

## Iterative Query Reformulation Berbasis LLM Untuk Temu Balik Artikel Hukum Indonesia

Muhammad Abdiel Al hafiz<sup>1</sup>, Bayu Kusuma Wardana<sup>2</sup>, Surya Harwindu Putra Wibawa<sup>3</sup>, Muhammad Zhiya Ulhaq<sup>4</sup>, Muhami Jauza Alma Ramadhan<sup>5</sup>, Noval Esa Ramdani<sup>6</sup>, Evan Prima Mangku Leksana<sup>7</sup>

<sup>1,2,3,4,5,6</sup>Prodi Informatika, Fakultas Ilmu Komputer, Universitas Amikom Purwokerto

<sup>7</sup>Prodi Teknik Informatika, Fakultas Teknik dan Sains, Universitas Muhammadiyah Purwokerto

surel: [hafd324@gmail.com](mailto:hafd324@gmail.com), [bwardana406@gmail.com](mailto:bwardana406@gmail.com), [suryauyawangon@gmail.com](mailto:suryauyawangon@gmail.com), [zhiyaulhaq25@gmail.com](mailto:zhiyaulhaq25@gmail.com),

[muhamijauza12@gmail.com](mailto:muhamijauza12@gmail.com), [novalesa453@gmail.com](mailto:novalesa453@gmail.com), [primaevan8@gmail.com](mailto:primaevan8@gmail.com)

### Info Artikel

#### Sejarah artikel:

Diterima 30-01-2026

Revisi 11-02-2026

Diterima 23-02-2026

#### Kata kunci:

RAG

Iterative Query Reformulation

Hukum Indonesia

LLM

Information Retrieval

### ABSTRAK

Akses terhadap informasi hukum yang akurat sering kali terkendala oleh kesenjangan semantik antara bahasa awam yang digunakan masyarakat dan terminologi hukum formal. Penelitian terdahulu yang menerapkan Retrieval Augmented Generation (RAG) pada Kitab Undang-Undang Hukum Pidana (KUHP) menghasilkan tingkat keberhasilan (Hit Rate) sebesar 80%, namun masih memiliki keterbatasan dalam menangani variasi gaya bahasa pengguna yang implisit serta konteks hukum yang kompleks. Penelitian ini mengusulkan metode Iterative Query Reformulation berbasis Large Language Model (LLM) untuk mengatasi permasalahan tersebut. Sistem ini menggunakan mekanisme Double Hop Retrieval yang melibatkan pencarian awal (initial retrieval) untuk mendapatkan konteks, diikuti oleh reformulasi pertanyaan menggunakan model Qwen 3 32B menjadi kueri hukum baku, sebelum dilakukan pencarian final (final retrieval). Menggunakan dataset yang dikumpulkan dari HukumOnline sebanyak 750 artikel, evaluasi dilakukan menggunakan metrik Faithfulness, Answer Relevancy, Hit Rate, dan Mean Reciprocal Rank (MRR). Hasil pengujian menunjukkan peningkatan performa yang signifikan, dengan skor Faithfulness mencapai 98.6%, Answer Relevancy 100%, Hit Rate 100%, dan peningkatan MRR dari 90% (tanpa reformulasi) menjadi 93% (dengan reformulasi). Pendekatan ini terbukti efektif menjembatani kesenjangan bahasa dan meningkatkan presisi temu balik informasi hukum di Indonesia.

### Penulis yang sesuai:

Muhammad Abdiel Al hafiz

Program Studi Informatika Fakultas Ilmu Komputer Universitas Amikom Purwokerto

Email: [hafd324@gmail.com](mailto:hafd324@gmail.com)

## 1. PENDAHULUAN

Akses terhadap informasi hukum yang akurat dan relevan merupakan prasyarat penting bagi terwujudnya keadilan dan kepastian hukum dalam masyarakat. Dalam sistem hukum Indonesia yang menganut tradisi *civil law*, dokumen hukum tertulis seperti undang-undang, peraturan pemerintah, dan putusan pengadilan menjadi rujukan



utama dalam penyelesaian persoalan hukum. Seiring meningkatnya digitalisasi dokumen hukum, tantangan utama yang dihadapi masyarakat bukan lagi pada ketersediaan informasi, melainkan pada kemampuan menemukan dan memahami dokumen hukum yang sesuai dengan kebutuhan secara tepat dan memahami dokumen hukum yang sesuai dengan kebutuhan secara tepat dan efisien. Sejumlah penelitian menunjukkan bahwa kompleksitas struktur dokumen hukum serta pengguna non-ahli dalam mengakses informasi hukum secara efektif [1] [2] [3] [4]. permasalahan tersebut berdampak langsung pada proses temu balik informasi hukum (*legal information retrieval*). Dokumen hukum memiliki karakteristik yang berbeda dibandingkan teks umum, baik dari segi panjang, struktur hirarkis, maupun definisi relevansi yang sangat berantung pada konteks hukum. Ketika pengguna mengajukan pertanyaan dengan bahasa alami atau istilah awam, sistem temu balik sering kali gagal mengidentifikasi dokumen yang relevan akibat ketidaksesuaian kosakata dan makna (*vocabulary mismatch*) antara kueri pengguna dan representasi dokumen hukum [5] [6] [7]. Akibatnya, meskipun informasi hukum tersedia dalam basis data, hasil pencarian yang diperoleh sering kali kurang presisi atau tidak menjawab kebutuhan pengguna secara memadai.

Berbagai penelitian telah berupaya meningkatkan aksesibilitas informasi hukum melalui pemanfaatan kecerdasan buatan, khususnya *Natural Language Processing* dan *Large Language Model (LLM)*, dalam bentuk chatbot hukum dan sistem tanya jawab berbasis dokumen. Pendekatan *Retrieval-Augmented Generation (RAG)* diperkenalkan untuk mengatasi keterbatasan pengetahuan parametrik LLM dan mengurangi risiko halusinasi dengan mengkaitkan proses generasi jawaban pada dokumen eksternal yang relevan [8] [9] [10]. Pada penelitian sebelumnya oleh [11]. Yang mengembangkan chatbot hukum pidana menggunakan model LLaMA dengan dataset KUHP. Penelitian tersebut berhasil mencapai Hit Rate sebesar 80% dalam relevansi dokumen. Meskipun hasil tersebut menjanjikan, penelitian tersebut memiliki keterbatasan, yaitu cakupan dataset yang hanya bersumber dari satu dokumen hukum (KUHP) dan metode evaluasi yang belum mengakomodasi variasi gaya bahasa pengguna. Dalam praktiknya, pengguna sering kali mengajukan pertanyaan dengan istilah awam (contoh: "nyolong wifi") yang tidak secara langsung cocok dengan istilah hukum formal ("akses komputer tanpa hak"), sehingga menyebabkan kegagalan dalam proses temu balik (*retrieval*) dokumen yang relevan. Fenomena ini sejalan dengan temuan penelitian di bidang *query reformulation* yang menyatakan bahwa kualitas hasil temu balik sangat dipengaruhi oleh bagaimana kueri pengguna dirumuskan sebelum proses pencarian dilakukan [12] [13].

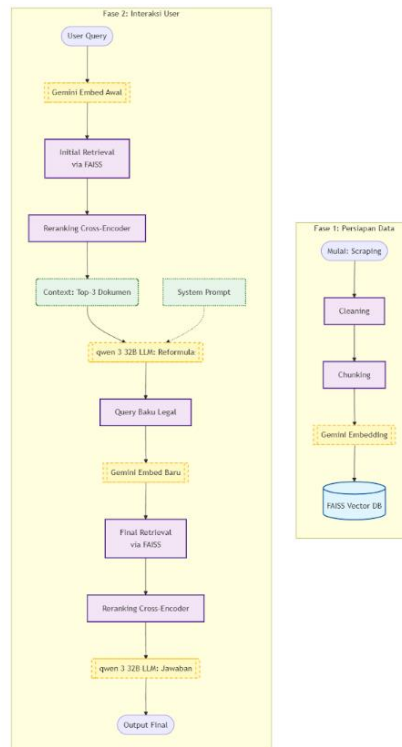
Untuk mengatasi keterbatasan tersebut, penelitian ini mengusulkan arsitektur *Iterative Query Reformulation* berbasis LLM yang terinspirasi dari metode *Corrective RAG*. Metode ini tidak langsung menggunakan kueri mentah pengguna untuk melakukan pencarian dokumen. Sebaliknya, sistem terlebih dahulu melakukan reformulasi kueri secara iteratif dengan memanfaatkan kemampuan pemahaman bahasa LLM untuk mengubah pertanyaan pengguna ke dalam bentuk kueri hukum yang lebih baku, eksplisit, dan presisi sebelum dilakukan proses temu balik akhir. Metode iteratif semacam ini terbukti mampu meningkatkan kualitas *retrieval* dan mengurangi kegagalan pencarian pada kebutuhan informasi yang kompleks [9] [10] [14].

Selain itu, penelitian ini memperluas cakupan dataset dengan memanfaatkan artikel hukum dari HukumOnline sehingga tidak terbatas pada domain hukum pidana saja. Perluasan dataset ini bertujuan untuk meningkatkan kemampuan generalisasi sistem dalam menangani berbagai domain hukum dan variasi kebutuhan informasi pengguna. Dengan mengkombinasikan reformulasi kueri berbasis LLM secara iteratif dan cakupan data hukum yang lebih luas, penelitian ini menawarkan kontribusi baru dalam pengembangan sistem temu balik artikel hukum Indonesia yang lebih adaptif, presisi, dan inklusif, sekaligus mendukung peningkatan kualitas akses masyarakat terhadap informasi hukum yang dapat dipertanggungjawabkan [15].

## 2. METODE

Penelitian ini menggunakan pendekatan kuantitatif eksperimental untuk mengukur efektivitas metode reformulasi kueri dalam arsitektur *Retrieval Augmented Generation (RAG)*. Pendekatan ini dilakukan dengan membandingkan kinerja sistem RAG standar yang menggunakan kueri mentah pengguna dengan sistem RAG yang dilengkapi mekanisme *Iterative Query Reformulation* berbasis *Large Language Model*. Evaluasi dilakukan secara terukur menggunakan metrik kuantitatif pada hasil temu balik dokumen guna menilai peningkatan relevansi dan presisi pencarian yang dihasilkan oleh metode yang diusulkan.





Gambar 2. Flowchart Persiapan Data dan Interaksi User

## 2.1. Pengumpulan Dataset

Data yang digunakan dalam penelitian ini diperoleh melalui proses *web scraping* dari situs HukumOnline. Dataset mencakup 15 tema hukum, termasuk pidana, perdata, teknologi keluarga, dan bisnis. Total data yang dikumpulkan berjumlah 750 artikel, dengan masing-masing tema terdiri dari 50 artikel. Atribut data meliputi tema, judul artikel, tautan sumber, tanggal publikasi, tag, dan konten artikel yang berisi analisis jawaban pakar hukum.

## 2.2. Pra-pemrosesan Data (Preprocessing)

Tahapan pengolahan data dilakukan secara terintegrasi untuk memastikan kualitas input bagi model:

### 2.2.1. Cleaning

Menggunakan alat bantu Firecrawl untuk menghapus elemen *noise* (seperti iklan dan navigasi situs) dan mengonversi format halaman web menjadi Markdown yang bersih.

### 2.2.2. Chunking

Data diproses menggunakan metode *Recursive Character Chunking* dengan konfigurasi ukuran *chunk* 1000 karakter dan *overlap* 200 karakter. Metode ini dipilih untuk menjaga keutuhan makna pasal atau argumen hukum agar tidak terputus, serta memastikan kesinambungan konteks antar potongan teks.

### 2.2.3. Embedding & Penyimpanan

Teks yang telah dipotong diubah menjadi representasi vektor menggunakan Gemini Embedding dan disimpan dalam basis data vektor FAISS (*Facebook AI Similarity Search*) untuk memungkinkan pencarian semantik yang cepat dan efisien.

## 2.3. Arsitektur Sistem: Retrieval-Enhanced Query Reformulation

Penelitian ini menerapkan mekanisme Double-Hop Retrieval yang terdiri dari tahapan berikut:

### 2.3.1. Initial Retrieval (Pencarian Awal)

Sistem menerima input pengguna dalam bahasa alami (awam) dan melakukan pencarian awal ke FAISS. Tujuan tahap ini bukan untuk menemukan jawaban akhir, melainkan untuk mendapatkan dokumen konteks awal yang mengandung kata kunci relevan secara luas.

### 2.3.2. Query Reformulation (Reformulasi Kueri)

Model LLM Qwen 3 32B menerima input berupa pertanyaan pengguna beserta 3 artikel hasil pencarian awal. Model menganalisis konteks hukum dari artikel tersebut dan mereformulasi pertanyaan pengguna menjadi

Standardized Legal Query. Contohnya, pertanyaan "tetangga nyolong wifi" direformulasi menjadi "Pasal 30 UU ITE tentang Akses Ilegal dan Pasal 362 KUHP tentang Pencurian".

### 2.3.3. Final Retrieval & Generation

Sistem melakukan pencarian ulang ke basis data vektor menggunakan kueri yang telah dibakukan. Hasil pencarian (Top 8 dokumen) diurutkan ulang (reranking) dan diserahkan ke model Llama untuk menghasilkan jawaban nal yang komprehensif dan mudah dipahami.

## 2.4. Evaluasi Model

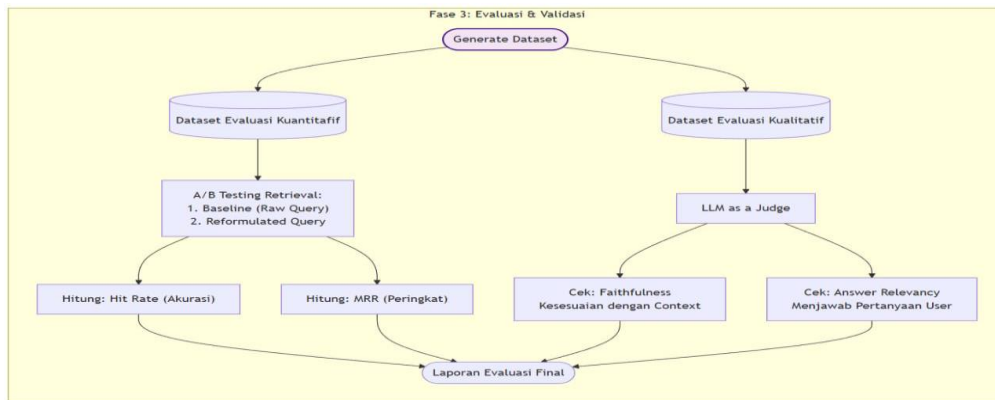
Evaluasi dilakukan melalui dua pendekatan utama:

### 2.4.1. Evaluasi Kuantitatif (Retrieval Metrics)

Dilakukan A/B Testing untuk membandingkan performa pencarian antara *Baseline* (menggunakan raw query) dan metode usulan (reformulated query). Metrik yang digunakan adalah *Hit Rate* (akurasi penemuan dokumen relevan) dan *Mean Reciprocal Rank* (MRR) untuk mengukur kualitas peringkat dokumen.

### 2.4.2. Evaluasi Kualitatif (Generation Quality)

Menggunakan metode *LLM as a Judge* untuk menilai kualitas jawaban teks. Aspek yang dinilai meliputi Faithfulness (konsistensi jawaban terhadap konteks dokumen untuk mendeteksi halusinasi) dan Answer Relevancy (relevansi jawaban terhadap inti pertanyaan pengguna).



Gambar 2. Flowchart Evaluasi Model

## 3. HASIL DAN PEMBAHASAN

### 3.1. Hasil Implementasi Sistem

Sistem yang dibangun berhasil mengintegrasikan proses scraping, penyimpanan vektor, dan orkestrasi LLM untuk reformulasi kueri. Penggunaan Qwen 3 32B sebagai model reformulasi terbukti efisien dengan context window yang memadai untuk memproses konteks artikel awal tanpa terpotong, serta didukung infrastruktur API Groq yang stabil.

### 3.2. Evaluasi Performa

Pengujian dilakukan untuk mengukur dampak dari teknik reformulasi kueri terhadap akurasi dan relevansi sistem. Ringkasan hasil evaluasi disajikan pada Tabel 1.

Tabel 1. Hasil Evaluasi Metrik Performa Sistem

No	Matrik Evaluasi	Skor	Predikat
1	Faithfulness	98.6%	Sangat Andal
2	Answer Relevancy	100%	Sangat Relevan
3	Hit Rate	100%	Sangat Efektif
4	MRR (Baseline - Tanpa Reformulasi)	90%	Presisi Tinggi
5	MRR (Proposed - Dengan Reformulasi)	93%	Presisi Tinggi

### 3.3. Pembahasan

Berdasarkan Tabel 3.1, sistem menunjukkan performa yang sangat baik dalam tiga aspek utama: pemahaman konteks, kepatuhan regulasi, dan keselarasan dengan intensi pengguna (User Intent Alignment).



Peningkatan skor MRR dari 90% menjadi 93% membuktikan bahwa metode Iterative Query Reformulation efektif dalam memperbaiki peringkat dokumen yang relevan. Dengan mereformulasi bahasa awam menjadi istilah hukum baku, sistem mampu menempatkan dokumen hukum yang paling tepat di urutan teratas hasil pencarian. Hal ini sangat krusial dalam sistem temu balik informasi hukum di mana presisi adalah prioritas.

Skor Faithfulness sebesar 98.6% mengindikasikan bahwa sistem memiliki tingkat halusinasi yang sangat rendah. Jawaban yang dihasilkan sangat konsisten dengan dokumen sumber yang ditemukan. Sementara itu, skor Answer Relevancy yang mencapai 100% menunjukkan keberhasilan inovasi reformulasi kueri dalam menjembatani kesenjangan bahasa (gap between layman language and legal terminology). Sistem mampu memahami pertanyaan implisit pengguna dan memberikan jawaban yang secara hukum akurat namun tetap mudah dipahami.

Meskipun demikian, penelitian ini memiliki keterbatasan pada cakupan data yang terbatas pada 15 kategori hukum (~750 artikel). Selain itu, ketergantungan pada API pihak ketiga (Groq/Google) menimbulkan tantangan terkait privasi data, sehingga implementasi untuk kasus yang sangat rahasia memerlukan pertimbangan infrastruktur On-Premise.

#### 4. KESIMPULAN

Penelitian ini berhasil mengembangkan dan mengevaluasi sistem temu balik informasi hukum yang memanfaatkan teknik *Iterative Query Reformulation* berbasis LLM. Hasil eksperimen menunjukkan bahwa pendekatan ini secara signifikan meningkatkan relevansi dan akurasi sistem dibandingkan dengan metode pencarian konvensional. Dengan skor MRR 93% dan *Answer Relevancy* 100%, sistem terbukti mampu menerjemahkan pertanyaan berbahasa awam menjadi kueri hukum yang presisi. Hal ini memberikan kontribusi paling signifikan dalam upaya demokratisasi akses informasi hukum di Indonesia, menjadikan hukum lebih mudah diakses dan dipahami oleh masyarakat luas. Pengembangan selanjutnya disarankan untuk memperluas korpus data ke skala nasional dan mengeksplorasi penggunaan model lokal untuk meningkatkan keamanan privasi data.

#### REFERENSI

- [1] L. Septian, T. Aljauza, and C. Juliane, "Indonesian Journal of Computer Science," vol. 12, no. 1, pp. 4428–4439, 2024.
- [2] I. D. Saputra, N. S. Harahap, S. Agustian, M. Fikry, and L. Oktavia, "Aplikasi Web Question Answering Menggunakan Langchain OpenAI Tentang Peraturan Perundang-Undangan Bidang Pendidikan," vol. 6, no. 1, pp. 293–304, 2024, doi: 10.47065/josyc.v6i1.6182.
- [3] D. Raevan, F. Darari, and R. Adha, "Granularity-aware legal question answering : a case study of indonesian government regulations," vol. 10, no. 3, pp. 359–378, 2024.
- [4] F. Lagioia, M. Lippi, G. Sartor, A. Passerini, and B. Sayin, "Towards Reliable Retrieval in RAG Systems for Large Legal Datasets".
- [5] Y. Zhou, H. Huang, and Z. Wu, "Boosting legal case retrieval by query content selection with large language models", doi: 10.1145/3624918.3625328.
- [6] E. Yulianti, N. Bhary, J. Abdurrohmam, and F. W. Dwitilas, "Named entity recognition on Indonesian legal documents : a dataset and study using transformer-based models," vol. 14, no. 5, pp. 5489–5501, 2024, doi: 10.11591/ijece.v14i5.pp5489-5501.
- [7] E. Q. Nuranti and E. Yulianti, "Predicting the Category and the Length of Punishment in Indonesian Courts Based on Previous Court Decision Documents," pp. 1–20, 2022.
- [8] F. Cuconasu, G. Trappolini, F. Siciliano, S. Filice, I. R. May, and N. Tonello, "The Power of Noise : Redefining Retrieval for RAG Systems", doi: 10.1145/3626772.3657834.
- [9] D. R. Faisal *et al.*, "A Hybrid Virtual Assistant for Legal Domain Based on Information Retrieval and Knowledge Graphs," vol. 2, pp. 125–140, 2023.
- [10] Z. Shao, Y. Gong, M. Huang, N. Duan, and W. Chen, "Enhancing Retrieval-Augmented Large Language Models with Iterative Retrieval-Generation Synergy".
- [11] A. Z. Pratama, A. Marinta, B. Triyudanto, M. Saman, and T. Noor, "Retrieval-Augmented Generation for Indonesian Criminal Law Information Using the LLaMA Model," vol. 1, no. 1, pp. 35–41, 2025.
- [12] K. D. Dhole and E. Agichtein, "GenQREnsemble : Zero-Shot LLM Ensemble Prompting for".
- [13] E. S. Optimization, D. A. Nguyen, R. K. Mohan, and V. Yang, "RL-based Query Rewriting with Distilled LLM for online E-Commerce Systems," 2022.
- [14] G. Dong, *Leveraging LLM-Assisted Query Understanding for Live Retrieval-Augmented Generation*, vol. 1, no. 1. arXiv, 2025.
- [15] X. Ma, Y. Gong, P. He, H. Zhao, and N. Duan, "Query Rewriting for Retrieval-Augmented Large Language Models," no. 2022, pp. 5303–5315, 2023.

