UNIVERSITAS
GADJAH MADA
Pengembangan Model Passage Reranking Berbasis BERT Cross Encoder untuk Bahasa Indonesia
Sigit Hanafi, Prof. Drs. Edi Winarko, M.Sc., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

DAFTAR PUSTAKA

Askari, A., Abolghasemi, A., Pasi, G., Kraaij, W., and Verberne, S. (2023). Injecting the bm25 score as text improves bert-based re-rankers. In *Advances in Information Retrieval: 45th European Conference on Information Retrieval, ECIR 2023, Dublin, Ireland, April 2–6, 2023, Proceedings, Part I*, page 66–83, Berlin, Heidelberg. Springer-Verlag.

Bajaj, P., Campos, D., Craswell, N., Deng, L., Gao, J., Liu, X., Majumder, R., McNamara, A., Mitra, B., Nguyen, T., Rosenberg, M., Song, X., Stoica, A., Tiwary, S., and Wang, T. (2018). Ms marco: A human generated machine reading comprehension dataset.

Bonifacio, L., Jeronymo, V., Abonizio, H. Q., Campiotti, I., Fadaee, M., Lotufo, R., and Nogueira, R. (2022). mmarco: A multilingual version of the ms marco passage ranking dataset.

Bowman, S. R., Angeli, G., Potts, C., and Manning, C. D. (2015). A large annotated corpus for learning natural language inference. In Màrquez, L., Callison-Burch, C., and Su, J., editors, *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 632–642, Lisbon, Portugal. Association for Computational Linguistics.

Busolin, F., Lucchese, C., Nardini, F. M., Orlando, S., Perego, R., Trani, S., and Veneri, A. (2025). Efficient re-ranking with cross-encoders via early exit. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '25, page 2534–2544, New York, NY, USA. Association for Computing Machinery.

Craswell, N., Mitra, B., Yilmaz, E., Campos, D., Voorhees, E. M., and Soboroff, I. (2021). Trec deep learning track: Reusable test collections in the large data regime. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '21, page 2369–2375, New York, NY, USA. Association for Computing Machinery.

Dai, L., Liu, H., and Xiong, H. (2024). Improve dense passage retrieval with entailment tuning. In Al-Onaizan, Y., Bansal, M., and Chen, Y.-N., editors, *Proceedings of the 2024 Conference on Empirical Methods in Natural Language*

UNIVERSITAS
GADJAH MADA
Pengembangan Model Passage Reranking Berbasis BERT Cross Encoder untuk Bahasa Indonesia
Sigit Hanafi, Prof. Drs. Edi Winarko, M.Sc., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

Processing, pages 11375–11387, Miami, Florida, USA. Association for Computational Linguistics.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In Burstein, J., Doran, C., and Solorio, T., editors, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Déjean, H., Clinchant, S., and Formal, T. (2024). A thorough comparison of cross-encoders and llms for reranking splade. *arXiv preprint arXiv:2403.10407*.

Gao, L., Dai, Z., and Callan, J. (2021a). Rethink training of bert rerankers in multi-stage retrieval pipeline. In *Advances in Information Retrieval: 43rd European Conference on IR Research, ECIR 2021, Virtual Event, March 28 – April 1, 2021, Proceedings, Part II*, page 280–286, Berlin, Heidelberg. Springer-Verlag.

Gao, L., Dai, Z., Chen, T., Fan, Z., Van Durme, B., and Callan, J. (2021b). Complement lexical retrieval model with semantic residual embeddings. In *Advances in Information Retrieval: 43rd European Conference on IR Research, ECIR 2021, Virtual Event, March 28 – April 1, 2021, Proceedings, Part I*, page 146–160, Berlin, Heidelberg. Springer-Verlag.

Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., Guo, Q., Wang, M., and Wang, H. (2023). Retrieval-augmented generation for large language models: A survey. *arXiv preprint arXiv:2312.10997*.

Geigle, G., Pfeiffer, J., Reimers, N., Vulić, I., and Gurevych, I. (2022). Retrieve fast, rerank smart: Cooperative and joint approaches for improved cross-modal retrieval. *Transactions of the Association for Computational Linguistics*, 10:503–521.

Gupta, S., Ranjan, R., and Singh, S. N. (2024). A comprehensive survey of retrieval-augmented generation (rag): Evolution, current landscape and future directions. *arXiv preprint arXiv:2410.12837*.

Han, S., Wang, X., Bendersky, M., and Najork, M. (2020). Learning-to-rank with bert in tf-ranking. *arXiv preprint arXiv:2004.08476*.

UNIVERSITAS
GADJAH MADA
Pengembangan Model Passage Reranking Berbasis BERT Cross Encoder untuk Bahasa Indonesia
Sigit Hanafi, Prof. Drs. Edi Winarko, M.Sc., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

Karpukhin, V., Oguz, B., Min, S., Lewis, P., Wu, L., Edunov, S., Chen, D., and Yih, W.-t. (2020). Dense passage retrieval for open-domain question answering. In Webber, B., Cohn, T., He, Y., and Liu, Y., editors, *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 6769–6781, Online. Association for Computational Linguistics.

Khattab, O. and Zaharia, M. (2020). Colbert: Efficient and effective passage search via contextualized late interaction over bert. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '20, page 39–48, New York, NY, USA. Association for Computing Machinery.

Lin, J., Alfonso-Hermelo, D., Jeronymo, V., Kamalloo, E., Lassance, C., Nogueira, R., Ogundepo, O., Rezagholizadeh, M., Thakur, N., Yang, J.-H., and Zhang, X. (2023). Simple yet effective neural ranking and reranking baselines for cross-lingual information retrieval.

Lin, J., Nogueira, R., and Yates, A. (2021). *Pretrained Transformers for Text Ranking: BERT and Beyond*. Springer, Cham.

Lu, J., Hall, K., Ma, J., and Ni, J. (2024). HYRR: Hybrid infused reranking for passage retrieval. In Calzolari, N., Kan, M.-Y., Hoste, V., Lenci, A., Sakti, S., and Xue, N., editors, *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 8528–8534, Torino, Italia. ELRA and ICCL.

Manning, C. D., Raghavan, P., and Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press, Cambridge, UK.

Nogueira, R. and Cho, K. (2019). Passage re-ranking with bert. *arXiv preprint arXiv:1901.04085*.

Nogueira, R., Jiang, Z., Pradeep, R., and Lin, J. (2020). Document ranking with a pretrained sequence-to-sequence model. In Cohn, T., He, Y., and Liu, Y., editors, *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 708–718, Online. Association for Computational Linguistics.

Pradeep, R., Nogueira, R., and Lin, J. (2021). The expando-mono-duo design pattern for text ranking with pretrained sequence-to-sequence models.

UNIVERSITAS
GADJAH MADA
Pengembangan Model Passage Reranking Berbasis BERT Cross Encoder untuk Bahasa Indonesia
Sigit Hanafi, Prof. Drs. Edi Winarko, M.Sc., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

Putra, I. M. S., Siahaan, D., and Saikhu, A. (2024). Snli indo: A recognizing textual entailment dataset in indonesian derived from the stanford natural language inference dataset. *Data in Brief*, 52:109998.

Qu, Y., Ding, Y., Liu, J., Liu, K., Ren, R., Zhao, W. X., Dong, D., Wu, H., and Wang, H. (2021). RocketQA: An optimized training approach to dense passage retrieval for open-domain question answering. In Toutanova, K., Rumshisky, A., Zettlemoyer, L., Hakkani-Tur, D., Beltagy, I., Bethard, S., Cotterell, R., Chakraborty, T., and Zhou, Y., editors, *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 5835–5847, Online. Association for Computational Linguistics.

Reimers, N. and Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In Inui, K., Jiang, J., Ng, V., and Wan, X., editors, *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3982–3992, Hong Kong, China. Association for Computational Linguistics.

Robertson, S. and Zaragoza, H. (2009). The probabilistic relevance framework: Bm25 and beyond. *Foundations and Trends in Information Retrieval*, 3(4):333–389.

Santhanam, K., Khattab, O., Potts, C., and Zaharia, M. (2022a). Plaid: An efficient engine for late interaction retrieval. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, CIKM '22, page 1747–1756, New York, NY, USA. Association for Computing Machinery.

Santhanam, K., Khattab, O., Saad-Falcon, J., Potts, C., and Zaharia, M. (2022b). ColBERTv2: Effective and efficient retrieval via lightweight late interaction. In Carpuat, M., de Marneffe, M.-C., and Meza Ruiz, I. V., editors, *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3715–3734, Seattle, United States. Association for Computational Linguistics.

Song, J., Jin, C., Zhao, W., McCallum, A., and Lee, J.-Y. (2024). Comparing neighbors together makes it easy: Jointly comparing multiple candidates for

UNIVERSITAS
GADJAH MADA
Pengembangan Model Passage Reranking Berbasis BERT Cross Encoder untuk Bahasa Indonesia
Sigit Hanafi, Prof. Drs. Edi Winarko, M.Sc., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

efficient and effective retrieval. In Al-Onaizan, Y., Bansal, M., and Chen, Y.-N., editors, *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 22255–22269, Miami, Florida, USA. Association for Computational Linguistics.

Wang, L., Yang, N., Huang, X., Jiao, B., Yang, L., Jiang, D., Majumder, R., and Wei, F. (2022). Text embeddings by weakly-supervised contrastive pre-training. *arXiv preprint arXiv:2212.03533*.

Warner, B., Chaffin, A., Clavié, B., Weller, O., Hallström, O., Taghadouini, S., Gallagher, A., Biswas, R., Ladhak, F., Aarsen, T., Adams, G. T., Howard, J., and Poli, I. (2025). Smarter, better, faster, longer: A modern bidirectional encoder for fast, memory efficient, and long context finetuning and inference. In Che, W., Nabende, J., Shutova, E., and Pilehvar, M. T., editors, *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2526–2547, Vienna, Austria. Association for Computational Linguistics.

Wilie, B., Vincentio, K., Winata, G. I., Cahyawijaya, S., Li, X., Lim, Z. Y., Soleman, S., Mahendra, R., Fung, P., Bahar, S., and Purwarianti, A. (2020). IndoNLU: Benchmark and resources for evaluating Indonesian natural language understanding. In Wong, K.-F., Knight, K., and Wu, H., editors, *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, pages 843–857, Suzhou, China. Association for Computational Linguistics.

Winata, G. I., Cahyawijaya, S., Liu, Z., Lin, Z., Madotto, A., and Fung, P. (2021). Are multilingual models effective in code-switching? In Solorio, T., Chen, S., Black, A. W., Diab, M., Sitaram, S., Soto, V., Yilmaz, E., and Srinivasan, A., editors, *Proceedings of the Fifth Workshop on Computational Approaches to Linguistic Code-Switching*, pages 142–153, Online. Association for Computational Linguistics.

Xiong, L., Xiong, C., Li, Y., Tang, K.-F., Liu, J., Bennett, P., Ahmed, J., and Overwijk, A. (2020). Approximate nearest neighbor negative contrastive learning for dense text retrieval.

UNIVERSITAS
GADJAH MADA
Pengembangan Model Passage Reranking Berbasis BERT Cross Encoder untuk Bahasa Indonesia
Sigit Hanafi, Prof. Drs. Edi Winarko, M.Sc., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

Zhan, J., Mao, J., Liu, Y., Guo, J., Zhang, M., and Ma, S. (2021). Optimizing dense retrieval model training with hard negatives. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '21, page 1503–1512, New York, NY, USA. Association for Computing Machinery.

Zhuang, H., Qin, Z., Jagerman, R., Hui, K., Ma, J., Lu, J., Ni, J., Wang, X., and Bendersky, M. (2023). Rankt5: Fine-tuning t5 for text ranking with ranking losses. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '23, page 2308–2313, New York, NY, USA. Association for Computing Machinery.