RANCANG BANGUN SEARCH ENGINE PADA PT XYZ MENGGUNAKAN ALGORITMA BERT-RAG DAN GPT-4

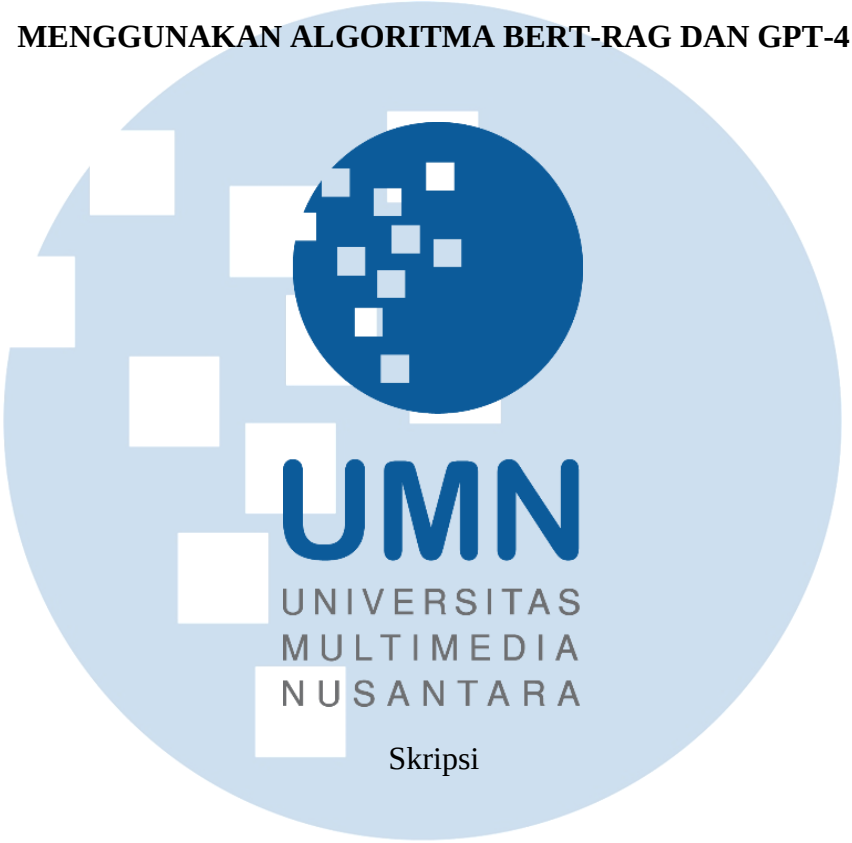
Skripsi

Virginia Lim
000000055667

PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2024

RANCANG BANGUN SEARCH ENGINE PADA PT XYZ

MENGGUNAKAN ALGORITMA BERT-RAG DAN GPT-4



Skripsi

Diajukan sebagai Salah Satu Syarat untuk Memperoleh

Gelar Sarjana Komputer (S.Kom)

Virginia Lim

00000055667

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS TEKNIK DAN INFORMATIKA

UNIVERSITAS MULTIMEDIA NUSANTARA

TANGERANG

2024

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,
Nama : Virginia Lim
Nomor Induk Mahasiswa : 00000055667
Program studi : Sistem Informasi

Skripsi dengan judul:
Rancang Bangun *Search Engine* pada PT XYZ Menggunakan Algoritma BERT-RAG dan GPT-4
merupakan hasil karya saya sendiri bukan plagiat dari karya ilmiah yang ditulis oleh orang lain, dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.
Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan skripsi maupun dalam penulisan laporan skripsi, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk Tugas Akhir yang telah saya tempuh.

Tangerang, 16 Desember 2024

Virginia Lim

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Yang bertanda tangan di bawah ini:

Nama : Virginia Lim

NIM : 00000055667

Program Studi : Sistem Informasi

Jenjang : S1

Judul Karya Ilmiah : Rancang Bangun *Search Engine* pada PT XYZ Menggunakan Algoritma BERT-RAG dan GPT-4

Menyatakan dengan sesungguhnya bahwa saya bersedia* **(pilih salah satu):**

☐ Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya ke dalam repositori Knowledge Center sehingga dapat diakses oleh Sivitas Akademika UMN/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial.

☑ Saya tidak bersedia mempublikasikan hasil karya ilmiah ini ke dalam repositori Knowledge Center, dikarenakan: dalam proses pengajuan publikasi ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*) **.

☐ Lainnya, pilih salah satu:

☐ Hanya dapat diakses secara internal Universitas Multimedia Nusantara

☐ Embargo publikasi karya ilmiah dalam kurun waktu 3 tahun.

Tangerang, 7 Januari 2024

(Virginia Lim)

* Pilih salah satu

** Jika tidak bisa membuktikan LoA jurnal/HKI, saya bersedia mengizinkan penuh karya ilmiah saya untuk dipublikasikan ke KC UMN dan menjadi hak institusi UMN.

KATA PENGANTAR

Puji dan Syukur saya panjatkan kepada Tuhan Yang Maha Esa atas segala rahmat-Nya, yang memungkinkan saya menyelesaikan penulisan skripsi ini dengan judul Rancang Bangun Search Engine pada PT XYZ menggunakan Algoritma BERT-RAG dan GPT-4. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana di Program Studi Sistem Informasi, Fakultas Teknik dan Informatika, Universitas Multimedia Nusantara. Saya menyadari bahwa pencapaian ini tidak lepas dari bimbingan dan dukungan berbagai pihak. Oleh karena itu, saya mengucapkan terima kasih kepada:

1. Bapak Dr. Andrey Andoko, M.Sc, selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng. Niki Prastomo, selaku Dekan Fakultas Teknik Informatika Universitas Multimedia Nusantara.
3. Ibu Ririn Ikana Desanti, S.Kom., M.Kom., selaku Ketua Program Studi Sistem Informasi Universitas Multimedia Nusantara.
4. Ibu Dinar Ajeng Kristiyanti, S.Kom., M.Kom., sebagai Pembimbing yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
5. Keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tugas akhir ini.

Semoga karya ilmiah ini ini bisa menjadi referensi penelitian yang baik umtuk berbagai penelitian di kemudian hari.

Tangerang, 16 Desember 2024

Virginia Lim

Rancang Bangun *Search Engine* pada PT XYZ menggunakan Algoritma BERT-RAG dan GPT-4

Virginia Lim

ABSTRAK

Industri 4.0 telah membawa transformasi besar dalam pengelolaan informasi, dengan tantangan utama berupa peningkatan volume data, termasuk dokumen tidak terstruktur seperti PDF. PT. XYZ menghadapi kesulitan dalam memastikan informasi yang dapat diakses secara cepat, akurat, dan relevan oleh karyawan. Metode pencarian tradisional seperti dokumen dalam bentuk kertas sering kali menghasilkan waktu pencarian yang lama dan hasil yang kurang sesuai, sehingga memengaruhi efisiensi kerja serta pengambilan keputusan. Penelitian ini bertujuan merancang *search engine* berbasis *web* menggunakan *hybrid* model, yaitu *Bidirectional Encoder Representations from Transformers* (BERT) dan *Retrieval-Augmented Generation* (RAG), serta *Application Program Interface* (API) GPT-4 untuk menghasilkan pencarian yang relevan, akurat, dan kontekstual.

Data yang kompleks memerlukan pendekatan terstruktur untuk pengelolaan dan analisis yang efektif. Penelitian ini menggunakan *framework data mining Cross-Industry Standard Process for Data Mining* (CRISP-DM), dimulai dari memahami proses bisnis perusahaan, pemilihan data, *data preparation* dengan tahapan ekstrasi hingga *encoding* PDF. Selanjutnya *data modelling* menggunakan *hybrid* model BERT-RAG dan GPT-4, evaluasi model, dan *deployment*. Pada tahap *deployment* menggunakan metode *prototype* untuk membangun antarmuka *web*, memungkinkan partisipasi aktif pengguna dalam proses pengembangan. Kombinasi CRISP-DM dan *prototyping* memastikan sistem *search engine* yang inovatif, relevan, dan *user-friendly*.

Hasil dari penelitian ini berupa *search engine* berbasis *website* yang dapat meningkatkan kemampuan pencarian data. Hal ini ditandai dengan akurasi model mencapai 86% meningkat 4% dari penelitian sebelumnya dalam kesamaan antara *keyword* yang dimasukkan dengan hasil yang ditampilkan. Hasil akurasi yang didapatkan dihitung dengan menggunakan *consine similiraty*. Pengujian *website* juga dilakukan menggunakan *User Acceptance Testing* (UAT) dengan evaluasi berhasil.

Kata kunci: API GPT-4, BERT , *Retrieval-Augmented Generation, Search Engine.*

Design and Development of a Search Engine at PT XYZ using BERT-RAG Algorithm and GPT-4

Virginia Lim

ABSTRACT

Industry 4.0 has brought significant transformations in information management, with a primary challenge being the increasing volume of data, including unstructured documents such as PDFs. PT. XYZ faces difficulties in ensuring information can be accessed quickly, accurately, and relevantly by employees. Traditional search methods, such as paper-based documents, often result in prolonged search times and less relevant results, impacting work efficiency and decision-making. This research aims to design a web-based search engine using a hybrid model, namely Bidirectional Encoder Representations from Transformers (BERT) and Retrieval-Augmented Generation (RAG), along with the GPT-4 Application Programming Interface (API) to deliver relevant, accurate, and contextual searches.

Complex data requires a structured approach for effective management and analysis. This research employs the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework, starting with understanding the company's business processes, data selection, and data preparation involving extraction and encoding of PDFs. Subsequently, the data modeling phase uses the hybrid BERT-RAG and GPT-4 model, followed by model evaluation and deployment. The deployment phase applies the prototype method to build a web interface, enabling active user participation in the development process. The combination of CRISP-DM and prototyping ensures an innovative, relevant, and user-friendly search engine system.

The results of this research include a website-based search engine capable of enhancing data search capabilities. This is evidenced by a model accuracy rate of 86%, an improvement of 4% over previous studies, in terms of the similarity between input keywords and displayed results. The accuracy obtained was calculated using cosine similarity. Website testing was also conducted through User Acceptance Testing (UAT), with successful evaluations.

***Keywords**: API GPT-4, BERT, Retrieval-Augmented Generation, Search Engine.*

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR RUMUS

DAFTAR LAMPIRAN