

ABSTRAK

Marcelino Gavriel:

Skripsi

Implementasi Sistem API Chatbot untuk customer PT XYZ dengan Large Language Model menggunakan metode Hybrid Search RAG

PT XYZ adalah perusahaan agri-food di Indonesia yang memproduksi pakan ternak, bibit unggul, peternakan ayam, udang, ikan, hingga olahan hewan. Proses layanan informasi kepada customer masih dilakukan secara manual oleh admin customer service, sehingga sering terjadi keterlambatan respons dan pertanyaan yang berulang. Oleh karena itu, dalam tugas akhir kali ini dikembangkan sistem API chatbot sebagai solusi awal untuk membantu PT XYZ dalam menangani pertanyaan customer secara otomatis. Sistem ini juga dilengkapi web base admin untuk menambah dan memperbarui informasi pada database chatbot. Chatbot dibangun menggunakan model Llama-3.1 8B ArliAI IndoFormax dengan metode hybrid search Retrieval Augmented Generation (RAG), yang menggabungkan similarity search dan keyword search menggunakan pgvector, serta reranking sebelum konteks masuk ke dalam pipeline LLM.

Web base admin dikembangkan dengan framework Streamlit dan basis data PostgreSQL, serta menggunakan Sentence Transformer Indo-Sentence-BERT untuk embedding informasi. Pengujian sistem dilakukan bersama tim admin dan IT PT XYZ untuk mengevaluasi kemampuan chatbot dalam memberikan jawaban yang relevan serta kemudahan pengelolaan data melalui web base yang telah dibuat.

Kata kunci : chatbot, api, llm, hybrid search, similarity search, keyword search, reranking

ABSTRACT

Marcelino Gavriel:

Undergraduate thesis

Implementation of a Chatbot API System for PT XYZ Customers Using a Large Language Model with the Hybrid Search RAG Method

PT XYZ is an agri-food company in Indonesia producing animal feed, superior livestock, poultry, shrimp, fish, and processed animal products. Customer inquiries regarding products and sales information are still handled manually by customer service staff, often causing delayed responses and repeated questions. Therefore, in this final project, an API-based chatbot system was developed as an initial solution to help PT XYZ manage customer inquiries automatically. Additionally, a web-based admin system was provided to enable administrators to add and update information in the chatbot's database. The chatbot is built using the Llama-3.1 8B ArliAI IndoFormax model with a hybrid search Retrieval Augmented Generation (RAG) method, combining similarity search and keyword search using pgvector, as well as reranking before the context is processed through the LLM pipeline to generate relevant answers.

The admin web system was developed using the Streamlit framework and PostgreSQL database, utilizing the Indo-Sentence-BERT Sentence Transformer for information embedding. System testing was conducted together with PT XYZ's admin and IT teams to evaluate the chatbot's ability to provide accurate responses and the ease of managing chatbot data through the web-based system.

keyword sentences : chatbot, api, llm, hybrid search, similarity search, keyword search, reranking

DAFTAR ISI

HALAMAN JUDUL	i
LEMBAR PENGESAHAN.....	ii
LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIK	iii
KATA PENGANTAR.....	iv
ABSTRAK	v
ABSTRACT	vi
DAFTAR ISI	vii
DAFTAR GAMBAR	x
DAFTAR TABEL.....	xii
1.PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah	3
1.3 Tujuan	4
1.4 Ruang Lingkup.....	4
1.5 Metodologi Penelitian	6
1.6 Manfaat Penelitian	8
1.7 Sistematika Penulisan	8
2.LANDASAN TEORI	10
2.1 Tinjauan Pustaka.....	10
2.1.1 Chatbot.....	10
2.1.2 Generative AI.....	10
2.1.3 Large Language Model (LLM)	11
2.1.4 Word & sentence embedding	14
2.1.5 Retrieval Augmented Generation (RAG).....	15
2.1.6 Hybrid Search	15
2.1.7 Reranking	16

2.2	Tinjauan Studi	17
2.2.1	Tabel State of Art	20
3.	ANALISIS DAN DESAIN SISTEM	23
3.1	Analisis Kebutuhan.....	23
3.2	Pengolahan Data.....	24
3.3	Desain Sistem.....	27
3.3.1.	Desain sistem API chatbot	27
3.3.2.	Desain sistem web based admin.....	32
3.3.2.1	Activity diagram admin login.....	33
3.3.2.2	Flowchart diagram admin add new knowledge type product knowledge..	34
3.3.2.4	Flowchart diagram admin add new knowledge type agents sales.....	36
3.2.3.5	Flowchart admin add new knowledge (add new topic)	37
3.3.2.6	Flowchart admin edit knowledge	38
3.3.2.7	Flowchart save_data process	39
3.3.2.8	Flowchart chunking_data_process.....	40
3.3.2.9	Flowchart chunking_optimalization	41
3.3.2.10	Flowchart chunking_optimalization_PA.....	43
3.3.2.11	Flowchart split_sisatokens	44
3.3.2.12	Flowchart chunksplit	46
3.3.2.13	Flowchart chunking_optimalization (topic = FAQ)	47
3.3.2.14	Flowchart splitsubchunk process	48
3.3.2.15	Flowchart chunk_with_overlap.....	50
3.3.2.16	Activity diagram admin delete knowledge.....	52
3.4	Struktur database	52
4.	PENGUJIAN.....	57
4.1	Pengujian kestabilan API sistem chatbot.....	57
4.2	Pengujian hybrid search serta reranking process.....	58
4.3	Pengujian Sistem chatbot	60
5 .	KESIMPULAN DAN SARAN	66

5.1 Kesimpulan.....	66
5.2 Saran	66
DAFTAR REFRENSI	67

DAFTAR GAMBAR

Gambar 2.1 Arsitektur Transformer.....	12
Gambar 3.4 Flowchart sistem API chatbot.....	27
Gambar 3.5 activity diagram rag_chat process.....	29
Gambar 3.6 flowchart guide_knowledge.....	30
Gambar 3.7 flowchart classify jawaban	31
Gambar 3.8 Activity diagram admin login.....	33
Gambar 3.9 flowchart diagram admin add new knowledge type product knowledge	34
Gambar 3.10 flowchart admin add new knowledge type FAQ.....	35
Gambar 3.11 flowchart admin add new knowledge type agent sales.....	36
Gambar 3.12 flowchart admin add new knowledge topic.....	37
Gambar 3.13 flowchart admin edit knowledge	38
Gambar 3.14 flowchart save_data process.....	39
Gambar 3.15 flowchart chunking_data_process	40
Gambar 3.16 Flowchart chunking_optimalization (topic = Product knowledge or Agent knowledge).....	42
Gambar 3.17 flowchart chunking_optimalization_PA	43
Gambar 3.18 Flowchart subtemp_chunking	44
Gambar 3.19 flowchart split_sisatokens.....	45
Gambar 3.20 Flowchart chunk_split	46
Gambar 3.21 Flowchart chunking_optimalization (FAQ).....	48
Gambar 3.22 Flowchart splitsubchunk process	48
Gambar 3.23 Flowchart chunk_with_overlap.....	51
Gambar 3.24 Flowchart admin delete knowledge.....	52
Gambar 3.25 Entity relationship diagram	53
Gambar 4.1 Hasil survey tingkat relevan jawaban yang diberikan oleh sistem chatbot	62
Gambar 4.2 Hasil survey keakuratan jawaban yang diberikan sistem chatbot terhadap informasi yang terdapat di web based admin.....	63

Gambar 4.3 Hasil survey seberapa baik pemahaman chatbot dalam menanggapi pertanyaan user dari berbagai macam varian bahasa.	63
Gambar 4.4 Hasil survey seberapa cepat waktu yang dibutuhkan chatbot untuk menghasilkan jawaban dari pertanyaan?	64
Gambar 4.5 Hasil survey seberapa baik UI website	65
Gambar 4.6 Hasil survey tingkat kemudahan admin dalam menambahkan informasi baru.....	65

DAFTAR TABEL

Tabel 2.1 State of Art	20
Tabel 3.1 mtadmin	53
Tabel 3.2 mtadminknowledgebase	54
Tabel 3.3 mtknowledgebase	54
Tabel 3.4 mtlogquestionuser	55
Tabel 3.5 mtloguserating	55
Tabel 3.6 mtloguseresponse	56
Tabel 4.1 Pengujian API.....	57
Tabel 4.2 pengujian hybrid search, serta reranking result.....	58
Tabel 4.3 Pengujian API server.....	61