ACKNOWLEDGMENT

This work is supported by the School of Computing, Telkom University, Bandung.

REFERENCES

[1] Iskakova, S., Berkimbaeva, R., Akkoziyeva, R., and Yeleussinova, A., "E-Document Systems: Trends and Perspectives," *Bulletin Series of Physics & Mathematical Sciences*, vol. 2020-4, pp. 1728–7901.30, 2020. [Online]. Available: https://doi.org/10.51889/2020-4.1728-7901.30.

[2] Tate, A., and Smallwood, C., "Comparing the efficiency of paper-based and electronic data capture during face-to-face interviews," *PLoS ONE*, vol. 16, 2021. [Online]. Available: https://doi.org/10.1371/journal.pone.0247570.

[3] Hanifa, R. M., Khairuddin, N. K. A., Rosdi, N. Q., Thomas, L. B. A., and Mohamad, S., "Development of Document Management System (DMSys): A Step Forward to a Paperless Office," *Int. J. Bus. Technol. Manag.*, vol. 5, no. S5, pp. 94–102, 2023. [Online]. Available: https://doi.org/10.55057/ijbtm.2023.5.s5.11.

[4] Gamido, M., Gamido, H., and Macaspac, D., "Electronic document management system for local area network-based organizations," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 31, no. 2, pp. 1154–1163, 2023. [Online]. Available: https://doi.org/10.11591/ijeecs.v31.i2.pp1154-1163.

[5] Tang, M., Chen, J., Chen, H., Xu, Z., Wang, Y., Xie, M., and Lin, J., "An ontology-improved vector space model for semantic retrieval," *Electron. Libr.*, vol. 38, pp. 919–942, 2020. [Online]. Available: https://doi.org/10.1108/el-04-2020-0081.

[6] He, L., Huang, Z., Chen, E., Liu, Q., Tong, S., Wang, H., Lian, D., and Wang, S., "An Efficient and Robust Semantic Hashing Framework for Similar Text Search," *ACM Transactions on Information Systems*, vol. 41, pp. 1–31, 2023. [Online]. Available: https://doi.org/10.1145/3570725.

[7] Taipalus, T., "Vector database management systems: Fundamental concepts, use-cases, and current challenges," *Cognitive Systems Research*, vol. 85, 101216, 2024.

[8] Atlan, "Relational vs Document Database," [Online]. Available: https://atlan.com/relational-vs-document-database/#summarizing-it-all-together.

[9] Pan, J., Wang, J., and Li, G., "Survey of Vector Database Management Systems," *VLDB J.*, vol. 33, pp. 1591–1615, 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2310.14021.

[10] Orlov, I., "Role of document management system for business processes optimization," *Problems of Theory and Methodology of Accounting, Control and Analysis*, 2024. [Online]. Available: https://doi.org/10.26642/pbo-2024-2(58)-45-49.

[11] Microsoft, "Windows Libraries," [Online]. Available: https://learn.microsoft.com/en-us/windows/client-management/client-tools/windows-libraries.

[12] Atwal, T., Scanlon, M., and Le-Khac, N., "Shining a light on Spotlight: Leveraging Apple's desktop search utility to recover deleted file metadata on macOS," *Digit. Investig.*, vol. 28, Supplement, pp. S105–S115, 2019. [Online]. Available: https://doi.org/10.1016/j.diin.2019.01.019.

[13] Corral, K., LaBrie, R., and St. Louis, R., "Document Search Practices," in *Research on Technologies and Applications in Document Management*, pp. 209–217, 2011. [Online]. Available: https://doi.org/10.4018/978-1-59904-931-1.CH020.

[14] A. N. Istiqamah and K. R. S. Wiharja, "A schema extraction of document-oriented database for data warehouse," Int. Journal on Information and Communication Technology (IJoICT), vol. 7, no. 2, pp. 36-47, Dec. 2021 [Online]. Available: https:/doi.org/10.21108/ijoict.v7i2.584.

[15] Kc, A., and Tl, D., "Vector Databases," *Open Access Research Journal of Engineering and Technology*, vol. 7, no. 1, 2024. [Online]. Available: https://doi.org/10.53022/oarjet.2024.7.1.0043.

[16] Wang, J., Yi, X., Guo, R., Jin, H., Xu, P., Li, S., and Xie, C., "Milvus: A purpose-built vector data management system," in *Proc. 2021 International Conference on Management of Data*, pp. 2614–2627, Jun. 2021.

[17] Galli, C., Donos, N., and Calciolari, E., "Performance of 4 Pre-Trained Sentence Transformer Models in the Semantic Query of a Systematic Review Dataset on Peri-Implantitis," *Inf.*, vol. 15, no. 2, p. 68, 2024. [Online]. Available: https://doi.org/10.3390/info15020068.

[18] Lashkari, F., Bagheri, E., and Ghorbani, A., "Neural embedding-based indices for semantic search," *Inf. Process. Manag.*, vol. 56, pp. 733–755, 2019. [Online]. Available: https://doi.org/10.1016/J.IPM.2018.10.015.

[19] Palangi, H., Deng, L., et al., "Deep sentence embedding using deep belief networks," *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 10–18, 2015.

[20] Hugging Face, "all-MiniLM-L6-v2," *Hugging Face Model Hub*. [Online]. Available: https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2. [Accessed: Dec. 15, 2024].

[21] Malkov, Y. A., and Yashunin, D. A., "Efficient and robust approximate nearest neighbor search using hierarchical navigable small world graphs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 4, pp. 824–836, 2018. [Online]. Available: https://doi.org/10.1109/TPAMI.2018.2889473.

[22] Lin, J., "Operational advice for dense and sparse retrievers: HNSW, flat, or inverted indexes?," *arXiv preprint arXiv:2409.06464*, 2024. [Online]. Available: https://doi.org/10.48550/arXiv.2409.06464.

[23] Chai, C., "Comparison of text preprocessing methods," *Natural Language Engineering*, vol. 29, pp. 509–553, 2022. [Online]. Available: https://doi.org/10.1017/S1351324922000213.

[24] Masua, B., Masasi, N., Maziku, H., and Mbwilo, B., "The impact of applying different pre-processing techniques on Swahili textual data using Doc2Vec," *Natural Language Processing Research*, vol. 3, no. 1-2, pp. 1–13, 2023. [Online]. Available: https://doi.org/10.55060/j.nlpre.230606.001.

[25] Rezqa, E., and Baraka, R., "Document Classification Based on Metadata and Keywords Extraction," in *Proc. 2021 Palestinian International Conference on Information and Communication Technology (PICICT)*, pp. 18–24, 2021. [Online]. Available: https://doi.org/10.1109/PICICT53635.2021.00016.

[26] Wilianto, D., and Girsang, A., "Automatic short answer grading on high school's e-learning using semantic similarity methods," *TEM Journal*, 2023. [Online]. Available: https://doi.org/10.18421/tem121-37.

[27] Elliott, O., and Clark, J., "The impacts of data, ordering, and intrinsic dimensionality on recall in hierarchical navigable small worlds," *arXiv preprint arXiv:2405.17813*, 2024. [Online]. Available: https://doi.org/10.48550/arXiv.2405.17813.

[28] Qwen Team, "Qwen2.5: A Party of Foundation Models," Sep. 2024. [Online]. Available: https://qwenlm.github.io/blog/qwen2.5/. [Accessed: Dec. 01, 2024].

[29] Yang, A., Yang, B., Hui, B., Zheng, B., Yu, B., Zhou, C., Li, C., Li, C., Liu, D., Huang, F., Dong, G., Wei, H., Lin, H., Tang, J., Wang, J., Yang, J., Tu, J., Zhang, J., Xu, J., Zhou, J., Bai, J., He, J., Lin, J., Dang, K., Lu, K., Chen, K., Yang, K., Li, M., Xue, M., Ni, N., Zhang, P., Wang, P., Peng, R., Men, R., Gao, R., Lin, R., Wang, S., Bai, S., Tan, S., Zhu, T., Li, T., Liu, T., Ge, W., Deng, X., Zhou, X., Ren, X., Zhang, X., Wei, X., Ren, X., Fan, Y., Yao, Y., Zhang, Y., Wan, Y., Chu, Y., Cui, Z., Zhang, Z., and Fan, Z., "Qwen2 Technical Report," *ArXiv*, abs/2407.10671, 2024. [Online]. Available: https://doi.org/10.48550/arXiv.2407.10671.

[30] Zalkow, F., Brandner, J., and Müller, M., "Efficient Retrieval of Music Recordings Using Graph-Based Index Structures," *Signals*, vol. 2, pp. 336–352, 2021. [Online]. Available: https://doi.org/10.3390/SIGNALS2020021.

[31] Sitarz, M., "Extending the F1 metric with a probabilistic approach to better assess model performance in unbalanced datasets," *IEEE Access*, vol. 10, pp. 6571–6580, 2022. doi: 10.1109/ACCESS.2022.3142705.

[32] Pan, J., Wang, J., and Li, G., "Vector Database Management Techniques and Systems," in *Companion of the 2024 International Conference on Management of Data*, 2024. [Online]. Available: https://doi.org/10.1145/3626246.3654691.