

**RETRIEVAL SKRIPSI
MENGUNAKAN MODIFIKASI BM25 BERBASIS
POSISI KATA ADAPTIF**

TESIS



**OLEH:
PUTU AGUNG ANANTA WIJAYA
2329101029**

**PROGRAM STUDI ILMU KOMPUTER
PASCASARJANA
UNIVERSITAS PENDIDIKAN GANESHA
2024**

TESIS

DIAJUKAN UNTUK MELENGKAPI TUGAS DAN MEMENUHI SYARAT-SYARAT UNTUK MENCAPAI GELAR MAGISTER KOMPUTER

Menyetujui

Pembimbing I	Dr. Ir. I Made Gede Sunarya, S.Kom., M.Cs. NIP.198307252008011008
Pembimbing II	Dr. I Gede Aris Gunadi, S.Si. M.Kom. NIP.197703182008121004



- UU ITE No. 11 Tahun 2008 Pasal 5 Ayat 1 "Informasi Elektronik dan/atau hasil cetaknya merupakan alat bukti hukum yang sah"
- Dokumen ini telah ditandatangani secara elektronik menggunakan sertifikat elektronik yang diterbitkan BSrE - BSSN, validitas dokumen elektronik ini bisa dicek menggunakan aplikasi mobile VeryDS oleh BSrE
- Cetakan dokumen ini merupakan salinan dari file dokumen bertandatangan elektronik yang keabsahannya dapat diakses melalui scan QRCode yang terdapat pada sertifikat ini.

Tesis oleh Putu Agung Ananta Wijaya telah dipertahankan di depan tim penguji dan dinyatakan diterima sebagai salah satu persyaratan untuk memperoleh gelar Magister Komputer di Ilmu Komputer (S2), Program Pascasarjana, Universitas Pendidikan Ganesha
Pada tanggal 28 Januari 2026

Dewan Penguji

Ketua	Dr. Ir. I Made Gede Sunarya, S.Kom., M.Cs. NIP.198307252008011008
Anggota	Dr. I Gede Aris Gunadi, S.Si. M.Kom. NIP.197703182008121004
Anggota	Dr. Putu Hendra Suputra, S.Kom., M.Cs. NIP.198212222006041001
Anggota	Prof. Dr. Ir. Dewa Gede Hendra Divayana, S.Kom., M.Kom., MCE., IPU., ASEAN.Eng., APEC.Eng. NIP.198407242015041002

Mengetahui Direktur Pascasarjana
Universitas Pendidikan Ganesha,



Prof. Dr. I Nyoman Jampel, M.Pd.

NIP.195910101986031003



- UU ITE No. 11 Tahun 2008 Pasal 5 Ayat 1 "Informasi Elektronik dan/atau hasil cetaknya merupakan alat bukti hukum yang sah"
- Dokumen ini telah ditandatangani secara elektronik menggunakan sertifikat elektronik yang diterbitkan BSrE - BSSN, validitas dokumen elektronik ini bisa dicek menggunakan aplikasi mobile VeryDS oleh BSrE
- Cetakan dokumen ini merupakan salinan dari file dokumen bertandatangan elektronik yang keabsahannya dapat diakses melalui scan QRCode yang terdapat pada sertifikat ini.

Lembar Pernyataan

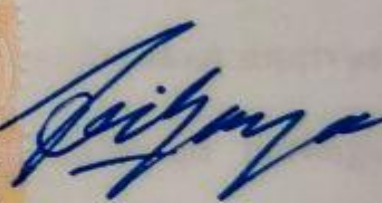
Saya menyatakan dengan sesungguhnya bahwa tesis yang saya susun sebagai syarat untuk memperoleh gelar Magister Pendidikan dari Pascasarjana Universitas Pendidikan Ganesha seluruhnya merupakan hasil karya saya sendiri. Bagian-bagian tertentu dalam penulisan tesis yang saya kutip dari hasil karya orang lain telah dituliskan sumbernya secara jelas dan sesuai dengan norma, kaidah, serta etika akademis.

Apabila di kemudian hari ditemukan seluruh atau sebagian tesis ini bukan hasil karya saya sendiri atau adanya plagiat dalam bagian-bagian tertentu, saya bersedia menerima sanksi pencabutan gelar akademik yang saya sandang dan sanksi-sanksi lainnya sesuai dengan peraturan perundang-undangan yang berlaku di wilayah Negara Kesatuan Republik Indonesia.



Denpasar, 8 September 2024

Yang memberi pernyataan,


(Putu Agung Ananta Wijaya)

PRAKATA

Puji dan syukur penulis panjatkan ke hadapan Tuhan Yang Maha Esa atas anugerah-Nya, sehingga tesis yang berjudul “RETRIEVAL SKRIPSI MENGGUNAKAN MODIFIKASI BM25 BERBASIS POSISI KATA ADAPTIF” dapat diselesaikan sesuai dengan yang direncanakan.

Tesis ini ditulis untuk memenuhi salah satu persyaratan untuk memperoleh gelar Magister Pendidikan Pascasarjana Universitas Pendidikan Ganesha pada Program Studi Ilmu Komputer. terselesaikannya tesis ini telah banyak memperoleh uluran tangan dari berbagai pihak. Untuk itu, ijin penulis menyampaikan terima kasih dan penghargaan kepada pihak berikut.

1. Bapak Dr. I Made Gede Sunarya, S.Kom., M.Cs. sebagai dosen pembimbing I yang dengan sabar membimbing, mengarahkan, dan memberikan motivasi yang demikian bermakna, sehingga penulis mampu melewati berbagai halangan dalam perjalanan studi dan penyelesaian tesis ini.
2. Bapak Dr. I Gede Aris Gunadi, S.Si. M.Kom. sebagai pembimbing II, yang telah memberikan saran, masukan, dan motivasi kepada penulis selama penelitian dan penulisan naskah laporan tesis ini, sehingga tesis ini dapat terwujud dengan baik sesuai harapan.
3. Bapak Dr. Putu Hendra Suputra, S.Kom., M.Cs. sebagai penguji I yang telah banyak memberikan masukan-masukan yang bermanfaat untuk penyempurnaan tesis ini.

4. Bapak Prof. Dr. Ir. Dewa Gede Hendra Divayana, S.Kom., M.Kom., MCE., IPU., ASEAN.Eng., APEC.Eng. sebagai penguji II yang telah banyak memberikan saran dan dorongan dalam penyelesaian tesis ini.
5. Koordinator Program Studi Magister Ilmu Komputer Universitas Pendidikan Ganesha dan staf dosen pengajar yang telah banyak membantu dan memotivasi penulis selama penyusunan tesis ini
6. Direktur Program Pascasarjana Universitas Pendidikan Ganesha, yang telah banyak memberikan pengarahan serta petunjuk hingga terselesaikannya tesis ini.
7. Rektor Universitas Pendidikan Ganesha, yang telah memberikan bantuan secara moril dan memfasilitasi berbagai kepentingan studi, selama Penulis menempuh pendidikan di Program Pascasarjana Universitas Pendidikan Ganesha.
8. Rekan-rekan seangkatan di Program Studi Magister Ilmu Komputer yang dengan karakternya masing-masing telah banyak berkontribusi membentuk kepribadian penulis selama menjalani studi dan penyelesaian tesis ini
9. Segenap keluarga tercinta, Bapak, Ibu, dan Saudara yang telah banyak membantu memberikan semangat dan motivasi serta dukungan moril dan materiil dalam penulisan tesis ini.

Semoga semua bantuan yang telah mereka berikan dalam menyelesaikan studi ini, mereka diberkati imbalan yang sepadan oleh Tuhan Yang Maha Esa, kesehatan, dan keharmonian dalam menjalani kehidupan. Penulis menyadari bahwa tesis ini belum sempurna. Namun, kehadirannya dalam konstelasi masyarakat akademis akan menambah perbendaharaan ilmu dalam perkembangan ilmu pengetahuan. Semoga tesis ini bermanfaat bagi masyarakat akademis, terutama mereka yang menyatakan diri bernaung di bawah kebesaran panji-panji pendidikan.



Singaraja, September 2024

Penulis

Daftar Isi

LEMBAR PERSETUJUAN PEMBIMBING_TESIS	iv
LEMBAR PERSETUJUAN TIM PENGUJI	v
Lembar Pernyataan.....	vi
PRAKATA	vii
ABSTRAK	x
ABSTRACT	xi
Daftar Isi.....	xii
Daftar Tabel	xv
Daftar Gambar.....	xvi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Identifikasi Masalah	4
1.3 Batasan Masalah.....	5
1.4 Rumusan Masalah	6
1.5 Tujuan Penelitian.....	6
1.6 Manfaat Penelitian.....	7
BAB II KAJIAN PUSTAKA	8
2.1 Kajian Hasil Penelitian Relevan.....	8
2.2 Kajian Teori.....	12
2.2.1 TF-IDF (<i>Term Frequency – Inverse Document Frequency</i>).....	12
2.2.2 BM25 (<i>Best Matching25</i>).....	13
2.2.3 Kueri Ekspansi	16
2.2.4 Word2Vec	19
2.3 <i>Library</i> Pengembangan Sistem	23
2.3.1 Sastrawi (<i>Stemmer</i> Bahasa Indonesia)	23
2.3.2 Django Framework.....	24
2.3.3 Genshim Word2Vec.....	25
2.4 Metode Usulan	26
2.4.1 Representasi Kueri dan Koefisien Bobot λ_i	27
2.4.2 Formula Dasar BM25 Modifikasi	28

2.5	Hipotesis Penelitian	31
BAB III METODELOGI PENELITIAN		33
3.1	Alur Penelitian	33
3.2	Perencanaan	34
3.3	Tahapan Pengumpulan Data	34
3.3.1	Teknik Pengumpulan Data	34
3.3.2	Jenis Data	35
3.4	Tahap Analisa dan Perancangan	36
3.4.1	Gambaran Pengujian Sistem	36
3.4.2	Perancangan Antarmuka Rekomendasi Skripsi	42
3.5	Alat Penelitian	43
3.5.1	<i>Hardware</i>	43
3.5.2	<i>Software</i>	44
3.6	Penyusunan Data <i>Ground Truth</i>	44
3.7	Pengujian Sistem	48
3.8	Jadwal Penelitian	50
BAB IV HASIL DAN PEMBAHASAN		51
4.1	Pengumpulan Data	51
4.1.1	Dokumen Skripsi	51
4.1.2	Pengumpulan Kueri	52
4.1.3	<i>Ground Truth</i>	53
4.2	<i>Preprocessing Text</i>	57
4.2.1	<i>Lowercasing</i> dan <i>Cleaning</i>	57
4.2.2	<i>Stopword Removal</i>	58
4.2.3	<i>Tokenizing</i>	59
4.2.4	<i>Stemming</i>	59
4.3	Skenario Pengujian	60
4.3.1	Metode Pencarian (<i>Retrieval Methods</i>)	60
4.3.2	Tipe Data Dokumen	61
4.3.3	Daftar Skenario Pengujian	62
4.4	Evaluasi dan Pembahasan Hasil	65
4.4.1	Evaluasi Metode	65

4.4.2	Evaluasi <i>Stemming</i>	70
4.4.3	Evaluasi Word2Vec	72
4.4.4	Evaluasi Per Sumber Data.....	76
4.5	Fitur Rekomendasi	77
BAB V_PENUTUP.....		81
5.1	Simpulan.....	81
5.2	Saran	83
DAFTAR PUSTAKA		86



Daftar Tabel

Tabel 3.1 Spesifikasi Hardware	43
Tabel 3.2 Spesifikasi Software.....	44
Tabel 3.3 Interpretasi Cohen's Kappa	46
Tabel 3.4 Jadwal Penelitian.....	50
Tabel 4.1 Daftar Skenario Pengujian	63
Tabel 4.2 Daftar F1-Score Tertinggi.....	67
Tabel 4.3 Daftar <i>MAP</i> Tertinggi	68
Tabel 4.4 Daftar Kueri Ekspansi.....	73



Daftar Gambar

Gambar 2.1 Pendekatan Ekspansi Kueri.....	18
Gambar 2.2 Arsitektur Word2Vec : Skip-Gram dan CBOW	21
Gambar 3.1 Alur Penelitian.....	33
Gambar 3.2 Flowchart Sistem Rekomendasi Skripsi.....	37
Gambar 3.3 Rancangan Antarmuka Rekomendasi Skripsi	43
Gambar 4.1 Data Tugas Akhir Skripsi.....	52
Gambar 4.2 Distribusi Data Tugas Akhir 2016-2024	52
Gambar 4.3 Kueri Uji.....	53
Gambar 4.4 Halaman Kuiser Dokumen Relevan	54
Gambar 4.5 Petunjuk Pengisian Kuiser	55
Gambar 4.6 Penilaian Relevansi Dokumen	56
Gambar 4.7 Hasil Proses Case Folding dan Cleaning.....	58
Gambar 4.8 Hasil Proses Stopword Removal	59
Gambar 4.9 Hasil Proses <i>Stemming</i>	60
Gambar 4.10 Grafik <i>MAP</i> Tiga Metode Data Abstrak <i>Stemming</i>	66
Gambar 4.11 Grafik <i>MAP@10 Stemming</i> Parameter.....	70
Gambar 4.12 Grafik <i>MAP@30 Stemming</i> Parameter.....	70
Gambar 4.13 Grafik <i>MAP@50 Stemming</i> Parameter.....	70
Gambar 4.14 Grafik <i>F1@10 Word2Vec</i> Parameter	72
Gambar 4.15 Grafik <i>F1@10 Word2Vec</i> Parameter	72
Gambar 4.16 Grafik <i>F1@50 Word2Vec</i> Parameter	72
Gambar 4.17 Grafik <i>MAP@10 Jenis Data</i> Parameter.....	76
Gambar 4.18 Grafik <i>MAP@30 Jenis Data</i> Parameter.....	76
Gambar 4.19 Grafik <i>MAP@50 Jenis Data</i> Parameter.....	76
Gambar 4.20 Flowchart Fitur Rekomendasi Skripsi.....	78

Daftar Lampiran

Lampiran 1. Identifikasi Masalah – Pencarian skripsi dengan kueri “sosial media” sistem perpustakaan Universitas Udayana (9 hasil pencarian)	97
Lampiran 2. Identifikasi Masalah – Pencarian skripsi dengan term yang mengandung “sosial media” pada database (299 hasil pencarian).....	97
Lampiran 3. Pembobotan term BM25 dan TF-IDF berdasarkan panjang dokumen	98

