

DAFTAR PUSTAKA

- Abdallah, A., Kasem, M. S., Abdalla, M., Mahmoud, M., Elkasaby, M., Elbendary, Y. dan Jatowt, A. (2024), 'Arabicaqa: A comprehensive dataset for arabic question answering', *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval* .
<https://api.semanticscholar.org/CorpusID:268691723>
- AbūSulaymān, dan DeLorenzo, Y. T. (1993), *Crisis in the muslim mind*.
<https://api.semanticscholar.org/CorpusID:142357545>
- Ahadi, Ridho; Harahap, N. S. F. M. K. F. (2025), 'Retrieval-augmented generation in a web-based question answering system for fiqh books', *Politeknik Negeri Lhokseumawe Vol 5, No 2 (2025): June*.
<https://e-jurnal.pnl.ac.id/JAISE/article/view/7005/4963>
- Al-Qaraāwī, Y. (1990), 'Kaifa nata 'āmal ma'a al-sunnah al-nabawiyah', *Ma'ālim wa āawābit, USA: al-Ma'had al-Alamī li al-Fikr al-Islāmī* .
- Antoun, W., Baly, F. dan Hajj, H. (2020), AraBERT: Transformer-based model for Arabic language understanding, in H. Al-Khalifa, W. Magdy, K. Darwish, T. El-sayed dan H. Mubarak, eds, 'Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection', European Language Resource Association, Marseille, France, pp. 9–15.
<https://aclanthology.org/2020.osact-1.2/>
- Azzopardi, L., de Rijke, M. dan Balog, K. (2007), Building simulated queries for known-item topics: an analysis using six european languages, in 'Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval', SIGIR '07, Association for Computing Machinery, New York, NY, USA, p. 455–462.
<https://doi.org/10.1145/1277741.1277820>
- Bonifacio, L. H., Campiotti, I., de Alencar Lotufo, R. dan Nogueira, R. (2021), mmarco: A multilingual version of the ms marco passage ranking dataset.
<https://api.semanticscholar.org/CorpusID:237364084>

Campos, D. F., Nguyen, T., Rosenberg, M., Song, X., Gao, J., Tiwary, S., Majumder, R., Deng, L. dan Mitra, B. (2016), 'Ms marco: A human generated machine reading comprehension dataset', *ArXiv* **abs/1611.09268**.

<https://api.semanticscholar.org/CorpusID:1289517>

Chen, Z. dan Eickhoff, C. (2021), 'The cross-lingual arabic information retrieval (claire) system', *ArXiv* **abs/2107.13751**.

<https://api.semanticscholar.org/CorpusID:236493337>

Chiu, Y.-T. dan Bai, Z.-H. (2025), 'Translation or multilingual retrieval? evaluating cross-lingual search strategies for traditional chinese financial documents', *FinTech in AI CUP Special Session*.

<https://research.nii.ac.jp/ntcir/workshop/>

[OnlineProceedings18/pdf/ai_cup/03-AICUP-AICUP-ChiuY.pdf](https://research.nii.ac.jp/ntcir/workshop/OnlineProceedings18/pdf/ai_cup/03-AICUP-AICUP-ChiuY.pdf)

Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F. P., Grave, E., Ott, M., Zettlemoyer, L. dan Stoyanov, V. (2019), 'Unsupervised cross-lingual representation learning at scale', *ArXiv* **abs/1911.02116**.

<https://api.semanticscholar.org/CorpusID:207880568>

Cormack, G. V., Clarke, C. L. A. dan Büttcher, S. (2009), 'Reciprocal rank fusion outperforms condorcet and individual rank learning methods', *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*.

<https://api.semanticscholar.org/CorpusID:12408211>

Darwish, K. dan Magdy, W. (2014), 'Arabic information retrieval', *Foundations and Trends® in Information Retrieval* **7(4)**, 239–342.

<http://dx.doi.org/10.1561/15000000031>

Day, R. dan Bamford, J. (1998), *Extensive Reading in the Second Language Classroom*, Cambridge Language Education, Cambridge University Press.

<https://books.google.co.id/books?id=qo91VUx1TroC>

Devlin, J., Chang, M.-W., Lee, K. dan Toutanova, K. (2019), BERT: Pre-training of deep bidirectional transformers for language understanding, in J. Burstein, C. Doran dan T. Solorio, eds, '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)', Association for Computational Linguistics, Minneapolis, Minnesota, pp. 4171–4186.

<https://aclanthology.org/N19-1423/>

Douze, M., Guzhva, A., Deng, C., Johnson, J., Szilvasy, G., Mazar'e, P.-E., Lomeli, M., Hosseini, L. dan J'egou, H. (2024), 'The faiss library', *ArXiv* **abs/2401.08281**.

<https://api.semanticscholar.org/CorpusID:267028372>

Fauzi, M. A., Arifin, A. Z. dan Yuniarti, A. (2017), 'Arabic book retrieval using class and book index based term weighting', *International Journal of Electrical and Computer Engineering* **7**, 3705–3710.

<https://api.semanticscholar.org/CorpusID:30084732>

Goworek, R., Macmillan-Scott, O. dan Özyigit, E. B. (2025), 'Bridging language gaps: Advances in cross-lingual information retrieval with multilingual llms', *ArXiv* **abs/2510.00908**.

<https://api.semanticscholar.org/CorpusID:281705770>

Habash, N. Y. (2010), *Introduction to Arabic natural language processing*, Morgan & Claypool Publishers.

<https://doi.org/10.1007/978-3-031-02139-8>

Jaleel, N. A., Allan, J., Croft, W. B., Diaz, F., Larkey, L. S., Li, X., Smucker, M. D. dan Wade, C. (2004), Umass at TREC 2004: Novelty and HARD, in E. M. Voorhees dan L. P. Buckland, eds, 'Proceedings of the Thirteenth Text REtrieval Conference, TREC 2004, Gaithersburg, Maryland, USA, November 16-19, 2004', Vol. 500-261 of *NIST Special Publication*, National Institute of Standards and Technology (NIST).

<http://trec.nist.gov/pubs/trec13/papers/umass.novelty.hard.pdf>

Karpukhin, V., Oguz, B., Min, S., Lewis, P., Wu, L., Edunov, S., Chen, D. dan Yih, W.-t. (2020), Dense passage retrieval for open-domain question answering, in B. Webber, T. Cohn, Y. He dan Y. Liu, eds, 'Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)', Association for Computational Linguistics, Online, pp. 6769–6781.

<https://aclanthology.org/2020.emnlp-main.550/>

Kong, A., Zhao, S., Chen, H., Li, Q., Qin, Y., Sun, R. dan Zhou, X. (2023), Better zero-shot reasoning with role-play prompting, in 'North American Chapter of the Association for Computational Linguistics'.

<https://api.semanticscholar.org/CorpusID:260900230>

Kudo, T. dan Richardson, J. (2018), SentencePiece: A simple and language independent subword tokenizer and detokenizer for neural text processing, in E. Blanco dan W. Lu, eds, 'Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations', Association for Computational Linguistics, Brussels, Belgium, pp. 66–71.

<https://aclanthology.org/D18-2012/>

Larkey, L. S., Ballesteros, L. dan Connell, M. E. (2007), Light stemming for arabic information retrieval.

<https://api.semanticscholar.org/CorpusID:17788645>

Linli, P. (2024), 'A cross-language information retrieval method based on multi-task learning', *MALCOM: Indonesian Journal of Machine Learning and Computer Science* 4(3), 853–862.

<https://journal.irpi.or.id/index.php/malcom/article/view/1384>

Littman, M. L., Dumais, S. T. dan Landauer, T. K. (1998), *Automatic Cross-Language Information Retrieval Using Latent Semantic Indexing*, Springer US, Boston, MA, pp. 51–62.

https://doi.org/10.1007/978-1-4615-5661-9_5

Macdonald, C., Tonellotto, N., MacAvaney, S. dan Ounis, I. (2021), Pyterrier: Declarative experimentation in python from bm25 to dense retrieval, in 'Proceedings of the 30th ACM International Conference on Information & Knowledge Management', CIKM '21, Association for Computing Machinery, New York, NY, USA, p. 4526–4533.

<https://doi.org/10.1145/3459637.3482013>

Manning, C. D. (2008), *Introduction to information retrieval*, Syngress Publishing,.

Marone, M., Weller, O., Fleshman, W., Yang, E., Lawrie, D. dan Durme, B. V. (2025), 'mmbert: A modern multilingual encoder with annealed language learning'.

<https://arxiv.org/abs/2509.06888>

Mayfield, J., Yang, E., Lawrie, D. J., Barham, S., Weller, O., Mason, M., Nair, S. dan Miller, S. (2023), ‘Synthetic cross-language information retrieval training data’, *ArXiv* **abs/2305.00331**.

<https://api.semanticscholar.org/CorpusID:258426878>

Narto, N. et al. (2019), PROBLEMATIKA KESULITAN BELAJAR PADA PEMBELAJARAN FIQH BERBASIS KITAB KUNING SISWA KELAS X DI MA IBNUL QOYYIM PUTRA BANTUL YOGYAKARTA, PhD thesis, UIN Sunan Kalijaga.

<http://digilib.uin-suka.ac.id/id/eprint/37967>

Nigst, L., Romanov, M., Savant, S. B., Seydi, M. dan Verkinderen, P. (2020), ‘Openiti: a machine-readable corpus of islamicate texts’.

<https://doi.org/10.5281/zenodo.4075046>

Ogilvie, P. dan Callan, J. (2003), Combining document representations for known-item search, in ‘Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval’, SIGIR ’03, Association for Computing Machinery, New York, NY, USA, p. 143–150.

<https://doi.org/10.1145/860435.860463>

Pan, X. (2025), ‘Exploring rwkv for sentence embeddings: Layer-wise analysis and baseline comparison for semantic similarity’, *ArXiv* **abs/2502.14620**.

<https://api.semanticscholar.org/CorpusID:276482395>

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E. Z., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J. dan Chintala, S. (2019), ‘Pytorch: An imperative style, high-performance deep learning library’, *CoRR* **abs/1912.01703**.

<http://arxiv.org/abs/1912.01703>

Pavlova, V. (2025), Multi-stage training of bilingual islamic LLM for neural passage retrieval, in S. Yagi, S. Yagi, M. Sawalha, B. A. Shawar, A. T. AlShdaifat, N. Abbas dan Organizers, eds, ‘Proceedings of the New Horizons in Computational Linguistics for Religious Texts’, Association for Computational Linguistics, Abu Dhabi, UAE, pp. 42–52.

<https://aclanthology.org/2025.clrel-1.4/>

- Purwarianti, A., Tsuchiya, M. dan Nakagawa, S. (2007), 'Indonesian-japanese transitive translation using english for clir', *Journal of Natural Language Processing* **14**(2), 95–123.
<https://cir.nii.ac.jp/crid/1390282679452710272>
- Rahmanda, R. A., Adriani, M. dan Tanaya, D. (2019), Cross language information retrieval using parallel corpus with bilingual mapping method, in '2019 International Conference on Asian Language Processing (IALP)', pp. 222–227.
<https://doi.org/10.1109/IALP48816.2019.9037705>
- Reimers, N. dan Gurevych, I. (2019), 'Sentence-bert: Sentence embeddings using siamese bert-networks', *ArXiv abs/1908.10084*.
<https://api.semanticscholar.org/CorpusID:201646309>
- Saleh, S. dan Pecina, P. (2020), Document translation vs. query translation for cross-lingual information retrieval in the medical domain, in D. Jurafsky, J. Chai, N. Schlueter dan J. Tetreault, eds, 'Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics', Association for Computational Linguistics, Online, pp. 6849–6860.
<https://aclanthology.org/2020.acl-main.613/>
- Sanh, V., Debut, L., Chaumond, J. dan Wolf, T. (2019), 'Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter', *ArXiv abs/1910.01108*.
<https://api.semanticscholar.org/CorpusID:203626972>
- Saracevic, T. (2017), *Effects of Inconsistent Relevance Judgments on Information Retrieval Test Results*, Springer International Publishing, Cham, pp. 81–87.
https://doi.org/10.1007/978-3-031-02302-6_8
- Sari, S. dan Adriani, M. (2008), Using mutual information technique in cross-language information retrieval, in 'International Conference on Asian Digital Libraries'.
<https://api.semanticscholar.org/CorpusID:37773664>
- Sari, S. dan Adriani, M. (2014), 'Learning to rank for determining relevant document in indonesian-english cross language information retrieval using bm25', *2014 International Conference on Advanced Computer Science and Information System* pp. 309–314.
<https://api.semanticscholar.org/CorpusID:34898389>

- Valentini, F., Kozlowski, D. dan Lariviere, V. (2025), CLIRudit: Cross-lingual information retrieval of scientific documents, *in* D. I. Adelani, C. Arnett, D. Ataman, T. A. Chang, H. Gonen, R. Raja, F. Schmidt, D. Stap dan J. Wang, eds, ‘Proceedings of the 5th Workshop on Multilingual Representation Learning (MRL 2025)’, Association for Computational Linguistics, Suzhuo, China, pp. 226–242.
<https://aclanthology.org/2025.mrl-main.16/>
- Wan-Chik, R. (2015), ‘Information credibility assessment of islamic and quranic information on the web’, *Proceedings of the 9th International Conference on Ubiquitous Information Management and Communication* .
<https://api.semanticscholar.org/CorpusID:16370459>
- Wang, B., Kuo, C.-C. J. dan Li, H. (2022), ‘Just rank: Rethinking evaluation with word and sentence similarities’, *ArXiv abs/2203.02679*.
<https://api.semanticscholar.org/CorpusID:247291808>
- Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M. dan Brew, J. (2019), ‘Huggingface’s transformers: State-of-the-art natural language processing’, *CoRR abs/1910.03771*.
<http://arxiv.org/abs/1910.03771>
- Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., Klingner, J., Shah, A., Johnson, M., Liu, X., Kaiser, L., Gouws, S., Kato, Y., Kudo, T., Kazawa, H., Stevens, K., Kurian, G., Patil, N., Wang, W., Young, C., Smith, J. R., Riesa, J., Rudnick, A., Vinyals, O., Corrado, G. S., Hughes, M. dan Dean, J. (2016), ‘Google’s neural machine translation system: Bridging the gap between human and machine translation’, *ArXiv abs/1609.08144*.
<https://api.semanticscholar.org/CorpusID:3603249>
- Yang, E., Lawrie, D. J., McNamee, P. dan Mayfield, J. (2024), ‘Extending translate-train for colbert-x to african language clir’, *ArXiv abs/2404.08134*.
<https://api.semanticscholar.org/CorpusID:269137335>
- Zerrouki, T. (2024), ‘Tashaphyne: A python package for arabic light stemming’, *Journal of Open Source Software* **9**(93), 6063.
<https://doi.org/10.21105/joss.06063>
- Zhelezniak, V., Savkov, A., Shen, A. dan Hammerla, N. Y. (2019), ‘Correlation coefficients and semantic textual similarity’, *ArXiv abs/1905.07790*.

<https://api.semanticscholar.org/CorpusID:159041087>