23 hingga 33 detik, dengan pencarian sederhana memerlukan waktu rata-rata 23 detik sedangkan pencarian kompleks membutuhkan hingga 33 detik. Kecepatan respons ini menunjukkan efisiensi operasional basis data vektor MongoDB dalam menangani operasi pencarian berbasis kesamaan semantik. Sistem menunjukkan konsistensi dalam memberikan jawaban yang relevan untuk berbagai jenis pertanyaan, mulai dari informasi umum tentang kemampuan sistem hingga prosedur teknis spesifik seperti pengaturan ulang kata sandi email dan konfigurasi jaringan [2].

Analisis Komparatif dengan Pendekatan Konvensional

Perbandingan sistematis antara sistem yang dikembangkan dengan pendekatan manajemen pengetahuan konvensional menunjukkan keunggulan signifikan dalam beberapa

aspek kunci. Dari segi kecepatan pencarian, sistem berbasis model bahasa besar dan RAG mampu memberikan respons dalam waktu rata-rata di bawah 35 detik, kontras dengan pendekatan tradisional yang bergantung pada pencocokan kata kunci eksplisit dan memerlukan waktu pemrosesan yang lebih lama. Aspek relevansi jawaban menunjukkan perbedaan mencolok, dimana sistem yang dikembangkan mampu menghasilkan respons kontekstual dengan skor relevansi 4,6 dari 5, sementara pendekatan konvensional terbatas pada pencarian berbasis kata kunci tanpa pemahaman konteks. Kemampuan pemahaman konteks menjadi keunggulan utama sistem ini, mampu memahami nuansa pertanyaan dan memberikan jawaban komprehensif, berbeda dengan sistem tradisional yang tidak mampu memahami konteks pertanyaan dengan baik. Skalabilitas sistem menunjukkan kemampuan menangani hingga 10 pengguna secara bersamaan dengan performa yang stabil, memberikan fleksibilitas yang lebih baik dibandingkan sistem konvensional yang terbatas dalam menangani jumlah pengguna besar [8].

Tantangan Implementasi dan Solusi Optimasi

Proses implementasi sistem menghadapi beberapa tantangan teknis yang memerlukan solusi optimasi khusus. Kendala utama yang diidentifikasi meliputi peningkatan waktu pemrosesan untuk pertanyaan kompleks yang berulang, yang diatasi melalui optimasi pengindeksan pada basis data vektor MongoDB dan implementasi mekanisme penyimpanan sementara untuk pertanyaan umum. Masalah akurasi jawaban yang kadang kurang relevan diselesaikan melalui penyesuaian halus pada model LLaMA3 dan peningkatan kualitas data pelatihan. Solusi teknis ini berhasil meningkatkan stabilitas dan efisiensi sistem dalam menangani berbagai jenis permintaan pengguna. Implementasi optimasi ini menunjukkan pentingnya pendekatan sistematis dalam mengatasi tantangan teknis yang muncul selama pengembangan sistem manajemen pengetahuan berbasis teknologi canggih [9].

Validasi Pengguna dan Penerimaan Sistem

Evaluasi pengguna yang melibatkan 10 karyawan PT Prima Integrasi Network menggunakan kuesioner skala Likert 1-5 menunjukkan tingkat penerimaan yang tinggi terhadap sistem yang dikembangkan. Hasil evaluasi menunjukkan skor rata-rata 4,6 untuk kemudahan penggunaan, 4,8 untuk kecepatan pencarian, 4,6 untuk akurasi jawaban, 4,4 untuk tampilan dan navigasi, serta 4,7 untuk kepuasan keseluruhan. Tanggapan kualitatif dari pengguna menunjukkan apresiasi positif terhadap kemudahan akses informasi melalui antarmuka chatbot yang mengeliminasi kebutuhan pencarian manual dokumentasi. Beberapa pengguna menyatakan bahwa sistem ini membantu mengatasi hambatan komunikasi bagi karyawan yang merasa segan bertanya langsung kepada personel yang bertanggung jawab. Meskipun demikian, evaluasi juga mengidentifikasi area perbaikan, khususnya terkait kompleksitas jawaban yang masih memerlukan peningkatan data pelatihan dan akurasi respons yang perlu ditingkatkan lebih lanjut. Secara keseluruhan, sistem manajemen pengetahuan berbasis LLM dan RAG ini berhasil memenuhi ekspektasi pengguna dalam meningkatkan efisiensi akses dan pemanfaatan informasi organisasi [4].

Analisis Kinerja Sistem Knowledge Management Berbasis LLM dan RAG

Implementasi sistem manajemen pengetahuan yang mengintegrasikan model bahasa besar LLaMA3 dengan teknologi Retrieval Augmented Generation (RAG) telah menunjukkan performa yang signifikan dalam mendukung aktivitas operasional PT Prima Integrasi Network. Berdasarkan hasil pengujian komprehensif yang dilakukan, sistem mampu memberikan respons yang relevan dengan tingkat akurasi rata-rata 4,6 dari skala 5, yang mengindikasikan keberhasilan integrasi antara komponen pencarian vektor dan model generatif dalam menghasilkan jawaban kontekstual [10].

Evaluasi mendalam terhadap kecepatan pemrosesan query menunjukkan bahwa sistem mampu menangani pencarian sederhana dalam waktu 23 detik, sementara untuk query kompleks membutuhkan durasi hingga 33 detik. Variasi waktu eksekusi ini menunjukkan adaptabilitas sistem dalam menangani berbagai tingkat kompleksitas pertanyaan pengguna, dengan rata-rata keseluruhan di bawah 35 detik yang masih berada dalam batas toleransi responsivitas yang dapat diterima. Implementasi basis data vektor MongoDB sebagai infrastruktur penyimpanan telah memungkinkan pencarian semantik yang efisien, memfasilitasi proses pengambilan informasi yang relevan berdasarkan kesamaan vektor representasi dokumen [11].

Aspek keunggulan sistem terletak pada kemampuan pemahaman kontekstual yang superior dibandingkan pendekatan tradisional berbasis kata kunci. Sistem mampu menginterpretasi maksud pengguna melalui pemrosesan bahasa alami yang canggih, menghasilkan jawaban yang tidak hanya akurat secara faktual tetapi juga kontekstual sesuai dengan kebutuhan spesifik pengguna. Hal ini tercermin dari tingkat relevansi jawaban yang konsisten tinggi across berbagai jenis pertanyaan, mulai dari pertanyaan sederhana mengenai topik bantuan hingga pertanyaan teknis kompleks terkait konfigurasi sistem [12].

Skalabilitas sistem telah diuji melalui simulasi beban kerja dengan kemampuan menangani hingga 10 pengguna secara bersamaan tanpa mengalami degradasi performa yang signifikan. Hasil pengujian menunjukkan bahwa arsitektur sistem yang dikembangkan memiliki fondasi yang solid untuk ekspansi lebih lanjut, meskipun optimisasi tambahan akan diperlukan untuk menangani volume pengguna yang lebih besar dalam lingkungan produksi skala enterprise [13].

Evaluasi Kepuasan Pengguna dan Dampak Implementasi

Evaluasi komprehensif terhadap pengalaman pengguna yang melibatkan 10 karyawan PT Prima Integrasi Network sebagai responden telah menghasilkan wawasan mendalam mengenai efektivitas sistem knowledge management yang dikembangkan. Metodologi evaluasi yang menggabungkan kuesioner terstruktur dengan skala Likert dan wawancara mendalam telah memberikan perspektif holistik terhadap dampak implementasi sistem [10].

Hasil evaluasi menunjukkan tingkat kepuasan pengguna yang tinggi dengan skor rata-rata 4,7 dari skala 5 untuk aspek kepuasan keseluruhan. Dimensi kecepatan pencarian memperoleh penilaian tertinggi dengan skor 4,8, yang mengindikasikan bahwa responsivitas sistem telah memenuhi ekspektasi pengguna dalam lingkungan kerja yang dinamis. Kemudahan penggunaan dan akurasi jawaban masing-masing memperoleh skor 4,6, menunjukkan keseimbangan yang baik antara usability dan reliability sistem [14].

Temuan kualitatif dari wawancara pengguna mengungkapkan dampak positif yang signifikan terhadap produktivitas kerja. Manajer divisi data enterprise menyatakan bahwa sistem telah "mempermudah mencari kebutuhan kantor hanya lewat chatbot, jadi tidak perlu mencari-cari dokumentasi yang ada", yang menunjukkan efisiensi operasional yang dihasilkan. Perspektif manajer divisi sales menegaskan bahwa sistem "akan lebih banyak membantu karyawan yang terkendala dan malu jika bertanya kepada PIC-nya", mengindikasikan nilai tambah sistem dalam mengatasi hambatan komunikasi organisasional [15].

Namun, evaluasi juga mengidentifikasi area yang memerlukan perbaikan. Leader database administrator menyarankan bahwa "hasil jawabannya masih kurang kompleks dan harus di training dengan data yang lebih banyak", yang menunjukkan kebutuhan akan peningkatan kualitas dataset pelatihan. Aspek tampilan dan navigasi memperoleh skor terendah dengan 4,4, mengindikasikan potensi perbaikan pada antarmuka pengguna yang dikembangkan menggunakan Streamlit. Perbandingan sistematis dengan pendekatan tradisional menunjukkan keunggulan signifikan sistem yang dikembangkan. Dalam aspek kecepatan pencarian, sistem baru mampu memberikan hasil dalam waktu rata-rata di bawah 35 detik, kontras dengan

pendekatan tradisional yang lambat karena ketergantungan pada pencocokan kata kunci eksplisit. Relevansi jawaban menunjukkan peningkatan dramatis dari pendekatan berbasis keyword yang terbatas menjadi sistem yang mampu menghasilkan jawaban kontekstual dengan skor relevansi 4,6 dari 5.

Implementasi sistem telah berhasil mengatasi beberapa kendala teknis melalui optimisasi yang tepat. Peningkatan waktu query untuk pertanyaan kompleks yang berulang telah diatasi melalui implementasi indexing yang dioptimalkan pada basis data vektor MongoDB dan penerapan mekanisme caching untuk pertanyaan umum. Masalah akurasi jawaban yang kurang relevan telah diminimalkan melalui fine-tuning model LLaMA3 dan peningkatan kualitas data pelatihan [16].

Secara keseluruhan, evaluasi pengguna menunjukkan bahwa sistem knowledge management berbasis LLM dan RAG telah berhasil memenuhi objektif penelitian dengan memberikan solusi yang efektif untuk pengelolaan dan pencarian informasi organisasional. Tingkat kepuasan pengguna yang tinggi dan peningkatan produktivitas yang terukur mengkonfirmasi validitas pendekatan teknologi yang dipilih dan implementasi yang dilakukan.

Implikasi Strategis dan Rekomendasi Pengembangan Sistem

Implementasi sistem knowledge management berbasis LLM dan RAG pada PT Prima Integrasi Network telah menghadirkan transformasi paradigma dalam pengelolaan informasi organisasional yang memiliki implikasi strategis jangka panjang. Keberhasilan sistem dalam mencapai tingkat akurasi 4,6 dari skala 5 dan kepuasan pengguna 4,7 dari skala 5 menunjukkan potensi signifikan untuk ekspansi aplikasi teknologi ini pada berbagai aspek operasional perusahaan. Analisis mendalam terhadap hasil implementasi mengindikasikan bahwa sistem tidak hanya berfungsi sebagai alat bantu pencarian informasi, tetapi telah berkembang menjadi platform strategis yang dapat meningkatkan efisiensi organisasional secara holistik.

Dari perspektif efisiensi operasional, sistem telah berhasil mengurangi waktu pencarian informasi dari metode tradisional yang memerlukan penelusuran manual dokumentasi menjadi proses otomatis dengan rata-rata waktu respons di bawah 35 detik. Transformasi ini memiliki implikasi ekonomis yang signifikan, mengingat pengurangan waktu pencarian dapat dialokasikan untuk aktivitas produktif lainnya. Perhitungan konservatif menunjukkan bahwa jika setiap karyawan menghemat rata-rata 30 menit per hari dari aktivitas pencarian informasi, maka perusahaan dapat memperoleh peningkatan produktivitas yang substansial dalam jangka waktu operasional tahunan [17].

Aspek skalabilitas sistem menunjukkan kemampuan menangani hingga 10 pengguna simultan dengan performa stabil, namun hasil pengujian ini mengindikasikan kebutuhan optimisasi lebih lanjut untuk mengakomodasi pertumbuhan organisasi. Rekomendasi strategis untuk pengembangan sistem mencakup implementasi arsitektur terdistribusi yang dapat menangani beban kerja yang lebih besar, serta integrasi dengan infrastruktur cloud computing untuk meningkatkan fleksibilitas dan skalabilitas sistem. Optimisasi basis data vektor MongoDB melalui implementasi sharding dan replikasi dapat meningkatkan kapasitas pemrosesan query secara signifikan [18].

Analisis terhadap kendala implementasi yang teridentifikasi memberikan wawasan penting untuk pengembangan sistem berkelanjutan. Peningkatan waktu query untuk pertanyaan kompleks yang berulang telah diatasi melalui optimisasi indexing dan implementasi caching, namun solusi jangka panjang memerlukan pengembangan algoritma prediktif yang dapat mengantisipasi pola pertanyaan pengguna. Implementasi machine learning untuk analisis pola penggunaan dapat memungkinkan sistem untuk melakukan pre-loading informasi yang sering diakses, sehingga mengurangi latensi respons secara keseluruhan. Kualitas jawaban yang kadang kurang akurat atau tidak relevan, sebagaimana diidentifikasi oleh leader database administrator, menunjukkan kebutuhan akan strategi peningkatan kualitas data pelatihan yang

sistematis. Rekomendasi pengembangan mencakup implementasi feedback loop yang memungkinkan sistem untuk belajar dari interaksi pengguna dan meningkatkan akurasi jawaban secara iteratif. Integrasi mekanisme active learning dapat memungkinkan sistem untuk mengidentifikasi area pengetahuan yang memerlukan penguatan dan secara otomatis meminta validasi dari ahli domain [19].

Dari perspektif antarmuka pengguna, skor terendah pada aspek tampilan dan navigasi (4,4) mengindikasikan potensi perbaikan yang signifikan dalam pengalaman pengguna. Rekomendasi pengembangan mencakup implementasi desain responsive yang dapat beradaptasi dengan berbagai perangkat, serta integrasi elemen interaktif yang dapat meningkatkan engagement pengguna. Pengembangan dashboard analitik yang memberikan wawasan tentang pola penggunaan sistem dapat membantu administrator dalam mengoptimalkan konfigurasi dan konten sistem.

Implikasi strategis jangka panjang dari implementasi sistem mencakup potensi ekspansi ke departemen lain dalam organisasi, serta integrasi dengan sistem enterprise resource planning existing. Pengembangan API yang robust dapat memungkinkan integrasi seamless dengan aplikasi bisnis lainnya, menciptakan ekosistem informasi yang terintegrasi dan komprehensif. Rekomendasi untuk fase pengembangan selanjutnya mencakup implementasi multi-modal capabilities yang dapat menangani berbagai jenis media informasi, termasuk dokumen visual, audio, dan video [20].

Kesimpulan strategis menunjukkan bahwa sistem knowledge management berbasis LLM dan RAG telah berhasil membuktikan viabilitas teknologi dalam konteks organisasional, dengan potensi pengembangan yang signifikan untuk meningkatkan value proposition sistem. Implementasi rekomendasi pengembangan yang sistematis dapat memposisikan PT Prima Integrasi Network sebagai pionir dalam adopsi teknologi artificial intelligence untuk pengelolaan pengetahuan organisasional.

KESIMPULAN

Implementasi sistem manajemen pengetahuan berbasis model bahasa besar dan mekanisme pengambilan-augmentasi generasi menggunakan LLaMA3 dengan basis data vektor MongoDB pada PT Prima Integrasi Network telah mencapai keberhasilan yang signifikan dalam transformasi digital pengelolaan informasi perusahaan. Penelitian ini mendemonstrasikan bahwa integrasi teknologi kecerdasan buatan terdepan dapat memberikan solusi komprehensif untuk tantangan manajemen pengetahuan di lingkungan korporat modern. Arsitektur sistem yang dikembangkan membuktikan efektivitas sinergi antara model LLaMA3, kerangka kerja LlamaIndex, dan teknologi basis data vektor MongoDB dalam menciptakan ekosistem pencarian informasi yang cerdas dan responsif. Mekanisme retrieval-augmented generation yang diimplementasikan berhasil meningkatkan relevansi dan akurasi respons sistem melalui pengambilan konteks yang tepat dari basis pengetahuan sebelum proses generasi jawaban. Pendekatan ini memungkinkan sistem untuk memberikan informasi yang tidak hanya akurat secara faktual, tetapi juga kontekstual sesuai dengan kebutuhan spesifik pengguna. Evaluasi performa sistem menunjukkan pencapaian yang menggembirakan dalam berbagai aspek kinerja operasional. Kecepatan pemrosesan kueri mengalami peningkatan substansial dibandingkan dengan metode pencarian konvensional, sementara tingkat akurasi jawaban mencapai standar yang memuaskan untuk aplikasi enterprise. Penggunaan teknologi pengindeksan vektor pada MongoDB terbukti memberikan keunggulan kompetitif dalam hal efisiensi pencarian, memungkinkan sistem untuk menangani volume data yang besar dengan tetap mempertahankan responsivitas yang optimal. Aspek penerimaan pengguna terhadap sistem yang dikembangkan menunjukkan tingkat kepuasan yang tinggi, dengan skor evaluasi rata-rata mencapai 4,7 dari skala 5 pada parameter kemudahan penggunaan, kecepatan pencarian, dan kepuasan keseluruhan. Hasil ini mengindikasikan bahwa sistem tidak hanya

unggul secara teknis, tetapi juga berhasil memenuhi ekspektasi dan kebutuhan praktis pengguna akhir. Meskipun demikian, identifikasi area perbaikan seperti peningkatan antarmuka pengguna, penyempurnaan mekanisme penyaringan pencarian, dan optimasi model melalui fine-tuning menunjukkan komitmen terhadap perbaikan berkelanjutan.

REFERENSI

- Ahn, Y., S. G. Lee, J. Shim, and J. Park, "Retrieval-Augmented Response Generation for Knowledge-Grounded Conversation in the Wild," *IEEE Access*, vol. 10, pp. 131374–131385, Nov. 2022, doi: 10.1109/ACCESS.2022.3228964.
- Chen, E., "Empowering artificial intelligence for knowledge management augmentation," *Issues Inf. Syst.*, vol. 25, no. 4, pp. 409–416, 2024, doi: 10.48009/4_iis_2024_132.
- Chawla, D., and H. Joshi, "Relationship between Knowledge Management Practices and Performance," *Int. J. Indian Cult. Bus. Manag.*, vol. 1, p. 1, Jan. 2020, doi: 10.1504/IJICBM.2020.10032110.
- Cheng, Y. *et al.*, "Exploring Large Language Model based Intelligent Agents: Definitions, Methods, and Prospects," 2024. [Online]. Available: <http://arxiv.org/abs/2401.03428>
- Glass, M., G. Rossiello, M. F. M. Chowdhury, and A. Gliozzo, "Robust Retrieval Augmented Generation for Zero-shot Slot Filling," *Proc. 2021 Conf. Empir. Methods Nat. Lang. Process. (EMNLP 2021)*, pp. 1939–1949, 2021, doi: 10.18653/v1/2021.emnlp-main.148.
- Hofstätter, S., J. Chen, K. Raman, and H. Zamani, "FiD-Light: Efficient and Effective Retrieval-Augmented Text Generation," *Proc. 46th Int. ACM SIGIR Conf. Res. Dev. Inf. Retr. (SIGIR 2023)*, pp. 1437–1447, 2023, doi: 10.1145/3539618.3591687.
- Inkpen, A., D. Minbaeva, and E. W. K. Tsang, "Unintentional, unavoidable, and beneficial knowledge leakage from the multinational enterprise," *J. Int. Bus. Stud.*, vol. 50, no. 2, pp. 250–260, 2019, doi: 10.1057/s41267-018-0164-6.
- Jakubczyc, J. A., and M. L. Owoc, "Knowledge Management and Artificial," *Argumenta Oeconomica*, vol. 1–2, no. 1, p. 6, 1998.
- Jiang, Z. *et al.*, "Active Retrieval Augmented Generation," pp. 7969–7992, 2023. [Online]. Available: <https://github.com/>
- Makin, A., "Ontology-Driven Knowledge Management Systems Enhanced by Large Language Models AI and Ethics MSc," pp. 0–24, Oct. 2024, doi: 10.13140/RG.2.2.24648.23043.
- Majumder, S., and N. Dey, "Artificial Intelligence and Knowledge Management," *Stud. Big Data*, vol. 107, pp. 85–100, 2022, doi: 10.1007/978-981-19-0316-8_5.
- Meilinda, and R. V. Imbar, "Perancangan Model Referral dengan Pendekatan Design Science Research Methodology," *J. Tek. Inform. dan Sist. Inf.*, vol. 10, pp. 430–438, 2024.
- Mundzir, A., D. G. Purnama, and I. Mayasari, "AI Implementation in Knowledge Management for Online News Media Case Study: GOnews.id," *Brill. Res. Artif. Intell.*, vol. 4, no. 2, pp. 771–778, 2024, doi: 10.47709/brilliance.v4i2.4995.
- Nakash, M., and E. Bolisani, "Knowledge Management Meets Artificial Intelligence: A Systematic Review and Future Research Agenda," *Proc. Eur. Conf. Knowl. Manag. (ECKM 2024)*, pp. 544–552, Sept. 2024, doi: 10.34190/eckm.25.1.2443.
- Noori, H., H. Al-Hashimy, Y. Jinfang, W. N. Hussein, and H. N. Hussain, "Knowledge Management in Accounting Information Systems with the Assistance of Artificial Intelligence: Bridging the Information Gap for Enhanced Decision-Making," *Multidimens. Res. J.*, pp. 46–56, 2023. [Online]. Available: <http://innosci.org>
- Pai, R. Y. *et al.*, "Integrating artificial intelligence for knowledge management systems – synergy among people and technology: a systematic review of the evidence," *Econ. Res. Istraživanja*, vol. 35, no. 1, pp. 7043–7065, Dec. 2022, doi:

10.1080/1331677X.2022.2058976.

Taipalus, T., “Vector database management systems: Fundamental concepts, use-cases, and current challenges,” *Cogn. Syst. Res.*, vol. 85, p. 101216, Jan. 2024, doi: 10.1016/j.cogsys.2024.101216.

Thakuri, S., M. Bon, N. Cavus, and N. Sancar, “Artificial Intelligence on Knowledge Management Systems for Businesses: A Systematic Literature Review,” *TEM J.*, vol. 13, no. 3, pp. 2146–2155, 2024, doi: 10.18421/TEM133-42.

Wawrzik, F., K. Adil, T. Urimubenshi, and C. Grimm, “KGG4SE: A Knowledge Graph Generation Framework for Systems Engineering,” vol. 0, no. 0.

Xian, J., T. Teofili, R. Pradeep, and J. Lin, “Vector Search with OpenAI Embeddings: Lucene Is All You Need,” *Proc. 17th ACM Int. Conf. Web Search Data Min.* (WSDM 2024), pp. 1090–1093, 2024, doi: 10.11