BIBLIOGRAPHY

[1] N. Matondang, "Sistem pencarian hasil penelitian skripsi menggunakan weighted tree similarity dan content based filtering skripsi," 2024.

[2] I. G. Nyoman, A. Jayarana, I. G. W. Darma, I. W. A. Juliantara, I. M. Agus, and W. Putra, "STUDY LITERATUR INFORMATION RETRIEVAL MODEL : TEKNIK DAN APLIKASI," vol. 03, no. 01, pp. 61–69, 2025.

[3] R. P. R. Tutuk Widowati, Zulmy Faqihuddin Putera, Nurul Shofiah, "Kemampuan Mengolah Bahan Bacaan pada Penyusunan Proposal Skripsi Mahasiswa," *Konstr. J. Pendidik. dan Pembelajaran*, vol. 15, no. 1, pp. 92–103, 2023, doi: 10.35457/konstruk.v15i1.2611.

[4] M. E. Siregar *et al.*, "Pengembangan Sistem Pencarian Resep Makanan dengan Implementasi Text Preprocessing dan BM25," vol. 8, no. 3, pp. 1458–1465, 2025.

[5] L. Heryawan, D. Novitaningrum, K. Rizqi, and S. Nurulfarah, "Medical Record Document Search with TF-IDF and Vector Space Model (VSM)," vol. 14, no. 3, pp. 847–852, 2024.

[6] D. A. Suryaningrum, R. Syaifudin, and H. R. P. Putra, "Integrasi Word Embeddings Dan Inverse Book Frequency Dalam Pembobotan Term Untuk Peningkatan Pencarian Dokumen," *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 9, no. 4, pp. 2529–2537, 2024, doi: 10.29100/jipi.v9i4.7557.

[7] M. P. Syah, A. P. Wardani, and M. Idhom, "Perbandingan Representasi Teks Tf-Idf Dan Bert Terhadap Akurasi," vol. 5, no. 1, pp. 47–59, 2025.

[8] D. Chandrasekaran and V. Mago, "Evolution of Semantic Similarity-A Survey," *ACM Comput. Surv.*, vol. 54, no. 2, pp. 1–35, 2022, doi: 10.1145/3440755.

[9] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, 2019. [Online]. Available: https://arxiv.org/abs/1908.10084

[10] S. Naseri, J. Dalton, A. Yates, and J. Allan, "CEQE: Contextualized Embeddings for Query Expansion," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 12656 LNCS, pp. 467–482, 2021, doi: 10.1007/978-3-030-72113-8_31.

[11] I. A. Mannix and E. Yulianti, "Academic expert finding using BERT pre-trained language model," *Int. J. Adv. Intell. Informatics*, vol. 10, no. 2, pp. 280–295, 2024, doi: 10.26555/ijain.v10i2.1497.

[12] A. Krishna, Y. Jin, C. Foster, G. Gabel, B. Hanley, and A. Youssef, "Query Expansion for Patent Searching using Word Embedding and Professional Crowdsourcing," *Present. AAAI FSS-19 Artif. Intell. Gov. Public Sect. Arlington, Virginia, USA*, 2019.

[13] T. Goyal, M. S. Bhadola, and M. K. Bhatia, "Automatic Query Expansion Using Word Embedding Based on Fuzzy Graph Connectivity Measures," *Int. J. Trend Sci. Res. Dev.*, vol. 5, no. 5, 2021.

[14] B. T. Ariesta, A. Romadhony, and Hasmawati, "Ekspansi Query Menggunakan Word2Vec pada Pencarian Artikel Ilmiah," *e-Proceeding*

Eng., vol. 10, no. 5, p. 4910, 2023.
[15] H. Viltres-sala and V. Estrada-sentí, "Information Retrieval Model with Query Expansion and User Preference Profile," *Rev. Fac. Ing.*, vol. 32, no. 64, pp. 0–3, 2023.
[16] R. Padaki, Z. Dai, and J. Callan, *Rethinking Query Expansion for BERT Reranking*. Springer International Publishing, 2020. doi: 10.1007/978-3-030-45442-5.
[17] K. Diana and M. L. Khodra, *IndoSBERT: Enhancing Indonesian Sentence Embeddings with Siamese Networks Fine-tuning*. 2023. doi: 10.1109/ICAICTA59291.2023.10390469.
[18] H. W. Rashid and S. H. Ahmed, "Fine-tuning SBERT for Semantic Research Title Classification in Trilingual University Repository," *Kurdistan J. Appl. Res.*, vol. 10, no. 2, pp. 119–135, 2025, doi: 10.24017/science.2025.2.9.
[19] R. Kaban, P. Sihombing, M. Pandia, and P. Simamora, "Pemrosesan Query dan Pemeringkatan Hasil dalam Information Retrieval: Sebuah Kajian Literatur," *J. Inf. Syst. Res.*, vol. 4, no. 3, pp. 748–754, 2023, doi: 10.47065/josh.v4i3.2867.
[20] G. Hirst, J. Lin, R. Nogueira, and A. Yates, "Pretrained Transformers for Text Ranking: BERT and beyond," *Synth. Lect. Hum. Lang. Technol.*, vol. 14, no. 4, pp. 1–325, 2021, doi: 10.2200/S01123ED1V01Y202108HLT053.
[21] M. Trabelsi, Z. Chen, B. D. Davison, and J. Heflin, "Neural ranking models for document retrieval," *Inf. Retr. J.*, vol. 24, no. 6, pp. 400–444, 2021, doi: 10.1007/s10791-021-09398-0.
[22] K. A. Hambarde and H. Proenca, "Information Retrieval: Recent Advances and beyond," *IEEE Access*, vol. 11, pp. 76581–76604, 2023, doi: 10.1109/ACCESS.2023.3295776.
[23] S. Bruch, F. M. Nardini, C. Rulli, and R. Venturini, *Efficient Inverted Indexes for Approximate Retrieval over Learned Sparse Representations*, vol. 1, no. 1. Association for Computing Machinery, 2024. doi: 10.1145/3626772.3657769.
[24] B. Biswas and R. Ramnath, "Efficient and Interpretable Information Retrieval for Product Question Answering with Heterogeneous Data," *7th Work. e-Commerce NLP, ECNLP 2024 Lr. 2024 - Work. Proc.*, no. Ecnlp 7, pp. 19–28, 2024.
[25] P. Christen, D. J. Hand, and N. Kirielle, "A Review of the F-Measure: Its History, Properties, Criticism, and Alternatives," *ACM Comput. Surv.*, vol. 56, no. 3, 2024, doi: 10.1145/3606367.
[26] P. Fränti and R. Mariescu-Istodor, "Soft precision and recall," *Pattern Recognit. Lett.*, vol. 167, pp. 115–121, 2023, doi: 10.1016/j.patrec.2023.02.005.
[27] O. Jeunen, I. Potapov, and A. Ustimenko, "On (Normalised) Discounted Cumulative Gain as an Off-Policy Evaluation Metric for Top-n Recommendation," *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, pp. 1222–1233, 2024, doi: 10.1145/3637528.3671687.
[28] N. Craswell, B. Mitra, E. Yilmaz, D. Campos, E. M. Voorhees, and I. Soboroff, "TREC Deep Learning Track: Reusable Test Collections in the Large Data Regime," *SIGIR 2021 - Proc. 44th Int. ACM SIGIR Conf. Res. Dev. Inf. Retr.*, no. i, pp. 2369–2375, 2021, doi: 10.1145/3404835.3463249.

[29] Q. Ai *et al.*, "Information Retrieval meets Large Language Models: A strategic report from Chinese IR community," *AI Open*, vol. 4, pp. 80–90, 2023, doi: 10.1016/j.aiopen.2023.08.001.
[30] Y. Zhu *et al.*, "Large Language Models for Information Retrieval: A Survey," pp. 1–43, 2024, [Online]. Available: http://arxiv.org/abs/2308.07107
[31] Y. Faqih, Y. Rahmanto, A. Ari Aldino, and B. Waluyo, "Penerapan String Matching Menggunakan Algoritma Boyer-Moore Pada Pengembangan Sistem Pencarian Buku Online," *Bull. Comput. Sci. Res.*, vol. 2, no. 3, pp. 100–106, 2022, doi: 10.47065/bulletincsr.v2i3.172.
[32] Ira Zulfa, Eliyin Eliyin, Rayuwati Rayuwati, and Riski Wanda, "Data Search System for Thesis and Internship Reports in the Library of the Faculty of Engineering, Gajah Putih University Takengon," *Int. J. Econ. Commer. Manag.*, vol. 2, no. 1, pp. 234–250, 2024, doi: 10.62951/ijecm.v2i1.403.
[33] D. Khurana, A. Koli, K. Khatter, and S. Singh, "Natural language processing: state of the art, current trends and challenges," *Multimed. Tools Appl.*, vol. 82, no. 3, pp. 3713–3744, 2023, doi: 10.1007/s11042-022-13428-4.
[34] B. Min *et al.*, "Recent Advances in Natural Language Processing via Large Pre-trained Language Models: A Survey," *ACM Comput. Surv.*, vol. 56, no. 2, Sep. 2023, doi: 10.1145/3605943.
[35] A. R. Lubis, Y. Y. Lase, D. A. Rahman, and D. Witarsyah, "Improving Spell Checker Performance for Bahasa Indonesia Using Text Preprocessing Techniques with Deep Learning Models," *Ing. des Syst. d'Information*, vol. 28, no. 5, pp. 1335–1342, 2023, doi: 10.18280/isi.280522.
[36] R. B. Hadiprakoso, H. Setiawan, R. N. Yasa, and Girinoto, "Text Preprocessing for Optimal Accuracy in Indonesian Sentiment Analysis Using a Deep Learning Model with Word Embedding," *AIP Conf. Proc.*, vol. 2680, no. 1, 2023, doi: 10.1063/5.0126116.
[37] M. A. Rosid, A. S. Fitrani, I. R. I. Astutik, N. I. Mulloh, and H. A. Gozali, "Improving Text Preprocessing for Student Complaint Document Classification Using Sastrawi," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 874, no. 1, 2020, doi: 10.1088/1757-899X/874/1/012017.
[38] M. U. Albab, Y. K. P., and M. N. Fawaiq, "Optimization of the Stemming Technique on Text Preprocessing President 3 Periods Topic," *J. Transform.*, vol. 20, no. 2, pp. 1–12, 2023, doi: 10.26623/transformatika.v20i2.5374.
[39] Rizal Chandra Rivaldi and T.D. Wismarini, "Analisis Sentimen Pada Ulasan Produk Dengan Metode Natural Language Processing (NLP)," *Elkom J. Elektron. dan Komput.*, vol. 17, no. 1, pp. 120–128, 2024, doi: 10.51903/elkom.v17i1.1680.
[40] Y. Zhang, Z. Lin, C. C. chiu Tong, and S. W. yeung Ho, "Enhancing tokenization accuracy with dynamic patterns: cumulative logic for segmenting user-generated content in logographic languages," *J. Comput. Soc. Sci.*, vol. 8, no. 3, pp. 1–24, 2025, doi: 10.1007/s42001-025-00406-7.
[41] R. Rinandyaswara, Y. A. Sari, and M. T. Furqon, "Pembentukan Daftar Stopword Menggunakan Term Based Random Sampling Pada Analisis Sentimen Dengan Metode Naïve Bayes (Studi Kasus: Kuliah Daring Di Masa Pandemi)," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 9, no. 4, p. 717, 2022, doi: 10.25126/jtiik.2022934707.

[42] S. Sarica and J. Luo, “Stopwords in technical language processing,” *PLoS One*, vol. 16, no. 8 August, pp. 1–13, 2021, doi: 10.1371/journal.pone.0254937.
[43] J. Ji and G. Yang, “Do Stop Words Matter in Bug Report Analysis? Empirical Findings Using Deep Learning Models Across Duplicate, Severity, and Priority Classification,” *Appl. Sci.*, vol. 15, no. 16, 2025, doi: 10.3390/app15169178.
[44] C. H. Yutika, A. Adiwijaya, and S. Al Faraby, “Analisis Sentimen Berbasis Aspek pada Review Female Daily Menggunakan TF-IDF dan Naïve Bayes,” *J. Media Inform. Budidarma*, vol. 5, no. 2, p. 422, 2021, doi: 10.30865/mib.v5i2.2845.
[45] M. Alyousf and M. F. Alhalabi, “A Survey of Document Stemming Algorithms in Information Retrieval Systems,” *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 24, no. 4, pp. 1–28, 2025, doi: 10.1145/3715120.
[46] M. Ali, A. Baqir, H. H. Raza Sherazi, S. Khalid, P. Smith, and M. Lee, “An Extended Pattern Based Comprehensive Stemmer for the Urdu Language,” *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 23, no. 12, 2024, doi: 10.1145/3701231.
[47] D. D. Jasman Pardede, “Perbandingan Algoritma Stemming Porter, Sastrawi, Idris, Dan Arifin & Setiono Pada Dokumen Teks Bahasa Indonesia,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 12, no. 1, pp. 69–76, 2025, doi: 10.25126/jtiik.2025128860.
[48] R. Ulgasesa, A. B. P. Negara, and T. Tursina, “Pengaruh Stemming Terhadap Performa Klasifikasi Sentimen Masyarakat Tentang Kebijakan New Normal,” *J. Sist. dan Teknol. Inf.*, vol. 10, no. 3, p. 286, 2022, doi: 10.26418/justin.v10i3.53880.
[49] G. Göksel, A. Arslan, and B. T. Dincer, “A selective approach to stemming for minimizing the risk of failure in information retrieval systems,” *PeerJ Comput. Sci.*, vol. 9, pp. 1–37, 2023, doi: 10.7717/peerj-cs.1175.
[50] J. S. McCarley, “Should we Translate the Documents or the Queries in Cross-language Information Retrieval?,” *IBM T.J. Watson Res. Cent.*, vol. 218, 1999.
[51] D. W. Oard, “A Comparative Study of Query and Document Translation for Cross-Language Information Retrieval,” in *Machine Translation and the Information Soup*, D. Farwell, L. Gerber, and E. Hovy, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 1998, pp. 472–483.
[52] V. Petras, A. Lüschow, R. Ramthun, J. Stiller, C. España-Bonet, and S. Henning, “Query or Document Translation for Academic Search -- What’s the Real Difference?,” in *Experimental IR Meets Multilinguality, Multimodality, and Interaction*, A. Arampatzis, E. Kanoulas, T. Tsikrika, S. Vrochidis, H. Joho, C. Lioma, C. Eickhoff, A. Névéol, L. Cappellato, and N. Ferro, Eds., Cham: Springer International Publishing, 2020, pp. 28–42.
[53] A. Pak, A. Ziyaden, T. Saparov, I. Akhmetov, and A. Gelbukh, “Word Embeddings: A Comprehensive Survey,” *Comput. y Sist.*, vol. 28, no. 4, pp. 2005–2029, 2024, doi: 10.13053/CyS-28-4-5225.
[54] E. Sezerer and S. Tekir, “A Survey On Neural Word Embeddings,” 2021, [Online]. Available: http://arxiv.org/abs/2110.01804

[55] A. Kashyap, T.-T. Nguyen, V. Schlegel, S. Winkler, S.-K. Ng, and S. Poria, “Beyond Words: A Comprehensive Survey of Sentence Representations.” 2023. doi: 10.48550/arXiv.2305.12641.

[56] T. Gao, X. Yao, and D. Chen, “SimCSE.DanqiChen.Princeton.21: Simple Contrastive Learning of Sentence Embeddings,” pp. 6894–6910, 2021, [Online]. Available: https://github.com/princeton-nlp/SimCSE.

[57] K. F. Agiharta, B. R. Suteja, and M. Ayub, “Penerapan Sentence BERT Untuk Similaritas Kompetensi Pekerjaan dan Mata Kuliah,” *J. Tek. Inform. dan Sist. Inf.*, vol. 10, p. 449, 2024, [Online]. Available: http://dx.doi.org/10.28932/jutisi.vXiX.X

[58] W. X. Zhao, J. Liu, R. Ren, and J. R. Wen, “Dense Text Retrieval Based on Pretrained Language Models: A Survey,” *ACM Trans. Inf. Syst.*, vol. 42, no. 4, pp. 1–41, 2024, doi: 10.1145/3637870.

[59] J. Devlin, M.-W. Chang, K. Lee, K. T. Google, and A. I. Language, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” *Naacl-Hlt 2019*, no. Mlm, pp. 4171–4186, 2018, [Online]. Available: https://aclanthology.org/N19-1423.pdf

[60] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, “IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP,” *COLING 2020 - 28th Int. Conf. Comput. Linguist. Proc. Conf.*, pp. 757–770, 2020, doi: 10.18653/v1/2020.coling-main.66.

[61] T. Pratama and Suharjito, “IndoXLNet: Pre-Trained Language Model for Bahasa Indonesia,” *Int. J. Eng. Trends Technol.*, vol. 70, no. 5, pp. 366–380, 2022, doi: 10.14445/22315381/IJETT-V70I5P240.

[62] V. Karpukhin *et al.*, “Dense passage retrieval for open-domain question answering,” *EMNLP 2020 - 2020 Conf. Empir. Methods Nat. Lang. Process. Proc. Conf.*, pp. 6769–6781, 2020, doi: 10.18653/v1/2020.emnlp-main.550.

[63] H. Zhang *et al.*, *Efficient and Effective Retrieval of Dense-Sparse Hybrid Vectors using Graph-based Approximate Nearest Neighbor Search*, vol. 1, no. 1. Association for Computing Machinery, 2024. [Online]. Available: http://arxiv.org/abs/2410.20381

[64] A. W. Anggoro, P. Corcoran, D. De Widt, and Y. Li, “Harmonized system code classification using supervised contrastive learning with sentence BERT and multiple negative ranking loss,” *Data Technol. Appl.*, vol. 59, no. 2, pp. 276–301, 2024, doi: 10.1108/DTA-01-2024-0052.

[65] Y. Sisodia, “Semantic Textual Similarity on Contracts : Exploring Multiple Negative Ranking Losses for Sentence Transformers,” pp. 1–8, 2023.

[66] S. Regularization, J. Chang, and K. Grover, “" That was smooth ": Exploration of S-BERT with Multiple Negatives Ranking Loss and,” *Stanford CS224N Nat. Lang. Process. with Deep Learn.*, 2024, [Online]. Available: https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1244/final-projects/JohnnyChangKanuGroverKaushalAtulAlate.pdf

[67] S. Jawale, S. Nehete, H. Patil, S. Pathak, P. Sapale, and S. Zite, “Cosine Similarity: a Key Driver for Enhanced Recommendation Systems,” *Int. Res. J. Mod. Eng. Technol. Sci.*, no. 04, pp. 1466–1470, 2024, [Online]. Available: www.irjmets.com

[68] Supiyanto and Sriyono, “Metode Cosine Similarity Untuk Mendeteksi

Kemiripan Pada Dokumen Teks," *SAINS J. MIPA dan Pengajarannya*, vol. 1, no. 1, pp. 1–7, 2023, [Online]. Available: https://ejournal.uncen.ac.id/index.php/SAINS

[69] Z. Xu *et al.*, "A Survey of Model Architectures in Information Retrieval," pp. 1–50, 2025, [Online]. Available: http://arxiv.org/abs/2502.14822

[70] R. Jagerman, H. Zhuang, Z. Qin, X. Wang, and M. Bendersky, *Query Expansion by Prompting Large Language Models*, vol. 1, no. 1. Association for Computing Machinery, 2023. [Online]. Available: http://arxiv.org/abs/2305.03653

[71] M. Dwi Laxmi and M. Ali Fauzi, "Query Expansion Pada Sistem Temu Kembali Informasi Berbahasa Indonesia Dengan Metode Pembobotan TF-IDF Dan Algoritme Cosine Similarity Berbasis Wordnet," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 1, pp. 2548–964, 2019, [Online]. Available: http://j-ptiik.ub.ac.id

[72] A. Allahim, A. Cherif, and A. Imine, "Semantic approaches for query expansion: taxonomy, challenges, and future research directions," *PeerJ Comput. Sci.*, vol. 11, pp. 1–53, 2025, doi: 10.7717/peerj-cs.2664.

[73] E. Bassani, N. Tonellotto, and G. Pasi, "Personalized Query Expansion with Contextual Word Embeddings," *ACM Trans. Inf. Syst.*, vol. 42, no. 2, 2023, doi: 10.1145/3624988.

[74] L. Wang, N. Yang, and F. Wei, "Query2doc: Query Expansion with Large Language Models," *EMNLP 2023 - 2023 Conf. Empir. Methods Nat. Lang. Process. Proc.*, pp. 9414–9423, 2023, doi: 10.18653/v1/2023.emnlp-main.585.

[75] M. Gospodinov, S. MacAvaney, and C. Macdonald, "Doc2Query–: When Less is More," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 13981 LNCS, pp. 414–422, 2023, doi: 10.1007/978-3-031-28238-6_31.

[76] E. Tanuwijaya, S. Adam, M. F. Anggris, and A. Z. Arifin, "Query expansion menggunakan word embedding dan pseudo relevance feedback," *Regist. J. Ilm. Teknol. Sist. Inf.*, vol. 5, no. 1, pp. 47–54, 2019, doi: 10.26594/register.v5i1.1385.

[77] M. Douze *et al.*, "The Faiss library," 2025, [Online]. Available: http://arxiv.org/abs/2401.08281

[78] J. Johnson, M. Douze, and H. Jegou, "Billion-Scale Similarity Search with GPUs," *IEEE Trans. Big Data*, vol. 7, no. 3, pp. 535–547, 2021, doi: 10.1109/TBDATA.2019.2921572.

[79] T. I. Ramadhan, A. Supriatman, and T. R. Kurniawan, "Passage Retrieval untuk Question Answering Bahasa Indonesia Menggunakan BERT dan FAISS," *J. Algoritm.*, vol. 21, no. 2, pp. 156–163, 2024, doi: 10.33364/algoritma/v.21-2.2100.

[80] R. C. Maringka and R. J. N. Makarawung, "OPTIMALISASI ANALISIS UJARAN KEBENCIAN ULASAN E-COMMERCE Abstraksi Pendahuluan Tinjauan Pustaka," vol. 7, no. 1, 2025.

[81] A. A. Nur Hakim, A. C. Murti, and R. Nindyasari, "Implementasi Artificial Intelligence Dalam Sistem Pencarian Orang Hilang Dengan Face Recognition Studi Kasus Polres Kudus," *SKANIKA Sist. Komput. dan Tek. Inform.*, vol. 8, no. 1, pp. 168–180, 2025, doi: 10.36080/skanika.v8i1.3334.

[82] A. Jadon and A. Patil, “A Comprehensive Survey of Evaluation,” pp. 1–25, 2024.
[83] I. Ruthven, “Re-examining the Potential Effectiveness of Interactive Query Expansion,” 2003.
[84] N. Craswell, B. Mitra, E. Yilmaz, D. Campos, and E. M. Voorhees, “OVERVIEW OF THE TREC 2019 DEEP LEARNING TRACK,” pp. 1–22, 2019.
[85] O. Alonso and S. Mizzaro, “Using crowdsourcing for TREC relevance assessment,” *Inf. Process. Manag.*, vol. 48, no. 6, pp. 1053–1066, 2012, doi: https://doi.org/10.1016/j.ipm.2012.01.004.