SKRIPSI

PENERAPAN SENTENCE-BERT DAN COSINE SIMILARITY UNTUK PENCARIAN SEMANTIK DOKUMEN SKRIPSI DALAM FORMAT PDF

MUHAMMAD ABDUL HAFIZH FATHUDDIN

NPM 21081010225

DOSEN PEMBIMBING

Eka Prakarsa Mandyartha, S.T., M.Kom
Afina Lina Nurlaili, S.Kom., M.Kom

KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI INFORMATIKA
SURABAYA
2025

SKRIPSI

PENERAPAN SENTENCE-BERT DAN COSINE SIMILARITY UNTUK PENCARIAN SEMANTIK DOKUMEN SKRIPSI DALAM FORMAT PDF

MUHAMMAD ABDUL HAFIZH FATHUDDIN

NPM 21081010225

DOSEN PEMBIMBING

Eka Prakarsa Mandyartha, S.T., M.Kom
Afina Lina Nurlaili, S.Kom., M.Kom

KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI INFORMATIKA
SURABAYA
2025

Halaman ini sengaja dikosongkan

LEMBAR PENGESAHAN

PENERAPAN SENTENCE-BERT DAN COSINE SIMILARITY UNTUK PENCARIAN SEMANTIK DOKUMEN SKRIPSI DALAM FORMAT PDF

Oleh :
MUHAMMAD ABDUL HAFIZH FATHUDDIN
NPM. 21081010225

Telah dipertahankan dihadapan dan diterima oleh Tim Penguji Skripsi Prodi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur Pada tanggal 11 September 2025

Eka Prakarsa Mandyartha, S.T., M.Kom NIP. 198805252018031001	.....	**(Pembimbing I)**
Afina Lina Nurlaili, S.Kom., M.Kom NIP. 199312132022032010	.....	**(Pembimbing II)**
Made Hanindia Prami S, S.Kom., M.Cs NIP.198902052018032001	.....	**(Ketua Penguji)**
Achmad Junaidi, S.Kom., M.Kom NIP.378110401991	.....	**(Anggota Penguji)**

Mengetahui,
Dekan Fakultas Ilmu Komputer

Prof. Dr. Ir. Novirina Hendrasarie, MT
NIP. 19681126 199403 2 001

Dipindai dengan CamScanner

Halaman ini sengaja dikosongkan

LEMBAR PERSETUJUAN

PENERAPAN SENTENCE-BERT DAN COSINE SIMILARITY UNTUK PENCARIAN SEMANTIK DOKUMEN SKRIPSI DALAM FORMAT PDF

Oleh :
MUHAMMAD ABDUL HAFIZH FATHUDDIN
NPM. 21081010225

Menyetujui,
Koordinator Program Studi Informatika
Fakultas Ilmu Komputer

Fetty Tri Anggraeny, S.Kom., M.Kom
NIP. 19820211 2021212 005

Dipindai dengan CamScanner

Halaman ini sengaja dikosongkan

SURAT PERNYATAAN BEBAS PLAGIASI

Saya yang bertanda tangan dibawah ini :

Nama	: Muhammad Abdul Hafizh Fathuddin
NPM	: 21081010225
Program	: Sarjana(S1)
Program Studi	: Informatika
Fakultas	: Ilmu Komputer

Menyatakan bahwa dalam dokumen ilmiah Tugas Akhir Skripsi ini tidak terdapat bagian dari karya ilmiah lain yang telah diajukan untuk memperoleh gelar akademik di suatu lembaga Pendidikan Tinggi, dan juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang/lembaga lain, kecuali yang secara tertulis disitasi dalam dokumen ini dan disebutkan secara lengkap dalam daftar pustaka.

Dan saya menyatakan bahwa dokumen ilmiah ini bebas dari unsur-unsur plagiasi. Apabila dikemudian hari ditemulan indikasi plagiat pada Skripsi ini, saya bersedia menerima sanksi sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian surat pernyataan ini saya buat dengan sesungguhnya tanpa ada paksaan dari siapapun juga dan untuk dipergunakan sebagaimana mestinya.

Surabaya, 10 September 2025
Yang Membuat Pernyataan,

MUHAMMAD ABDUL HAFIZH FATHUDDIN
NPM. 21081010225

Dipindai dengan CamScanner

Halaman ini sengaja dikoso

ABSTRAK

Nama Mahasiswa / NPM	:	Muhammad Abdul Hafizh Fathuddin / 21081010225
Judul Skripsi	:	Penerapan Sentence-Bert dan Cosine Similarity Untuk Pencarian Semantik Dokumen Skripsi Dalam Format PDF
Dosen Pembimbing	:	1. Eka Prakarsa Mandyartha, S.T., M.Kom 2. Afina Lina Nurlaili, S.Kom., M.Kom

Pencarian dokumen skripsi pada repositori digital umumnya masih terbatas pada pencocokan kata kunci sehingga sering menghasilkan temuan yang kurang relevan. Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk membangun sistem pencarian semantik dokumen skripsi dalam format PDF dengan memanfaatkan Sentence-BERT (SBERT) dan metode Cosine Similarity yang dipadukan dengan ontologi untuk memperkaya pemahaman makna query. Sistem ini dirancang agar mampu memahami maksud pengguna secara lebih mendalam, baik ketika query diberikan dalam bentuk kata, frasa, kalimat, maupun paragraf. Tahapan penelitian meliputi ekstraksi teks dari dokumen PDF, preprocessing, tokenisasi WordPiece, serta pembentukan vektor representasi kalimat menggunakan SBERT. Skor relevansi dihitung dengan kombinasi bobot cosine similarity (0,7) dan ontologi (0,3) sehingga sistem dapat menampilkan dokumen dengan makna paling mendekati query. Hasil pengujian menunjukkan bahwa sistem mampu memberikan hasil pencarian yang relevan dengan nilai Mean Reciprocal Rank (MRR) konsisten sebesar 1.0 pada semua jenis query. Nilai Precision rata-rata mencapai 0,80 dan Recall rata-rata sebesar 0,92. Perbandingan dengan metode Keyword Matching menunjukkan bahwa metode semantik lebih unggul dengan Precision rata-rata 0,88 dan Recall 0,65 dibandingkan keyword yang hanya mencapai Precision 0,24 dan Recall 0,12. Temuan ini membuktikan bahwa sistem semantik efektif dalam menempatkan dokumen paling relevan di peringkat teratas dan lebih unggul dibandingkan pencarian berbasis kata kunci, meskipun cakupan hasil masih perlu ditingkatkan melalui pengayaan ontologi dan perluasan dataset.

Kata kunci : Pencarian Semantik, Sentence-BERT, Cosine Similarity, Ontology, Dokumen Skripsi.

Halaman ini sengaja dikosongkan

ABSTRACT

Student Name / NPM	:	Muhammad Abdul Hafizh Fathuddin / 21081010225
Thesis Title	:	PImplementation of Sentence-BERT and Cosine Similarity for Semantic Search of Thesis Documents in PDF Format
Advisor	:	1. Eka Prakarsa Mandyartha, S.T., M.Kom 2. Afina Lina Nurlaili, S.Kom., M.Kom

The search for thesis documents in digital repositories is generally limited to keyword matching, which often produces less relevant results. To address this issue, this study develops a semantic search system for thesis documents in PDF format by utilizing Sentence-BERT (SBERT) and the Cosine Similarity method, combined with ontology to enrich the understanding of query meanings. The research stages include text extraction from PDF documents, preprocessing, WordPiece tokenization, and sentence vector representation using SBERT, with relevance scores calculated by combining cosine similarity (0.7) and ontology (0.3) weights. The evaluation results show that the system is capable of producing relevant search results with a consistent Mean Reciprocal Rank (MRR) of 1.0 across all query types. The average Precision reached 0.80, while the average Recall was 0.92. A comparison with the Keyword Matching method shows that the semantic approach performs better, with an average Precision of 0.88 and Recall of 0.65, compared to keyword matching which only achieved 0.24 for Precision and 0.12 for Recall. These findings demonstrate that the semantic system effectively places the most relevant documents at the top rank and outperforms keyword-based search, although the coverage of relevant results still needs to be improved through ontology enrichment and dataset expansion.

Keywords: Semantic Search, Sentence-BERT, Cosine Similarity, Ontology, Thesis Documents.

Halaman ini sengaja dikosongkan

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT atas limpahan rahmat, hidayah, dan karunia-Nya, sehingga penulis dapat menyelesaikan skripsi dengan judul **"Penerapan Sentence-BERT dan Cosine Similarity untuk Pencarian Semantik Dokumen Skripsi dalam Format PDF"**. Skripsi ini disusun sebagai salah satu syarat untuk menyelesaikan studi Strata-1 pada Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional "Veteran" Jawa Timur.

Dalam proses penyusunan skripsi ini, penulis merasakan betul bahwa perjalanan ini bukanlah sesuatu yang mudah. Ada banyak hambatan, kesulitan, bahkan rasa lelah yang harus dilalui. Namun, semua itu bisa terlewati berkat doa, dukungan, serta bantuan dari banyak pihak. Untuk itu, dengan penuh rasa hormat dan ketulusan, penulis ingin menyampaikan ucapan terima kasih kepada:

1. Abi Wali Muhammad dan Umi Fadilah, orang tua tercinta. Terima kasih atas segala doa yang tidak pernah terputus, kasih sayang yang selalu tulus, serta dukungan moril maupun materiil yang diberikan sejak awal hingga sekarang. Tanpa doa dan pengorbanan beliau berdua, penulis tidak akan mampu menyelesaikan pendidikan ini. Setiap langkah yang penulis ambil tidak pernah lepas dari restu dan doa orang tua, yang menjadi kekuatan terbesar dalam menyelesaikan skripsi ini.
2. Bapak Dr. Ir. I Gede Susrama Mas Diyasa, ST., MT., IPU selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional "Veteran" Jawa Timur, atas segala dukungan serta fasilitas yang diberikan sehingga penulis dapat menjalani masa studi dengan baik.
3. Bapak Eka Prakarsa Mandyartha, S.T., M.Kom selaku Dosen Pembimbing I dan Ibu Afina Lina Nurlaili, S.Kom., M.Kom selaku Dosen Pembimbing II. Terima kasih yang sebesar-besarnya atas kesediaan waktu, ilmu, serta kesabaran dalam membimbing penulis. Berkat arahan dan masukan dari beliau berdua, penulis dapat belajar banyak hal baru, tidak hanya terkait penelitian, tetapi juga tentang bagaimana bersikap lebih teliti, sabar, dan konsisten.

4. Ibu Fetty Tri Anggraeny, S.Kom., M.Kom selaku Ketua Program Studi Informatika Fakultas Ilmu Komputer UPN "Veteran" Jawa Timur, yang telah memberikan arahan, kebijakan, serta dukungan selama penulis menempuh perkuliahan.
5. Seluruh dosen di Program Studi Informatika, Fakultas Ilmu Komputer UPN "Veteran" Jawa Timur, yang telah mendidik, membimbing, dan membekali penulis dengan ilmu pengetahuan dan pengalaman berharga selama masa kuliah.
6. Seluruh staf administrasi Fakultas Ilmu Komputer, yang selalu membantu kelancaran urusan akademik, baik selama masa perkuliahan maupun saat penulis menyusun skripsi ini.
7. Keluarga dari pihak ibu yang dengan penuh ketulusan telah membolehkan penulis untuk tinggal di rumahnya selama proses studi, serta selalu menyediakan makanan, bekal, dan perhatian. Dukungan tersebut sangat membantu penulis untuk tetap fokus dalam menyelesaikan perkuliahan dan skripsi ini.
8. Sahabat-sahabat penulis, Fawwas, Fahmi, dan Dzaki, yang selalu hadir memberikan dukungan, semangat, serta kebersamaan yang tidak ternilai. Terima kasih atas waktu yang dihabiskan bersama, diskusi, bantuan, dan canda tawa yang membuat perjalanan kuliah dan penyusunan skripsi ini terasa lebih ringan.
9. Teman-teman Discord Azriel, Sopa, Mbappe, Wigi, Bang Didi, Bang Jack, dan Ayyash. Terima kasih sudah sering menjadi tempat berbagi cerita, bercanda, maupun sekadar mengalihkan penat di sela-sela kesibukan. Kehadiran kalian membuat proses ini terasa lebih menyenangkan dan tidak terlalu berat dijalani.
10. Firdaus, terima kasih sudah menjadi teman yang selalu ada untuk berbagi cerita, canda, maupun sekadar mengisi waktu di sela-sela kesibukan. Kehadiranmu membuat penulis merasa lebih ringan menjalani hari-hari yang penuh tekanan. Perhatian dan semangat yang kamu berikan, meski sederhana, sangat berarti dan membantu penulis tetap fokus menyelesaikan skripsi ini. Dukunganmu akan

11. selalu penulis ingat sebagai bagian dari perjalanan panjang yang akhirnya membawa penulis sampai pada tahap ini.
12. Seluruh teman-teman Informatika angkatan 2021, yang telah menjadi bagian dari perjalanan ini. Terima kasih atas kebersamaan, saling berbagi ilmu, kerja sama, serta dukungan yang membuat masa perkuliahan jauh lebih berwarna.
13. Semua pihak lain yang tidak dapat penulis sebutkan satu per satu, namun yang kontribusinya tetap sangat berarti. Setiap bentuk bantuan, doa, maupun dukungan dari mereka turut membantu penulis dalam menyelesaikan skripsi ini.

Penulis menyadari sepenuhnya bahwa skripsi ini masih jauh dari sempurna. Untuk itu, penulis sangat mengharapkan kritik dan saran yang membangun dari semua pihak demi penyempurnaan karya ini di masa mendatang.

Akhir kata, semoga skripsi ini dapat memberikan manfaat, baik bagi pembaca, maupun bagi perkembangan ilmu pengetahuan, khususnya di bidang Informatika. Penulis berharap hasil penelitian ini juga bisa menjadi langkah kecil yang berguna, sekaligus menjadi bentuk penghargaan dan rasa terima kasih penulis kepada semua pihak yang telah mendukung.

Surabaya, 10 September 2025

Penulis

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR LISTING