



**CHATBOT INTERNAL MULTI-DIVISI BERBASIS
RETRIEVAL-AUGMENTED GENERATION
MENGUNAKAN HYBRID RETRIEVER BM25 DAN
APPROXIMATE NEAREST NEIGHBOR**

**TUGAS AKHIR
APLIKATIF**

**IBKA ANHAR FATCHA
41522010137**

**UNIVERSITAS
MERCU BUANA**

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2026**



**CHATBOT INTERNAL MULTI-DIVISI BERBASIS
RETRIEVAL-AUGMENTED GENERATION
MENGUNAKAN HYBRID RETRIEVER BM25 DAN
APPROXIMATE NEAREST NEIGHBOR**

**TUGAS AKHIR
APLIKATIF**

Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana

IBKA ANHAR FATCHA

UNIVERSITAS 41522010137

MERCU BUANA

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2026**

HALAMAN PERNYATAAN KARYA SENDIRI

Saya yang bertanda tangan di bawah ini:

Nama : Ibka Anhar Fatcha

NIM : 41522010137

Fakultas/Program Studi : Fakultas Ilmu Komputer / Teknik Informatika

Menyatakan dengan sebenar-benarnya bahwa Tugas Akhir berjudul:

“CHATBOT INTERNAL MULTI-DIVISI BERBASIS RETRIEVAL-AUGMENTED GENERATION MENGGUNAKAN HYBRID RETRIEVER BM25 DAN APPROXIMATE NEAREST NEIGHBOR” adalah hasil karya saya sendiri, tidak mengandung unsur plagiarisme, pelanggaran hak cipta, atau konten ilegal dalam bentuk apapun dan tidak melanggar hukum atau hak pihak manapun.

Apabila di kemudian hari ditemukan pelanggaran terhadap pernyataan ini, saya bersedia menanggung seluruh konsekuensi hukum dan membebaskan Universitas Mercu Buana dari segala bentuk tuntutan hukum dan saya siap mendapatkan sanksi akademis yang berlaku di Universitas Mercu Buana.

Demikian pernyataan ini saya buat dengan sebenar-benarnya untuk digunakan sebagaimana mestinya.

UNIVERSITAS
MERCU BUANA

Jakarta, 27 Juli 2026



Ibka Anhar Fatcha

PERNYATAAN SIMILARITY CHECK
/SIMILARITY CHECK STATEMENT

Saya yang bertanda tangan dibawah ini menyatakan, bahwa karya ilmiah yang ditulis oleh
/The undersigned, hereby declare that the scientific work written by

Nama /Name : Ibka Anhar Fatcha
NIM /Student ID Number : 41522010137
Program Studi /Study Program : Teknik Informatika

Dengan Judul Tugas Akhir

/The title:

“CHATBOT INTERNAL MULTI-DIVISI BERBASIS RETRIEVAL-AUGMENTED GENERATION MENGGUNAKAN HYBRID RETRIEVER BM25 DAN APPROXIMATE NEAREST NEIGHBOR”

telah dilakukan pengecekan *similarity* dengan sistem Turnitin pada tanggal:

/Similarity checks have been carried out with the Turnitin system on the date:

1 Juli 2026

dengan nilai persentase sebesar :

/with a percentage value of:

11%

dinyatakan memenuhi standar sesuai dengan ketentuan berlaku di **Fakultas Ilmu Komputer** Universitas Mercu Buana. */declared to meet standards in accordance with applicable regulations at the Faculty of Computer Science, Universitas Mercu Buana.*

File hasil cek similarity turnitin:

/Turnitin similarity report file

https://drive.google.com/file/d/1y1C9pvoH2KZkiV87Ke0z1WQdTu6uWu2s/view?usp=drive_link



Jakarta, 1 Juli 2026
Admin Turnitin Fasilkom UMB



Agung Prawoto, S.Kom., B.Sc
NIK : 322970503

HALAMAN PENGESAHAN

Tugas Akhir ini diajukan oleh:

Nama : Ibka Anhar Fatcha
NIM : 41522010137
Fakultas/Program Studi : Fakultas Ilmu Komputer / Teknik Informatika
Judul Tugas Akhir : CHATBOT INTERNAL MULTI-DIVISI BERBASIS
RETRIEVAL-AUGMENTED GENERATION
MENGUNAKAN HYBRID RETRIEVER BM25
DAN APPROXIMATE NEAREST NEIGHBOR

Telah berhasil dipertahankan pada sidang tanggal 27 Juli 2026 dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer Universitas Mercu Buana.

Disahkan oleh:

Pembimbing



Siti Maesaroh, S.Kom, M.T.

NIDN/NUPTK: 0413059003

Jakarta, 27 Juli 2026

Mengetahui,

Dekan Fakultas Ilmu Komputer



Dr. Bambang Jokonowo, S.Si., MTI
NIDN/NUPTK: 0320037002

Ketua Program Studi
Teknik Informatika



Dr. Hadi Santoso, S.Kom., M.Kom
NIDN/NUPTK: 0225067701

KATA PENGANTAR

Puji syukur saya panjatkan kepada Tuhan Yang Maha Esa, karena atas berkat dan rahmat-Nya, saya dapat menyelesaikan Tugas Akhir ini. Penulisan Tugas Akhir ini dilakukan dalam rangka memenuhi salah satu syarat untuk mencapai gelar Sarjana Teknik Informatika pada Fakultas Ilmu Komputer Universitas Mercu Buana. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan Tugas Akhir ini, sangatlah sulit bagi saya untuk menyelesaikan Tugas Akhir ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. Andi Adriansyah, M.Eng. selaku Rektor Universitas Mercu Buana.
2. Bapak Dr. Bambang Jokonowo, S.Si., MTI selaku Dekan Fakultas Ilmu Komputer.
3. Bapak Dr. Hadi Santoso, S.Kom., M.Kom. selaku Ketua Program Studi Teknik Informatika Universitas Mercu Buana.
4. Ibu Siti Maesaroh, S.Kom, M.T. selaku dosen pembimbing tugas akhir yang telah memberikan pengarahan, motivasi, menyediakan waktu, tenaga, dan pikiran sehingga selama pembuatan tugas akhir ini terjadwal dengan baik.
5. Kedua Orang Tua saya yang selalu mendukung saya selama menjalani masa studi sebagai mahasiswa Universitas Mercu Buana dan juga doa-doa nya.
6. Semua teman-teman yang selalu berbagi informasi dan memberikan dukungan dalam bentuk yang berbeda-beda.
7. Aulia Nur Soleha atas doa, motivasi serta kesabaran dalam memberikan dukungan moral dan menemani penulis hingga tugas akhir ini dapat diselesaikan

Akhir kata, saya berharap Tuhan Yang Maha Esa berkenan membalas segala kebaikan semua pihak yang telah membantu. Semoga Tugas Akhir ini membawa manfaat bagi pengembangan ilmu.

Jakarta, 27 Juli 2026

Ibka Anhar Fatcha

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR DI REPOSITORI UMB

Sebagai sivitas akademik Universitas Mercu Buana, saya yang bertanda tangan di bawah ini:

Nama : Ibka Anhar Fatcha
NIM : 41522010137
Fakultas/Program Studi : Fakultas Ilmu Komputer / Teknik Informatika
Judul Laporan Skripsi : CHATBOT INTERNAL MULTI-DIVISI
BERBASIS RETRIEVAL-AUGMENTED
GENERATION MENGGUNAKAN HYBRID
RETRIEVER BM25 DAN APPROXIMATE
NEAREST NEIGHBOR

Demi pengembangan ilmu pengetahuan, dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Non-Eksklusif** (*Non-exclusive Royalty-Free Right*) atas karya ilmiah saya yang berjudul di atas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti Non-Eksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan Tugas Akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

UNIVERSITAS
MERCU BUANA

Jakarta, 27 Juli 2026

Yang menyatakan,



Ibka Anhar Fatcha

CHATBOT INTERNAL MULTI-DIVISI BERBASIS RETRIEVAL-AUGMENTED GENERATION MENGGUNAKAN HYBRID RETRIEVER BM25 DAN APPROXIMATE NEAREST NEIGHBOR

IBKA ANHAR FATCHA

ABSTRAK

Perusahaan ataupun organisasi yang memiliki banyak divisi pastinya memiliki dokumen internal yang terbagi setiap divisi, seringkali informasi dan dokumen internal ini belum terintegrasi dengan baik, sehingga menyebabkan kelambatan dan juga kesulitan dalam mencari informasi yang relevan antar divisi. Permasalahan ini memperlambat proses penyebaran pengetahuan di lingkungan internal. Oleh sebab itu, penelitian ini merancang dan mengembangkan Chatbot Internal Multi-Divisi berbasis *RAG (Retrieval-Augmented Generation)* yang dapat menjawab pertanyaan pengguna sesuai dengan dokumen internal yang tersedia setiap divisi. Mekanisme *Hybrid Retrieval (BM25 dan ANN)* diimplementasikan dalam sistem sebagai mesin pencarian, dengan harapan meningkatkan relevansi jawaban yang dihasilkan *LLM*. Model *all-MiniLM-L6-v2* digunakan sebagai *embedder* untuk menghasilkan representasi vektor dari dokumen internal dan pertanyaan pengguna. Aplikasi yang dibuat berbasis web dengan struktur *backend* yang mengatur data dan proses pencarian, serta antarmuka pengguna untuk interaksi percakapan dan pengelolaan dokumen. Evaluasi sistem dilakukan menggunakan *framework* RAGAS untuk mengukur kualitas hasil *retrieval* dan respons chatbot, serta pengukuran *latensi* sistem untuk menilai kinerja *backend*. Hasil yang diharapkan dari penelitian ini adalah rancangan awal chatbot internal yang dapat meningkatkan efisiensi dalam mencari informasi dan mendukung pengelolaan dokumen antar divisi secara lebih terhubung.

Kata kunci: Chatbot, Multi-Divisi, *RAG*, *Hybrid Retrieval*, *LLM*

INTERNAL MULTI-DIVISION CHATBOT BASED ON RETRIEVAL-AUGMENTED GENERATION USING HYBRID RETRIEVER BM25 AND APPROXIMATE NEAREST NEIGHBOR

IBKA ANHAR FATCHA

ABSTRACT

Companies or organizations with multiple divisions often have internal documents separated by division. Frequently, this information and documentation are not well integrated, leading to delays and difficulties in finding relevant information across divisions. This issue slows down the internal knowledge dissemination process. Therefore, this study designs and develops a Multi-Division Internal Chatbot based on *RAG* (*Retrieval-Augmented Generation*) capable of answering user queries according to the internal documents available in each division. The *Hybrid Retrieval* method (*BM25* and *ANN*) is implemented in the system as a search engine to improve the relevance of responses generated by the *LLM*. The *all-MiniLM-L6-v2* model is used as the *embedder* to produce *vector* representations of internal documents and user queries. The application is web-based, with a *backend* structure managing data and *retrieval* processes, as well as a user interface for conversational interaction and document management. The system evaluation was conducted using the *RAGAS framework* to measure the quality of *retrieval* results and chatbot responses, as well as system *latency* measurements to assess *backend* performance. The expected outcome is an initial chatbot design that can enhance efficiency in *information retrieval* and support more connected document management across divisions.

Keywords: Chatbot, Multi-Divisi, *RAG*, *Hybrid Retrieval*, *LLM*

DAFTAR ISI

HALAMAN SAMPUL.....	0
HALAMAN JUDUL	i
HALAMAN PERNYATAAN KARYA SENDIRI	ii
HALAMAN SURAT KETERANGAN HASIL UJI TURNITIN	iii
HALAMAN PENGESAHAN.....	iv
KATA PENGANTAR.....	v
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR DI REPOSITORI UMB.....	vi
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI.....	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR.....	xii
DAFTAR LAMPIRAN	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah	3
1.3 Tujuan Penelitian	3
1.4 Batasan Penelitian.....	3
BAB II TINJAUAN PUSTAKA	5
2.1 Penelitian Terdahulu	5
2.2 Teori Pendukung.....	26
2.2.1 <i>LLM</i>	26
2.2.2 <i>Chatbot</i>	27
2.2.3 <i>RAG</i>	27
2.2.4 <i>BM25</i>	28
2.2.5 <i>ANN</i>	29
2.2.6 <i>Embedding</i>	30
2.2.7 <i>Reciprocal Rank Fusion (RRF)</i>	30
2.2.8 <i>Hybrid Retrieval</i>	31
2.2.9 <i>OpenSearch</i>	32
2.2.10 <i>PostgreSQL</i>	32

2.2.11	<i>MinIO</i>	32
2.2.12	Evaluasi	32
2.3	Gap Penelitian	36
BAB III METODE PENELITIAN		38
3.1	Pendekatan Penelitian	38
3.2	Desain Penelitian	39
3.2.1	Identifikasi Permasalahan.....	39
3.2.2	Analisis Kebutuhan Sistem	40
3.2.3	Perancangan Sistem <i>Backend</i>	41
3.2.4	Pengembangan Sistem.....	41
3.2.5	Pengujian dan Evaluasi Sistem.....	43
3.2.6	Konfigurasi Pengujian dan Skenario Evaluasi	44
3.3	Subjek Penelitian	46
3.4	Instrumen Penelitian	46
3.5	Teknik Pengumpulan Data	48
3.6	Analisis Data	49
3.7	Prosedur Penelitian	49
3.8	Evaluasi Hasil Penelitian	51
3.9	Timeline Penelitian	53
BAB IV PEMBAHASAN		55
4.1	Hasil Penelitian	55
4.1.1	Arsitektur Sistem.....	55
4.1.2	<i>Pipeline RAG</i>	57
4.1.3	<i>ERD (Entity Relationship Diagram)</i>	59
4.1.4	Hasil Implementasi Sistem.....	61
4.2	Pembahasan.....	83
4.2.1	Skenario Pengujian.....	84
4.2.2	Evaluasi Kualitas <i>Retrieval</i>	87
4.2.3	Evaluasi Kualitas Sistem RAG.....	90
4.2.4	Evaluasi Kinerja Sistem	95
BAB V KESIMPULAN DAN SARAN		98
5.1	Kesimpulan	98
5.2	Saran	99
DAFTAR PUSTAKA.....		101
LAMPIRAN.....		107

DAFTAR TABEL

Tabel 2.1 Penelitian Terkait	5
Tabel 3.1 Perencanaan <i>Sprint</i> 42	
Tabel 3.2 Contoh Skenario Pengujian dan Komponen Evaluasi	45
Tabel 3.3 Timeline Penelitian	54
Tabel 4.1 <i>Endpoint API</i> Layanan <i>ExpressJS</i> 80	
Tabel 4.2 <i>Endpoint API</i> Layanan <i>FastAPI</i>	82
Tabel 4.3 Konfigurasi Pengujian	85
Tabel 4.4 Skenario Pertanyaan Divisi Human Capital	85
Tabel 4.5 Skenario Pertanyaan Divisi Produksi	86
Tabel 4.6 Hasil Evaluasi Retrieval Divisi Human Capital	87
Tabel 4.7 Hasil Evaluasi Retrieval Divisi Produksi	88
Tabel 4.8 Rata-Rata Keseluruhan Hasil Evaluasi Retrieval	89
Tabel 4.9 Hasil Evaluasi RAGAS Divisi Human Capital	91
Tabel 4.10 Hasil Evaluasi RAGAS Divisi Produksi	92
Tabel 4.11 Rata-Rata Keseluruhan Hasil Evaluasi	94
Tabel 4.12 Hasil Evaluasi Kinerja Sistem	95



DAFTAR GAMBAR

Gambar 3.1 Desain Penelitian.....	39
Gambar 4.1 Arsitektur Sistem.....	55
Gambar 4.2 <i>RAG Pipeline</i>	58
Gambar 4.3 <i>ERD</i>	60
Gambar 4.4 Hasil <i>Retrieval</i> pada Divisi Human Capital	63
Gambar 4.5 Daftar Dokumen Divisi Human Capital yang Tersimpan pada <i>OpenSearch</i>	64
Gambar 4.6 Hasil <i>Retrieval</i> pada Divisi Produksi	64
Gambar 4.7 Hasil <i>Retrieval ANN</i>	66
Gambar 4.8 Hasil <i>Retrieval BM25</i>	68
Gambar 4.9 Hasil <i>Retrieval Hybrid</i>	71
Gambar 4.10 Hasil Log <i>Parsing</i> Dokumen	75
Gambar 4.11 Proses Generasi Jawaban	78
Gambar 4.12 Dokumentasi <i>API</i> Menggunakan <i>Swagger UI</i>	83



DAFTAR LAMPIRAN

Lampiran 1 Asistensi	107
Lampiran 2 Curriculum Vitae	108
Lampiran 3 Surat Pernyataan dan Penyerahan HAKI	109
Lampiran 4 BNSP	111

