Jurnal Ilmiah Sistem Informasi dan Ilmu Komputer
Volume 6, Nomor 3, November 2026
e-ISSN: 2827-7953; p-ISSN: 2827-8135, Hal. 01-16
DOI: https://doi.org/10.55606/juisik.v6i3.2614
Tersedia: http://journal.sinov.id/index.php/juisik

Analisis Pengaruh *Convex Combination* terhadap Performa *Hybrid* BM25 dan IndoSBERT pada Berita Berbahasa Indonesia

Irawati Kusumadewi[1*], Deni Arifianto[2], Nanda Kurnia Wardati[3]
[1-3]Program Studi Sistem Informasi, Universitas Muhammadiyah Jember, Indonesia
**Penulis Korespondensi: irrawkd12@gmail.com*

Abstract. *The increasing availability of digital news has made retrieving relevant information more challenging due to the rapid growth of online news content. BM25 provides effective lexical matching but is limited in capturing semantic similarity, whereas IndoSBERT better understands contextual semantics but is less effective for explicit lexical matching. This study investigates the effect of Convex Combination weighting on the performance of Hybrid Retrieval integrating BM25 and IndoSBERT using an Indonesian news corpus consisting of 83.967 articles derived from the Indonesian News Title Dataset and enriched through web scraping. The retrieval pipeline included text preprocessing, BM25 indexing, document chunking, IndoSBERT embedding generation, and score fusion using Convex Combination. Performance was evaluated using Precision@5, Precision@10, NDCG@5, NDCG@10, and latency based on an expert-judgment ground truth comprising 155 unique query-document pairs. Results show that Hybrid Retrieval consistently outperformed both BM25 and IndoSBERT. The α = 0,8 configuration achieved the best ranking quality, with NDCG@5 and NDCG@10 of 1,000, whereas α = 0,9 achieved the highest retrieval precision, with Precision@5 of 1.000 and Precision@10 of 0,975. These findings demonstrate that Convex Combination provides an effective and practical approach for developing Indonesian-language Information retrieval systems by balancing lexical and semantic retrieval while maintaining efficient response time.*

Keywords: *BM25; Convex Combination; Hybrid Retrieval; IndoSBERT; Information Retrieval.*

Abstrak. Pertumbuhan berita digital menyebabkan ketersediaan informasi semakin besar sehingga proses menemukan dokumen yang relevan menjadi semakin kompleks. BM25 memiliki keunggulan dalam pencocokan leksikal, tetapi terbatas dalam menangkap kesamaan semantik, sedangkan IndoSBERT lebih baik dalam memahami makna kontekstual, namun kurang optimal dalam mempertahankan kesesuaian leksikal. Penelitian ini menganalisis pengaruh pembobotan *Convex Combination* terhadap performa *Hybrid Retrieval* yang mengintegrasikan BM25 dan IndoSBERT menggunakan korpus berita berbahasa Indonesia yang terdiri atas 83.967 artikel yang disusun dari *Indonesian News Title Dataset* dan diperkaya melalui proses *web scraping*. Tahapan penelitian meliputi *text preprocessing*, pengindeksan BM25, *document chunking*, pembentukan *embedding* menggunakan IndoSBERT, serta penggabungan skor menggunakan *Convex Combination*. Evaluasi dilakukan menggunakan *Precision@5, Precision@10,* NDCG@5, NDCG@10, dan *latency* berdasarkan *ground truth* hasil *expert judgment* yang terdiri atas 155 pasangan kueri-dokumen unik. Hasil penelitian menunjukkan bahwa *Hybrid Retrieval* secara konsisten mengungguli BM25 dan IndoSBERT. Konfigurasi α=0,8 menghasilkan kualitas pemeringkatan terbaik dengan nilai NDCG@5 dan NDCG@10 masing-masing sebesar 1,000, sedangkan α=0,9 menghasilkan presisi pencarian tertinggi dengan nilai *Precision@5* sebesar 1,000 dan *Precision@10* sebesar 0,975. Temuan ini menunjukkan bahwa *Convex Combination* merupakan pendekatan yang efektif dan praktis untuk meningkatkan kinerja *information retrieval* berita berbahasa Indonesia melalui keseimbangan antara pencocokan leksikal dan semantik dengan tetap mempertahankan efisiensi waktu respons.

Kata Kunci: BM25; *Convex Combination*; *Hybrid Retrieval*; IndoSBERT; Temu Kembali Informasi.

1. LATAR BELAKANG

Perkembangan teknologi digital telah meningkatkan volume berita daring secara signifikan sehingga proses pencarian dokumen yang relevan menjadi semakin menantang (Pratiwi & Nugraha, 2022). Tantangan ini semakin besar pada bahasa dengan keterbatasan sumber daya (*low-resource languages*), termasuk bahasa Indonesia yang masih memiliki keterbatasan dataset, model, dan *benchmark* untuk mendukung pengembangan sistem pencarian informasi (Cahyawijaya et al., 2021). Oleh karena itu, *information retrieval* menjadi

Naskah Masuk: 20 Mei 2026; Revisi: 26 Juni 2026; Diterima: 30 Juli 2026; Terbit: 1 November 2026

komponen penting dalam membantu pengguna memperoleh informasi yang sesuai dengan kebutuhan pencarian (Sharma & Panda, 2023). Selain itu, tren penelitian menunjukkan bahwa *information retrieval* masih menjadi salah satu bidang yang terus berkembang untuk meningkatkan efektivitas sistem pencarian informasi (Hambarde & Proenca, 2023).

Salah satu metode *information retrieval* yang banyak digunakan adalah *Best Matching 25* (BM25). Metode ini efektif dalam mencocokkan istilah kueri dengan dokumen dan mampu memberikan performa yang baik pada pencarian berbasis kata kunci (Lasminingrum et al., 2026). Namun, BM25 memiliki keterbatasan dalam memahami hubungan semantik sehingga kurang mampu menemukan dokumen yang memiliki makna serupa tetapi menggunakan istilah berbeda (Sharma & Panda, 2023). Sebaliknya, model berbasis *Transformer* seperti IndoSBERT mampu merepresentasikan makna teks secara kontekstual sehingga lebih baik dalam menangkap kesamaan semantik, tetapi cenderung kurang efektif dalam mempertahankan kecocokan istilah eksplisit, terutama pada kueri pendek (Reimers & Gurevych, 2019; Ubaidillah et al., 2025). Berdasarkan penelitian Thamrin et al. (2024), model *denaya/IndoSBERT-large* menunjukkan performa yang lebih baik dibandingkan varian IndoSBERT lainnya pada tugas pemrosesan bahasa Indonesia, sehingga model tersebut dipilih sebagai representasi semantik dalam penelitian ini.

Perbedaan karakteristik tersebut menunjukkan bahwa pendekatan leksikal dan semantik saling melengkapi sehingga mendorong pengembangan *hybrid retrieval*. Salah satu teknik yang banyak digunakan adalah *Convex Combination*, yaitu metode *score fusion* yang menggabungkan skor BM25 dan model berbasis *Transformer* melalui parameter pembobotan tertentu. Wang et al. (2021) menunjukkan bahwa interpolasi antara BM25 dan *dense retriever* mampu meningkatkan kualitas pemeringkatan dibandingkan metode tunggal, sedangkan Bruch et al. (2024) membuktikan bahwa *Convex Combination* memberikan performa yang lebih baik dan lebih stabil dibandingkan pendekatan *rank-based fusion* pada berbagai domain. Selain itu, penelitian Kodri et al. (2025) dan Merchant et al. (2026) juga menunjukkan bahwa integrasi pendekatan leksikal dan semantik mampu meningkatkan efektivitas *information retrieval* pada domain hukum maupun data multilingual. Temuan-temuan tersebut menunjukkan bahwa penggabungan sinyal relevansi leksikal dan semantik merupakan pendekatan yang menjanjikan dalam meningkatkan kualitas hasil pencarian. Dalam penelitian ini, korpus berita berbahasa Indonesia dipilih karena memiliki keragaman topik, kosakata, dan gaya penulisan yang lebih tinggi dibandingkan domain spesifik, sehingga lebih representatif untuk mengevaluasi performa sistem *information retrieval* pada kondisi penggunaan nyata.

Meskipun demikian, penelitian terdahulu masih memiliki beberapa keterbatasan. Sebagian besar penelitian dilakukan pada domain hukum, dokumen akademik, atau data multilingual sehingga belum merepresentasikan karakteristik berita berbahasa Indonesia yang lebih heterogen. Selain itu, penelitian Lasminingrum et al. (2026) menunjukkan bahwa BM25 masih mengungguli IndoSBERT ketika digunakan secara tunggal pada dataset berbahasa Indonesia, namun belum mengimplementasikan *hybrid retrieval* untuk menggabungkan keunggulan kedua metode tersebut. Di sisi lain, kajian yang secara khusus mengintegrasikan BM25 dengan model bahasa Indonesia seperti IndoSBERT serta menganalisis pengaruh variasi pembobotan *Convex Combination* terhadap performa *Hybrid Retrieval* pada korpus berita berbahasa Indonesia masih relatif terbatas. Oleh karena itu, penelitian ini bertujuan menganalisis performa BM25, IndoSBERT, dan *Hybrid Retrieval*, sekaligus mengevaluasi pengaruh variasi pembobotan *Convex Combination* terhadap efektivitas sistem *information retrieval* menggunakan metrik *Precision@K*, NDCG@K, dan *latency*. Penelitian ini diharapkan dapat memberikan kontribusi sebagai referensi dalam pengembangan sistem *information retrieval* berita berbahasa Indonesia yang lebih akurat dan efisien melalui pemanfaatan pendekatan *hybrid retrieval.*

2. KAJIAN TEORITIS

Information Retrieval (IR) merupakan bidang yang berfokus pada proses pencarian dokumen yang relevan berdasarkan kebutuhan pengguna (Manning et al., 2008). Perkembangan IR menunjukkan pergeseran dari pendekatan berbasis pencocokan kata (*lexical retrieval*) menuju pendekatan berbasis representasi makna (*semantic retrieval)* untuk mengatasi *lexical gap* dan *semantic gap* (Hambarde & Proenca, 2023; Atoum, 2025). Oleh karena itu, penelitian terkini banyak mengembangkan *hybrid retrieval* yang mengintegrasikan kedua pendekatan guna meningkatkan efektivitas pencarian informasi (Wang et al., 2021).

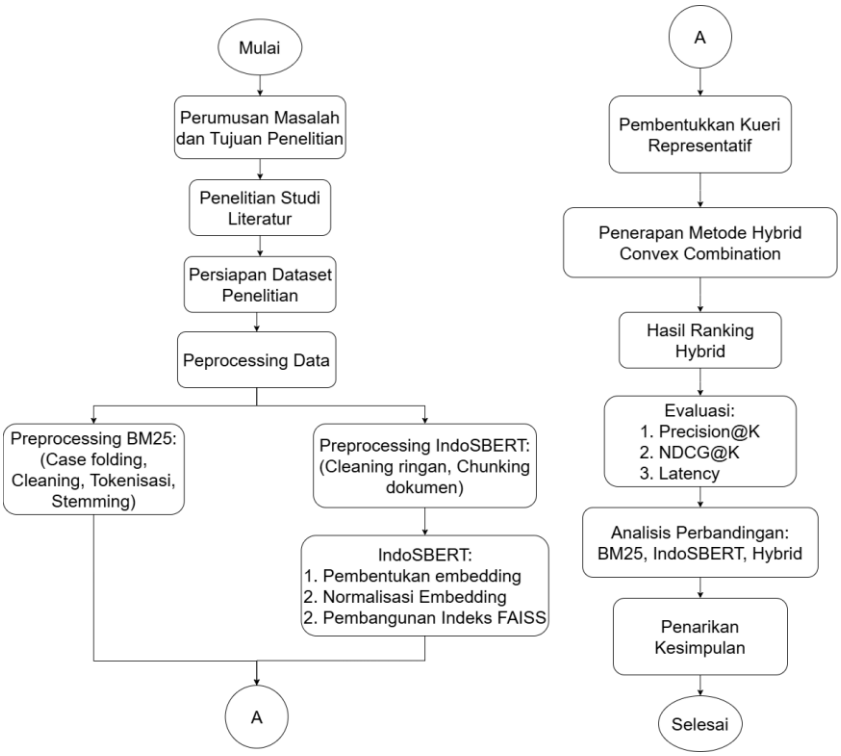
Pada pendekatan leksikal, BM25 merupakan metode pemeringkatan yang mengukur relevansi berdasarkan frekuensi istilah, panjang dokumen, dan tingkat keunikan istilah sehingga masih banyak digunakan sebagai *first-stage retrieval* pada sistem pencarian modern (Robertson & Zaragoza, 2009; Hambarde & Proenca, 2023). Sementara itu, pendekatan semantik memanfaatkan model berbasis *Transformer* seperti IndoSBERT untuk merepresentasikan teks ke dalam *dense embedding* sehingga mampu memahami hubungan kontekstual antar kata (Devlin et al., 2019; Reimers & Gurevych, 2019). Meskipun demikian, BM25 memiliki keterbatasan dalam memahami makna semantik, sedangkan IndoSBERT

kurang sensitif terhadap kecocokan istilah eksplisit, terutama pada kueri yang pendek (Lasminingrum et al., 2026).

Perbedaan karakteristik tersebut menjadikan *hybrid retrieval s*ebagai pendekatan yang mampu menggabungkan keunggulan BM25 dan IndoSBERT. Salah satu teknik *score fusion* yang banyak digunakan adalah *Convex Combination*, yaitu penggabungan skor relevansi kedua metode menggunakan koefisien pembobotan setelah proses normalisasi skor (Bruch et al., 2024). Penelitian sebelumnya menunjukkan bahwa pendekatan ini mampu meningkatkan efektivitas temu kembali informasi dibandingkan penggunaan metode tunggal, namun performanya dipengaruhi oleh nilai bobot yang digunakan (Wang et al., 2021; Bruch et al., 2024). Oleh karena itu, penelitian ini menganalisis pengaruh variasi pembobotan *Convex Combination* terhadap performa *hybrid retrieval* BM25 dan IndoSBERT pada koleksi berita berbahasa Indonesia.

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen untuk menganalisis pengaruh variasi bobot *Convex Combination* terhadap performa *Hybrid Retrieval* berbasis BM25 dan IndoSBERT pada korpus berita berbahasa Indonesia. Eksperimen dilakukan dengan membandingkan performa BM25, IndoSBERT, dan *Hybrid Retrieval* pada berbagai nilai pembobotan α. Alur penelitian secara keseluruhan ditunjukkan pada Gambar 1.



Gambar 1. Alur Penelitian.

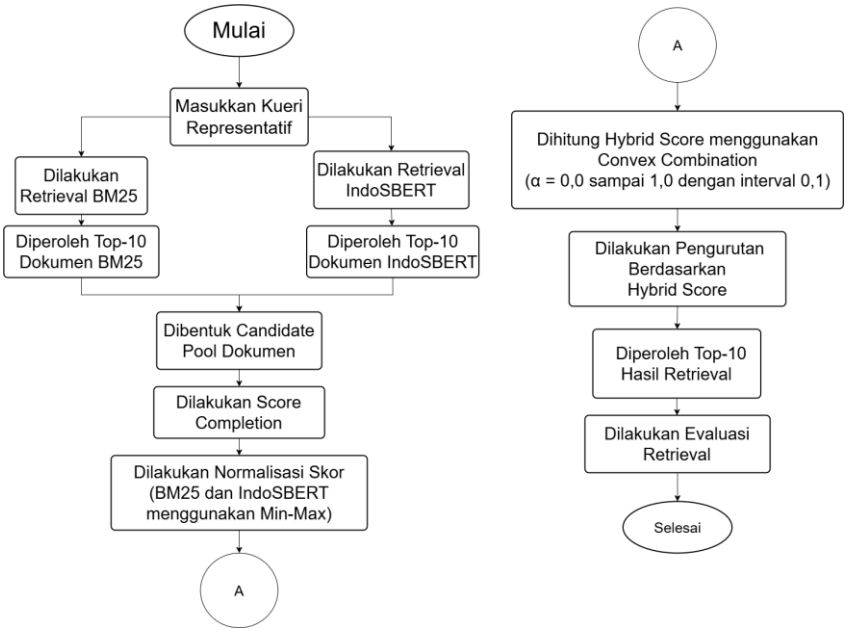
Dataset penelitian dibangun dari *Indonesian News Title Dataset* yang dipublikasikan melalui *Kaggle* oleh Ibrahim (2021), kemudian diperkaya melalui proses *web scraping* untuk memperoleh isi artikel secara lengkap. Dataset awal berisi metadata artikel berupa tanggal publikasi, URL, judul, dan kategori berita. Artikel yang digunakan dalam penelitian ini merupakan berita yang dipublikasikan pada periode 1 Januari 2020 hingga 13 Juni 2020, sedangkan proses *web scraping* menghasilkan korpus akhir sebanyak 83.967 artikel berita berbahasa Indonesia yang telah melalui proses pembersihan data, meliputi penghapusan *missing value, dokumen duplikat*, artikel dengan panjang konten kurang dari 100 karakter, serta elemen non-konten seperti teks navigasi dan iklan.

Tahap *preprocessing* disesuaikan dengan karakteristik masing-masing metode. Pada BM25 dilakukan *general cleaning, case folding, cleaning, tokenization, stopword removal*, dan *stemming* (Wijaya et al., 2021). Sementara itu, pada IndoSBERT dilakukan *general cleaning* dan *whitespace normalization* untuk mempertahankan struktur kalimat (Diana, 2023). Selanjutnya dokumen dibagi menjadi beberapa *chunk* menggunakan *tokenizer* bawaan model *denaya/IndoSBERT-large* dengan ukuran maksimum 256 token dan *overlap* 50 token. Setiap *chunk* kemudian direpresentasikan menjadi *embedding*, dinormalisasi menggunakan *L2 Normalization*, dan diindeks menggunakan FAISS *IndexFlatIP* untuk mendukung proses *semantic retrieval* (Douze et al., 2025).

Proses *Hybrid Retrieval* diawali dengan melakukan pencarian dokumen menggunakan BM25 dan IndoSBERT secara independen. Sepuluh dokumen teratas dari masing-masing metode digabungkan menjadi *candidate pool*, kemudian dilakukan *score completion* sehingga setiap dokumen memiliki skor BM25 dan IndoSBERT. Kedua skor dinormalisasi menggunakan *Min-Max Normalization* agar berada pada rentang yang sama sebelum digabungkan menggunakan metode *Convex Combination*. Perhitungan *hybrid score* mengacu pada Bruch et al. (2024) yang dirumuskan sebagai berikut:

$$Score_{hybrid}(d) = \alpha . S_{sem}(d) + (1 - \alpha) . S_{lex}(d) \quad \text{(i)}$$

Dengan $S_{sem}(d)$ menyatakan skor relevansi IndoSBERT, $S_{lex}(d)$ menyatakan skor relevansi BM25 yang telah dinormalisasi, sedangkan α merupakan koefisien pembobotan yang mengatur kontribusi kedua metode. Pada penelitian ini digunakan variasi nilai α dari 0,0 hingga 1,0 dengan *interval* 0,1 untuk mengevaluasi berbagai proporsi kontribusi skor semantik dan leksikal. Dokumen kemudian diurutkan berdasarkan nilai *hybrid score* sehingga diperoleh *Top-10* hasil pencarian pada setiap kombinasi bobot. Tahapan *Hybrid Retrieval* ditunjukkan pada Gambar 2.



Gambar 2. Tahapan *Hybrid Retrieval Menggunakan Convex Combination.*

Evaluasi dilakukan menggunakan *ground truth* yang dibentuk melalui *expert judgment* oleh satu orang ahli bahasa Indonesia terhadap 155 pasangan kueri-dokumen unik yang diperoleh dari hasil penggabungan dokumen kandidat BM25 dan IndoSBERT setelah penghapusan dokumen duplikat. Setiap pasangan kueri-dokumen diberi label relevan atau tidak relevan berdasarkan kesesuaian substansi dokumen terhadap topik kueri. Efektivitas sistem diukur menggunakan *Precision@5, Precision@10,* NDCG@5, dan NDCG@10, sedangkan efisiensi sistem diukur menggunakan *latency*.

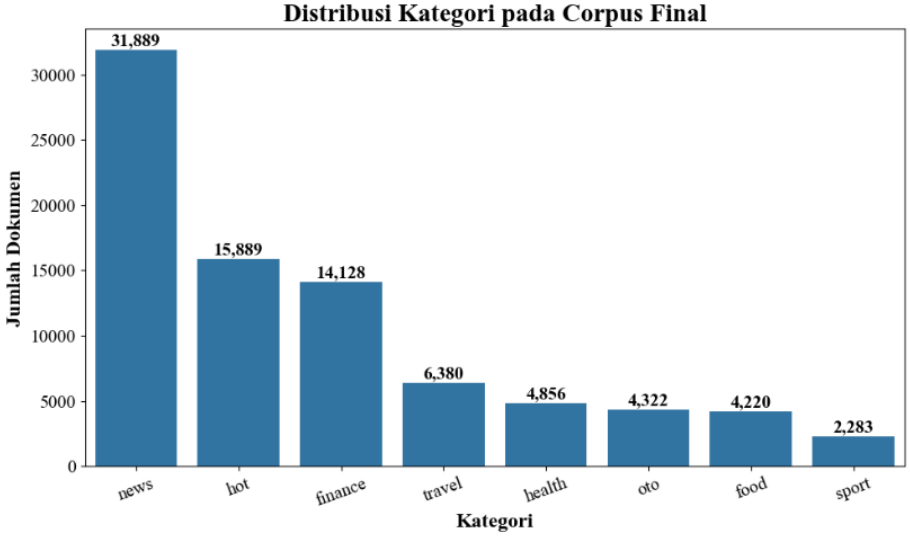
Seluruh eksperimen diimplementasikan menggunakan bahasa pemrograman *Python* dengan pustaka *SentenceTransformers, Transformers*, FAISS, *rank_bm25*, *Pandas, NumPy,* dan *Scikit-learn*. Eksperimen dijalankan pada komputer dengan prosesor *Intel Core i5 Gen-11*, RAM 16GB, GPU terintegrasi *Intel Iris Xe Graphics,* dan sistem operasi Windows, sehingga hasil pengukuran *latency* diperoleh pada konfiguasi perangkat tersebut. Hasil evaluasi digunakan untuk membandingkan performa BM25, IndoSBERT, dan *Hybrid Retrieval*, serta menganalisis pengaruh variasi nilai α terhadap kualitas temu kembali informasi.

4. HASIL DAN PEMBAHASAN

Persiapan Data Penelitian

Dataset penelitian dibangun dari *Indonesian News Title Dataset* yang diperkaya melalui proses *web scraping* untuk memperoleh isi lengkap artikel berita. Dataset awal terdiri atas 91.017 URL berita yang dipublikasikan pada periode 1 Januari 2020 hingga 13 Juni 2020. Proses *web scraping* berhasil mengekstraksi 85.231 artikel dengan tingkat keberhasilan

93,64%, kemudian dilakukan proses finalisasi melalui penghapusan *missing value*, artikel duplikat, artikel dengan panjang konten kurang dari 100 karakter, serta elemen non-konten. Hasil proses tersebut menghasilkan korpus akhir sebanyak 83.967 artikel yang terdiri atas delapan kategori berita. Distribusi Kategori pada Korpus Akhir disajikan pada Gambar 3.



Gambar 3. Distribusi Kategori pada Korpus Final.

Berdasarkan Gambar 3, distribusi dokumen antar kategori menunjukkan kondisi yang tidak seimbang, dengan kategori *news* memiliki jumlah artikel terbanyak, sedangkan kategori *sport* memiliki jumlah artikel paling sedikit. Distribusi tersebut mencerminkan karakteristik data berita asli sehingga tidak dilakukan proses penyeimbangan (*balancing*) agar representasi topik tetap sesuai dengan kondisi nyata. Korpus akhir selanjutnya digunakan sebagai sumber dokumen pada proses *retrieval*. *Preprocessing* dilakukan sesuai karakteristik masing-masing metode, yaitu BM25 menggunakan *case folding, cleaning, tokenization, stopword removal,* dan *stemming,* sedangkan IndoSBERT hanya menerapkan *general cleaning* untuk mempertahankan informasi kontekstual. Seluruh dokumen kemudian dibagi menjadi beberapa *chunk* berukuran 256 token dengan *overlap* 50 token sebelum direpresentasikan menjadi *embedding* menggunakan model *denaya/IndoSBERT-large*. Ringkasan representasi dokumen ditunjukkan pada Tabel 1.

Tabel 1. Ringkasan Representasi Dokumen.

No	Karakteristik	Nilai
1.	Jumlah Dokumen	83.967
2.	Jumlah Kategori	8
3.	Rata-rata Panjang Artikel	1.801,96 karakter
4.	Jumlah *Chunk*	170.328
5.	*Chunk Size*	256 token
6.	*Overlap*	50 token
7.	Rata-rata *Chunk* per Dokumen	2.03
8.	Jumlah *Embedding*	170.328
9.	Dimensi *Embedding*	256

Proses *chunking* menghasilkan 170.328 *chunk* atau rata-rata 2,03 *chunk* untuk setiap dokumen. Hasil ini menunjukkan bahwa sebagian besar artikel memiliki panjang melebihi batas maksimum token model IndoSBERT sehingga diperlukan pembagian dokumen menjadi beberapa *chunk* agar seluruh informasi tetap dapat direpresentasikan pada proses pembentukan *embedding*.

Kueri Evaluasi dan *Ground Truth*

Evaluasi sistem dilakukan menggunakan delapan kueri representatif yang masing-masing mewakili satu kategori berita pada korpus penelitian. Kueri dibentuk secara otomatis menggunakan pendekatan TF-IDF untuk mengidentifikasi istilah yang paling representatif pada setiap kategori. Selanjutnya dilakukan proses penyaringan menggunakan *cross-category analysis*, *automatic blacklist*, *metadata blacklist*, dan *generic terms blacklist* untuk menghilangkan istilah yang bersifat umum atau muncul pada banyak kategori. Dari hasil tersebut dipilih 20 istilah dengan bobot TF-IDF tertinggi pada setiap kategori yang kemudian digabungkan menjadi satu kueri representatif (Papagiannopoulou & Tsoumakas, 2019). Hasil pembentukan kueri representatif disajikan pada Tabel 2.

Tabel 2. Hasil Kueri Representatif.

No	Kategori	Kueri
1.	*Finance*	menteri usaha perintah kerja ekonomi uang jakarta turun harga negara covid dampak masyarakat dagang triliun pt bank naik tambah bumn
2.	*Food*	makan enak rasa masak daging saji menu ayam bahan goreng restoran kuliner santap buah manis nasi resep bumbu minum sayur
3.	*Health*	covid sakit pasien kasus dr infeksi sehat tubuh positif sembuh tular dokter gejala tingkat tangan teliti sebab obat sebar cegah
4.	*Hot*	aku film sang anak lagu Instagram diri tulis cerita suami rasa istri kabar soal nikah dunia pernah keluarga tampil kini
5.	*News*	covid jakarta warga kasus perintah polisi masyarakat positif kota pasien minta tangan kait tugas jumlah korban sehat dki pihak kerja
6.	*Oto*	mobil kendara motor mesin jual honda otomotif toyota listrik unit model produksi Jakarta harga pabrik juta kemudi merk mulai pasar
7.	*Sport*	juara balap motogp latih tanding atlet main gelar olimpiade menang musim tim olahraga ambulans kalah dunia rossi marquez tarung yamaha
8.	*Travel*	wisata tempat traveler wisatawan kota unjung terbang pariwisata indah turis pantai destinasi kunjung negara mulai tutup buka taman dunia tumpang

Berdasarkan Tabel 2, setiap kueri terdiri atas kumpulan istilah yang merepresentasikan karakteristik topik pada masing-masing kategori berita. Sebagai contoh, kategori *Finance* didominasi oleh istilah seperti menteri, usaha, ekonomi, bank, dan BUMN, sedangkan kategori *Sport* memuat istilah seperti juara, MotoGP, atlet, dan olimpiade. Perbedaan komposisi istilah

tersebut menunjukkan bahwa proses seleksi berhasil menghasilkan kueri yang memiliki daya pembeda antarkategori sehingga layak digunakan sebagai kueri evaluasi.

Ground truth diperoleh melalui proses *expert judgment* oleh satu orang ahli terhadap hasil retrieval BM25 dan IndoSBERT. Untuk setiap kueri diambil masing-masing 10 dokumen teratas (*Top-10*) dari kedua metode, kemudian seluruh pasangan kueri-dokumen dinilai sebagai relevan atau tidak relevan berdasarkan kesesuaian substansi dokumen terhadap topik kueri. Selanjutnya hasil penilaian dari kedua metode digabungkan dan dilakukan penghapusan dokumen duplikat sehingga diperoleh 155 pasangan kueri-dokumen unik yang digunakan sebagai *ground truth* pada proses evaluasi. Seluruh metode dievaluasi menggunakan *ground truth* yang sama sehingga perbandingan performa dilakukan berdasarkan acuan relevansi yang konsisten.

Hasil Evaluasi *Hybrid Retrieval*

Evaluasi Precision@K

Precision@K digunakan untuk mengukur proporsi dokumen relevan yang muncul pada K dokumen teratas hasil *retrieval*, sehingga metrik ini menggambarkan kemampuan sistem dalam menyajikan dokumen yang relevan pada posisi awal hasil pencarian (Manning et al., 2008). Semakin tinggi nilai *Precision@K*, semakin besar proporsi dokumen relevan yang berhasil ditempatkan pada peringkat teratas. Nilai *Precision* dihitung menggunakan Persamaan (2) sebagai berikut:

$$P@K = \frac{Jumlah\ dokumen\ relevan\ dalam\ k\ teratas}{k} \quad \text{..... (ii)}$$

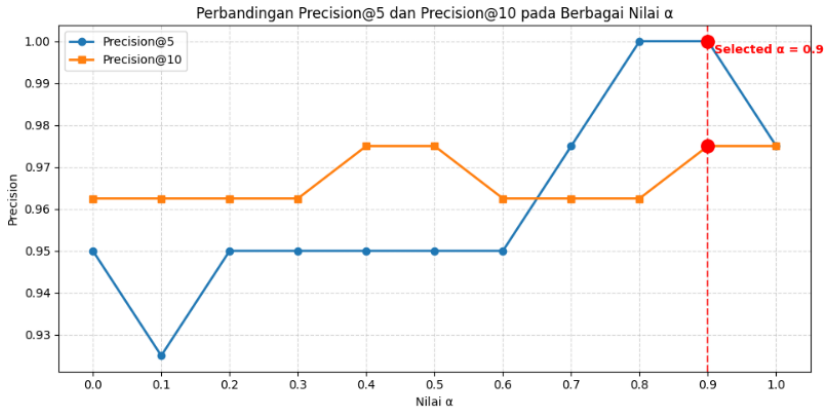
Hasil evaluasi *Precision@5* dan *Precision@10* pada setiap variasi nilai α disajikan pada Tabel3.

Tabel 3. Hasil Perhitungan *Precision@5* dan *Precision@10.*

No	*Alpha*	*Precision@5*	*Precision@10*
1.	0.0	0.950	0.9625
2.	0.1	0.925	0.9625
3.	0.2	0.950	0.9625
4.	0.3	0.950	0.9625
5.	0.4	0.950	0.9750
6.	0.5	0.950	0.9750
7.	0.6	0.950	0.9625
8.	0.7	0.975	0.9625
9.	0.8	1.000	0.9625
10.	0.9	1.000	0.9750
11.	1.0	0.975	0.9750

Berdasarkan Tabel 3, nilai *Precision@5* menunjukkan kecenderungan meningkat seiring bertambahnya nilai α hingga mencapai nilai maksimum sebesar 1,000 pada α=0,8 dan α=0,9. Sementara itu, *Precision@10* memperoleh nilai tertinggi sebesar 0,975 pada α=0,4,

α=0,5, α=0,9, α=1,0. Di antara kombinasi tersebut, α=0,9 dipilih sebagai konfigurasi terbaik berdasarkan metrik *Precision* karena mampu menghasilkan nilai maksimum secara bersamaan pada *Precision@5 dan Precision@10*.



Gambar 4. Grafik *Precision@5* dan *Precision@10* Terhadap Variasi Nilai α.

Berdasarkan Gambar 4, *Precision@5* mengalami peningkatan yang lebih signifikan dibandingkan *Precision@10* seiring meningkatnya nilai α. Hal ini menunjukkan bahwa semakin besar kontribusi skor semantik dari IndoSBERT, semakin baik sistem dalam menempatkan dokumen yang benar-benar relevan pada lima peringkat teratas. Representasi semantik memungkinkan sistem menemukan dokumen yang memiliki makna serupa meskipun tidak selalu menggunakan istilah yang sama dengan kueri, sedangkan BM25 tetap berperan dalam mempertahankan kecocokan kata kunci secara eksplisit. Oleh karena itu, kombinasi kedua metode melalui *Convex Combination* mampu memanfaatkan keunggulan pendekatan leksikal dan semantik secara bersamaan sehingga meningkatkan ketepatan hasil pencarian.

Temuan tersebut menunjukkan bahwa peningkatan bobot semantik memberikan pengaruh yang lebih besar terhadap kualitas hasil pada peringkat awal dibandingkan pada sepuluh dokumen teratas. Hasil ini juga konsisten dengan penelitian Wang et al. (2021) yang melaporkan bahwa penggabungan pendekatan *lexical retrieval* dan *dense retrieval* melalui interpolasi skor mampu meningkatkan *efektivitas* temu kembali informasi dibandingkan penggunaan masing-masing metode secara terpisah. Dengan demikian, hasil penelitian ini memperkuat bahwa keseimbangan antara sinyal leksikal dari BM25 dan representasi semantik dari IndoSBERT berkontribusi terhadap peningkatan kualitas *retrieval*, khususnya dalam menampilkan dokumen yang paling relevan pada posisi teratas.

Evaluasi NDCG@K

Normalized Discounted Cumulative Gain (NDCG) digunakan untuk mengevaluasi kualitas pemeringkatan dokumen dengan mempertimbangkan tingkat relevansi serta posisi dokumen pada hasil *retrieval*. Nilai NDCG berada pada rentang 0 hingga 1, di mana nilai yang

semakin mendekati 1 menunjukkan bahwa urutan dokumen yang dihasilkan semakin mendekati urutan relevansi ideal (Järvelin & Kekäläinen, 2002). Nilai NDCG dihitung menggunakan Persamaan (3) sebagai berikut:

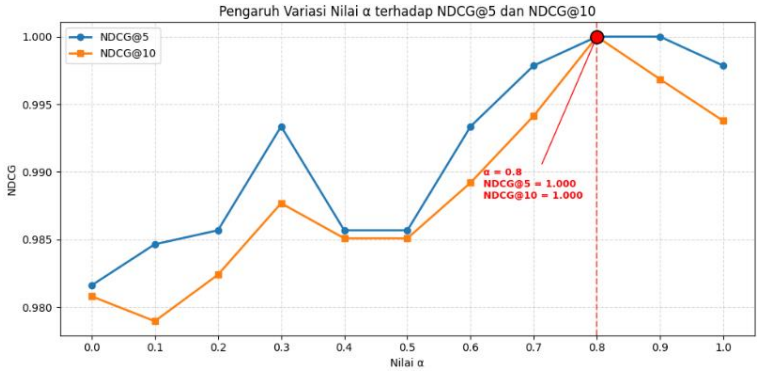
$$N\,DCG_k = \frac{DCG_k}{I\,DCG_k} \quad \text{(iii)}$$

Hasil evaluasi NDCG@5 dan NDCG@10 untuk setiap variasi nilai α disajikan pada Tabel 4.

Tabel 4. Hasil Evaluasi NDCG@5 dan NDCG@10.

No	*Alpha*	*ndcg@5*	*ndcg@10*
1.	0.0	0.981616	0.980803
2.	0.1	0.984652	0.978964
3.	0.2	0.985682	0.982399
4.	0.3	0.993363	0.987684
5.	0.4	0.985682	0.985092
6.	0.5	0.985682	0.985092
7.	0.6	0.993363	0.989205
8.	0.7	0.997862	0.994146
9.	0.8	1.000000	1.000000
10.	0.9	1.000000	0.996856
11.	1.0	0.997862	0.993778

Berdasarkan Tabel 4, nilai NDCG@5 dan NDCG@10 menunjukkan kecenderungan meningkat seiring bertambahnya nilai α hingga mencapai nilai maksimum pada α=0,8. Pada kombinasi tersebut, memiliki kesesuaian penuh dengan urutan relevansi berdasarkan *ground truth* hasil *expert judgment*. Nilai tersebut merupakan hasil rata-rata evaluasi terhadap delapan kueri representatif yang digunakan pada penelitian ini. Ketika nilai α ditingkatkan menjadi 0,9 dan 1,0, NDCG@10 mengalami sedikit penurunan menjadi 0,9969 dan 0,9938, sedangkan NDCG@5 tetap mencapai nilai maksimum pada α =0,9 sebelum kembali menurun pada α =1,0.



Gambar 5. Grafik NDCG@5 dan NDCG@10 terhadap Variasi Nilai α.

Berdasarkan Gambar 5, peningkatan nilai NDCG hingga α=0,8 menunjukkan bahwa penambahan kontribusi skor semantik dari IndoSBERT mampu menghasilkan urutan dokumen yang semakin mendekati urutan relevansi ideal. Namun, ketika bobot semantik ditingkatkan lebih lanjut, kualitas pemeringkatan tidak lagi mengalami peningkatan. Kondisi ini

menunjukkan bahwa dominasi skor semantik belum tentu menghasilkan urutan dokumen terbaik.

Hasil tersebut mengindikasikan bahwa konfigurasi α=0,8 memberikan keseimbangan yang optimal antara kontribusi BM25 dan IndoSBERT. Pada konfigurasi ini, representasi semantik dari IndoSBERT telah cukup dominan untuk memperbaiki urutan dokumen, sementara BM25 masih mampu mempertahankan kecocokan istilah yang penting pada proses *retrieval*. Oleh karena itu, kombinasi kedua metode menghasilkan kualitas pemeringkatan yang lebih baik dibandingkan ketika salah satu metode mendominasi proses pencarian.

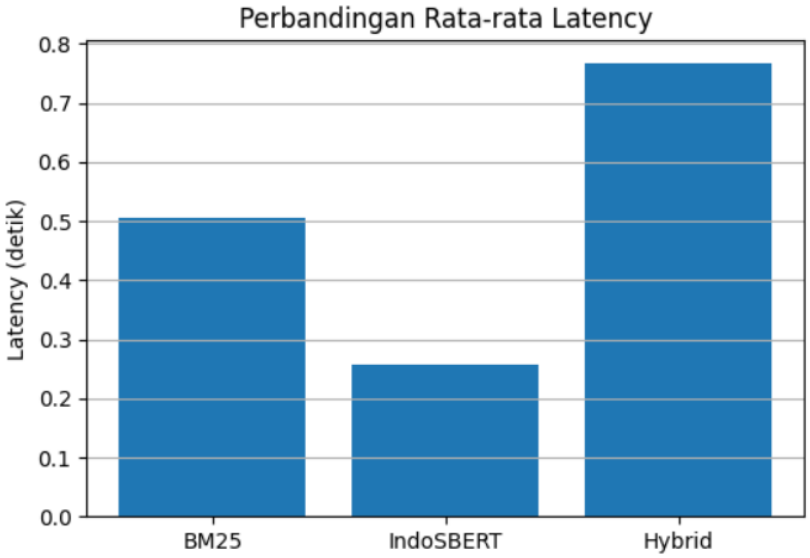
Temuan ini sejalan dengan penelitian Bruch et al. (2024) yang menunjukkan bahwa metode *score fusion* berbasis *Convex Combination* mampu menghasilkan kualitas pemeringkatan yang lebih stabil dibandingkan pendekatan *rank-based fusion*. Dengan demikian, hasil penelitian ini memperkuat bahwa keseimbangan kontribusi antara BM25 dan IndoSBERT melalui *Convex Combination* berperan penting dalam menghasilkan urutan dokumen yang semakin mendekati relevansi ideal.

Evaluasi Latency

Latency digunakan untuk mengukur efisiensi sistem, yaitu waktu yang dibutuhkan sejak kueri diproses hingga sistem menghasilkan daftar dokumen hasil *retrieval* (Gao et al., 2021). Metrik ini digunakan untuk menganalisis *trade-off* antara peningkatan kualitas *retrieval* dan waktu respons sistem.

Tabel 5. Ringkasan Rata-Rata *Latency Retrieval*.

No	Metode	*Latency* Rata-Rata (detik)
1.	BM25	0.5066
2.	IndoSBERT	0.2561
3.	*Hybrid*	0.7669



Gambar 6. Perbandingan Rata-Rata *Latency Retrieval*.

Berdasarkan Tabel 5 dan Gambar 6, IndoSBERT menghasilkan waktu *retrieval* tercepat dengan rata-rata *latency* sebesar 0,2561 detik, diikuti BM25 sebesar 0,5066 detik, sedangkan *Hybrid Retrieval* memerlukan waktu rata-rata 0,7669 detik. Rendahnya *latency* pada IndoSBERT dipengaruhi oleh penggunaan indeks FAISS sehingga proses pencarian cukup dilakukan melalui pembentukan embedding kueri dan pencarian *nearest neighbor*. Sebaliknya, *Hybrid Retrieval* memerlukan waktu lebih lama karena sistem harus menjalankan *retrieval* BM25 dan IndoSBERT sebelum melakukan penggabungan skor menggunakan *Convex Combination*. Meskipun memiliki *latency* tertinggi, *Hybrid Retrieval* hanya membutuhkan tambahan waktu sekitar 0,26 detik dibandingkan BM25 dan sekitar 0,51 detik dibandingkan IndoSBERT. Seluruh proses *retrieval* tetap dapat diselesaikan dalam waktu kurang dari satu detik sehingga tambahan waktu komputasi tersebut masih sebanding dengan peningkatan kualitas hasil pencarian yang diperoleh.

Perbandingan Performa BM25, IndoSBERT, dan Hybrid Retrieval

Perbandingan performa BM25, IndoSBERT, dan *Hybrid Retrieval* berdasarkan metrik *Precision@5, Precision@10,* NDCG@5, NDCG@10, dan *latency* disajikan pada Tabel 6. Pada *Hybrid Retrieval* digunakan dua konfigurasi terbaik, yaitu α=0,8 sebagai konfigurasi dengan nilai NDCG tertinggi dan α=0,9 sebagai konfigurasi dengan nilai *Precision* tertinggi.

Tabel 6. Perbandingan Performa BM25, IndoSBERT, dan *Hybrid Retrieval*.

No	Metode	*Precision@5*	*Precision@10*	*ndcg@5*	*ndcg@10*	*Latency*
1.	BM25	0.950	0.9625	0.9816	0.9808	0.5066
2.	*Hybrid* ($\alpha = 0.8$)	1.000	0.9625	1.0000	1.0000	0.7669
3.	*Hybrid* ($\alpha = 0.9$)	1.000	0.9750	1.0000	0.9969	0.7669
4.	IndoSBERT	0.975	0.9750	0.9979	0.9938	0.2561

Berdasarkan Tabel 6, BM25 menghasilkan performa yang baik dalam *retrieval* berbasis kecocokan istilah, sedangkan IndoSBERT memberikan kualitas *retrieval* yang lebih tinggi sekaligus waktu respons tercepat. Penggunaan representasi semantik memungkinkan IndoSBERT menemukan dokumen yang memiliki makna serupa meskipun menggunakan kosakata yang berbeda dari kueri. *Hybrid Retrieval* berhasil menggabungkan keunggulan kedua metode tersebut. Pada konfigurasi α=0,8, *Hybrid Retrieval* menghasilkan nilai NDCG@5 dan NDCG@10 sebesar 1,000, sedangkan pada konfigurasi α=0,9 diperoleh nilai *Precision@5* sebesar 1,000 dan *Precision@10* sebesar 0,975, yang merupakan performa terbaik berdasarkan metrik *Precision*. Hasil ini menunjukkan bahwa penggabungan sinyal leksikal dari BM25 dan representasi semantik dari IndoSBERT melalui *Convex Combination* mampu meningkatkan kualitas hasil *retrieval* dibandingkan penggunaan masing-masing metode secara individual.

Meskipun *Hybrid Retrieval* memiliki *latency* lebih tinggi dibandingkan BM25 maupun IndoSBERT, waktu respons yang dihasilkan masih berada di bawah satu detik. Dengan demikian, peningkatan efektivitas *retrieval* yang diperoleh masih sebanding dengan tambahan waktu komputasi, sehingga metode yang diusulkan berpotensi diterapkan pada sistem pencarian berita berbahasa Indonesia yang membutuhkan keseimbangan antara kualitas hasil pencarian dan efisiensi waktu respons. Temuan ini sejalan dengan Bruch et al. (2024) yang menyatakan bahwa pendekatan *hybrid* memanfaatkan sifat komplementer antara *retrieval* leksikal dan semantik, sehingga penggabungan keduanya melalui *Convex Combination* mampu meningkatkan kualitas retrieval. Selain itu, hasil penelitian ini juga konsisten dengan Kodri et al. (2025) yang menunjukkan bahwa integrasi BM25 dan Sentence-BERT mampu meningkatkan performa *retrieval* dibandingkan penggunaan metode tunggal.

5. KESIMPULAN DAN SARAN

Penelitian ini menganalisis pengaruh variasi nilai pembobotan pada metode *Convex Combination* terhadap performa *Hybrid Retrieval* berbasis BM25 dan IndoSBERT pada korpus berita berbahasa Indonesia. Hasil evaluasi menunjukkan bahwa perubahan nilai pembobotan α memengaruhi kualitas *retrieval* yang dihasilkan. Semakin besar kontribusi skor semantik dari IndoSBERT, kualitas *retrieval* cenderung meningkat hingga mencapai konfigurasi optimal, namun peningkatan bobot semantik secara berlebihan tidak selalu menghasilkan performa terbaik.

Berdasarkan metrik evaluasi, konfigurasi α=0,8 menghasilkan kualitas pemeringkatan terbaik dengan nilai NDCG@5 dan NDCG@10 sebesar 1,000, sedangkan konfigurasi α=0,9 memberikan nilai *Precision@5* sebesar 1,000 dan *Precision@10* sebesar 0,975. Hasil tersebut menunjukkan bahwa kombinasi bobot yang memberikan keseimbangan antara sinyal leksikal dari BM25 dan representasi semantik dari IndoSBERT mampu menghasilkan performa *retrieval* yang lebih baik dibandingkan ketika salah satu metode mendominasi proses penggabungan skor. Meskipun *Hybrid Retrieval* memiliki *latency* yang lebih tinggi dibandingkan metode tunggal, waktu respons yang dihasilkan tetap berada di bawah satu detik sehingga peningkatan kualitas *retrieval* masih sebanding dengan tambahan waktu komputasi.

Penelitian ini masih memiliki keterbatasan pada jumlah kueri evaluasi dan *ground truth* yang dibangun melalui *expert judgment* terhadap sejumlah dokumen terbatas. Oleh karena itu, penelitian selanjutnya disarankan menggunakan dataset yang memiliki pasangan kueri-dokumen relevan (*query-document relevance*) dalam jumlah yang lebih besar agar evaluasi

dapat dilakukan secara lebih komprehensif. Selain itu, metode *score fusion* lain maupun metrik evaluasi tambahan, seperti *Recall*, *Mean Reciprocal Rank* (MRR), dan *Mean Average Precision* (MAP), dapat dieksplorasi untuk memperoleh gambaran performa sistem yang lebih menyeluruh.

DAFTAR REFERENSI

Atoum, I. (2025). Evolving information retrieval: From traditional models to emerging paradigms. *Journal of Advances in Information Technology, 16*(9), 1277-1294. https://doi.org/10.12720/jait.16.9.1277-1294

Bruch, S., Gai, S., & Ingber, A. (2024). An analysis of fusion functions for hybrid retrieval. *ACM Transactions on Information Systems, 42*(1), Article 20, 1-35. https://doi.org/10.1145/3596512

Cahyawijaya, S., Winata, G. I., Wilie, B., Vincentio, K., Li, X., Kuncoro, A., Ruder, S., Lim, Z. Y., Bahar, S., Khodra, M. L., Purwarianti, A., & Fung, P. (2021). IndoNLG: Benchmark and resources for evaluating Indonesian natural language generation. *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 8875-8898. https://doi.org/10.18653/v1/2021.emnlp-main.699

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 4171-4186. https://doi.org/10.18653/v1/N19-1423

Diana, D. (2023). *Indonesian SBERT for semantic textual similarity task*. Hugging Face. https://huggingface.co/denaya/indoSBERT-large

Douze, M., Guzhva, A., Deng, C., Johnson, J., Szilvasy, G., Mazaré, P.-E., Lomeli, M., Hosseini, L., & Jégou, H. (2025). *The Faiss library*. *arXiv*. https://arxiv.org/abs/2401.08281

Gao, L., Dai, Z., Chen, T., Fan, Z., Van Durme, B., & Callan, J. (2021). *Complementing lexical retrieval with semantic residual embeddings*. *arXiv*. https://arxiv.org/abs/2004.13969

Hambarde, K. A., & Proença, H. (2023). Information retrieval: Recent advances and beyond. *IEEE Access, 11*, 76581-76604. https://doi.org/10.1109/ACCESS.2023.3295776

Ibrahim, I. (2021). *Indonesian news title dataset*. Kaggle. https://www.kaggle.com/datasets/ibamibrahim/indonesian-news-title

Järvelin, K., & Kekäläinen, J. (2002). Cumulated gain-based evaluation of IR techniques. *ACM Transactions on Information Systems, 20*(4), 422-446. https://doi.org/10.1145/582415.582418

Kodri, W. A. G., Haris, M., & Fitriadi, R. (2025). Fine-Hybrid: Integration of BM25 and fine-tuned SBERT to enhance search relevance. *TEKNIKA, 14*(2), 213-222. https://doi.org/10.34148/teknika.v14i2.1229

Lasminingrum, D. P., Puspaningrum, E. Y., & Mulyo, B. M. (2026). Perbandingan metode lexical dan semantic retrieval: BM25, IndoSBERT baseline, dan IndoSBERT fine-tuned pada pencarian dokumen berbahasa Indonesia. *Prosiding Seminar Nasional Penelitian dan Pengabdian kepada Masyarakat (LPPM Universitas 'Aisyiyah Yogyakarta)*, *4*, 245-253. https://repository.upnjatim.ac.id/48873/

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. Cambridge University Press.

Merchant, B., Khazi, A., & Sonawane, S. S. (2026). Reciprocal rank fusion based hybrid dense-sparse information retrieval on code-mixed Banglish social media text. *CEUR Workshop Proceedings* (FIRE 2025: Forum for Information Retrieval Evaluation).

Papagiannopoulou, E., & Tsoumakas, G. (2019). A review of keyphrase extraction. *WIREs Data Mining and Knowledge Discovery*. https://arxiv.org/abs/1905.05044

Pratiwi, I. Y. R., & Nugraha, A. F. (2022). Hoax news identification using machine learning model from online media in Bahasa Indonesia. *MATRIX: Jurnal Manajemen Teknologi dan Informatika, 12*(2), 58-67. https://doi.org/10.31940/matrix.v12i2.58-67

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 3982-3992. https://doi.org/10.18653/v1/D19-1410

Robertson, S., & Zaragoza, H. (2009). The probabilistic relevance framework: BM25 and beyond. *Foundations and Trends in Information Retrieval, 3*(4), 333-389. https://doi.org/10.1561/1500000019

Sharma, S., & Panda, S. P. (2023). Efficient information retrieval model: Overcoming challenges in search engines—An overview. *Indonesian Journal of Electrical Engineering and Computer Science, 32*(2), 925-932. https://doi.org/10.11591/ijeecs.v32.i2.pp925-932

Thamrin, H., Oktafiani, D., Rasyid, I. I., & Fauzi, I. M. (2024). Classification of SWOT statements employing BERT pre-trained model embedding. *Jurnal Sistem Informasi Bisnis, 14*(2), 143-152. https://doi.org/10.21456/vol14iss2pp143-152

Ubaidillah, M., Adikara, P. P., & Perdana, R. S. (2025). Pemerolehan informasi terjemahan Al-Quran berbahasa Indonesia menggunakan IndoSBERT. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (J-PTIIK), 9*(8). http://j-ptiik.ub.ac.id

Wang, S., Zhuang, S., & Zuccon, G. (2021). BERT-based dense retrievers require interpolation with BM25 for effective passage retrieval. *Proceedings of the 2021 ACM SIGIR International Conference on Theory of Information Retrieval*, 317-324. https://doi.org/10.1145/3471158.3472233

Wijaya, I. N. S. W., Seputra, K. A., & Parwita, W. G. S. (2021). Comparison of the BM25 and Rabin-Karp algorithm for plagiarism detection. *Journal of Physics: Conference Series, 1810*(1), Article 012032. https://doi.org/10.1088/1742-6596/1810/1/012032