UNIVERSITAS
GADJAH MADA
**Rancang Bangun Sistem Manajemen Dokumen dengan Fitur Pencarian Cerdas Berbasis Retrieval-Augmented
Generation Menggunakan Large Language Model pada PT Parama Data Unit**
Muhammad Hanif Sauqi, Dr.Eng. Ir. Ganjar Alfian, S.T., M.Eng.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

DAFTAR PUSTAKA

[1] R. Kurniah, "Penerapan Sistem Informasi Berbasis Web Untuk Pengelolaan dan Pengarsipan Dokumen," *Infotek J. Inform. Dan Teknol.*, vol. 6, no. 2, pp. 258–267, Jul. 2023, doi: 10.29408/jit.v6i2.11946.

[2] S. Jordan, S. Sternad Zabukovšek, and I. Šišovska Klančnik, "Document Management System – A Way to Digital Transformation," *Naše Gospodarstvo/Our Economy*, vol. 68, no. 2, pp. 43–54, 2022, doi: 10.2478/ngoe-2022-0010.

[3] C. D. R. Amirillah, S. Andriyana, and B. Benrahman, "Perancangan Aplikasi Document Management System Berbasis Web Universitas Nasional dengan Metode Waterfall," *STRING Satuan Tulisan Ris. Dan Inov. Teknol.*, vol. 5, no. 1, p. 45, Aug. 2020, doi: 10.30998/string.v5i1.6353.

[4] Y. Luan, J. Eisenstein, K. Toutanova, and M. Collins, "Sparse, Dense, and Attentional Representations for Text Retrieval," *Trans. Assoc. Comput. Linguist.*, vol. 9, pp. 329–345, Apr. 2021, doi: 10.1162/tacl_a_00369.

[5] Y. Zhu et al., "Large Language Models for Information Retrieval: A Survey," *ACM Trans. Inf. Syst.*, vol. 44, no. 1, pp. 1–54, Jan. 2026, doi: 10.1145/3748304.

[6] P. Lewis et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," in *Proc. 34th Int. Conf. Neural Information Processing Systems (NeurIPS)*, vol. 33, 2020, pp. 9459–9474.

[7] N. F. Liu, K. Lin, J. Hewitt, A. Paranjape, M. Bevilacqua, F. Petroni, and P. Liang, "Lost in the Middle: How Language Models Use Long Contexts," arXiv:2307.03172, Jul. 2023.

[8] S. M. Alade, "Design and Implementation of a Web-based Document Management System," *International Journal of Information Technology and Computer Science (IJITCS)*, vol. 15, no. 2, pp. 35–53, Apr. 2023, doi: 10.5815/ijitcs.2023.02.04.

[9] K. Muludi, K. M. Fitria, J. Triloka, and Sutedi, "Retrieval-Augmented Generation Approach: Document Question Answering using Large Language Model," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 15, no. 3, 2024, doi: 10.14569/IJACSA.2024.0150379.

[10] B. Sarmah, B. Hall, R. Rao, S. Patel, S. Pasquali, and D. Mehta, "HybridRAG: Integrating Knowledge Graphs and Vector Retrieval Augmented Generation for Efficient Information Extraction," arXiv:2408.04948, 2024.

[11] M. L. Husain, Y. Wibisono, and A. Anisyah, "Development of an Academic Services Chatbot Based on Retrieval-Augmented Generation (RAG)," *Brilliance: Research of Artificial Intelligence*, vol. 5, no. 2, pp. 727–735, Aug. 2025, doi: 10.47709/brilliance.v5i2.6719.

[12] B. Saha, U. Saha, and M. Z. Malik, "QuIM-RAG: Advancing Retrieval-Augmented Generation with Inverted Question Matching for Enhanced QA Performance," *IEEE Access*, vol. 12, pp. 185401–185410, Dec. 2024, doi: 10.1109/ACCESS.2024.3513155.

UNIVERSITAS
GADJAH MADA
**Rancang Bangun Sistem Manajemen Dokumen dengan Fitur Pencarian Cerdas Berbasis Retrieval-Augmented
Generation Menggunakan Large Language Model pada PT Parama Data Unit**
Muhammad Hanif Sauqi, Dr.Eng. Ir. Ganjar Alfian, S.T., M.Eng.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

[13] R. Dayarathne, U. Ranaweera, and U. Ganegoda, "Comparing the Performance of LLMs in RAG-based Question-Answering: A Case Study in Computer Science Literature," arXiv:2511.03261, 2025.
[14] S. Krishnan and N. Subramanian, "Semantic Search in a Document Management System Using Retrieval Augmented Generation," in *2025 IEEE/ACIS 29th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, Busan, Korea, Jun. 2025, doi: 10.1109/SNPD65828.2025.11252585.
[15] L. Joos, D. A. Keim, and M. T. Fischer, "Leveraging LLMs for Semi-Automatic Corpus Filtration in Systematic Literature Reviews," arXiv:2510.11409, Oct. 2025.
[16] I. Sommerville, *Software Engineering*, 9th ed. Boston, MA, USA: Pearson, 2011.
[17] Object Management Group, OMG Unified Modeling Language (OMG UML), Version 2.5.1. Object Management Group, Dec. 2017. [Online]. Available: https://www.omg.org/spec/UML/2.5.1.
[18] P. P.-S. Chen, "The Entity-Relationship Model—Toward a Unified View of Data," *ACM Trans. Database Syst.*, vol. 1, no. 1, pp. 9–36, Mar. 1976, doi: 10.1145/320434.320440.
[19] Vercel, Next.js Documentation. [Online]. Available: https://nextjs.org/docs. Accessed: Jul. 3, 2026.
[20] OpenJS Foundation, Express.js: Fast, Unopinionated, Minimalist Web Framework for Node.js. [Online]. Available: https://expressjs.com. Accessed: Jul. 3, 2026.
[21] PostgreSQL Global Development Group, About PostgreSQL. [Online]. Available: https://www.postgresql.org/about/. Accessed: Aug. 11, 2026.
[22] A. Kane, pgvector: Open-Source Vector Similarity Search for Postgres. [Online]. Available: https://github.com/pgvector/pgvector. Accessed: Jul. 3, 2026.
[23] M. Jones, J. Bradley, and N. Sakimura, "JSON Web Token (JWT)," RFC 7519, Internet Engineering Task Force, May 2015. [Online]. Available: https://www.rfc-editor.org/rfc/rfc7519.
[24] S. R. Bhat, M. Rudat, J. Spiekermann, and N. Flores-Herr, "Rethinking Chunk Size For Long-Document Retrieval: A Multi-Dataset Analysis," arXiv:2505.21700, May 2025.
[25] J. Chen, S. Xiao, P. Zhang, K. Luo, D. Lian, and Z. Liu, "M3-Embedding: Multi-Linguality, Multi-Functionality, Multi-Granularity Text Embeddings Through Self-Knowledge Distillation," arXiv:2402.03216, Feb. 2024.
[26] M. Zhang, X. Zhang, X. Zhao, S. Huang, B. Hu, and M. Zhang, "On the Role of Pretrained Language Models in General-Purpose Text Embeddings: A Survey," *ACM Comput. Surv.*, arXiv:2507.20783, Jul. 2025.
[27] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. Cambridge, U.K.: Cambridge Univ. Press, 2008.
[28] P. Liu, W. Yuan, J. Fu, Z. Jiang, H. Hayashi, and G. Neubig, "Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural

UNIVERSITAS
GADJAH MADA
**Rancang Bangun Sistem Manajemen Dokumen dengan Fitur Pencarian Cerdas Berbasis Retrieval-Augmented
Generation Menggunakan Large Language Model pada PT Parama Data Unit**
Muhammad Hanif Sauqi, Dr.Eng. Ir. Ganjar Alfian, S.T., M.Eng.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

Language Processing," *ACM Comput. Surv.*, vol. 55, no. 9, Art. no. 195, pp. 1–35, Jan. 2023, doi: 10.1145/3560815.

[29] J. Wei, X. Wang, D. Schuurmans, M. Bosma, B. Ichter, F. Xia, E. Chi, Q. V. Le, and D. Zhou, "Chain-of-Thought Prompting Elicits Reasoning in Large Language Models," in *Proc. 36th Int. Conf. Neural Information Processing Systems (NeurIPS)*, New Orleans, LA, USA, Dec. 2022, pp. 24824–24837.

[30] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention Is All You Need," in *Proc. 31st Int. Conf. Neural Information Processing Systems (NeurIPS)*, Long Beach, CA, USA, Dec. 2017, pp. 5998–6008.

[31] Ollama, Ollama: Get Up and Running with Large Language Models Locally. [Online]. Available: https://ollama.com. Accessed: Jul. 3, 2026.

[32] A. Yang et al., "Qwen3 Technical Report," arXiv:2505.09388, May 2025.

[33] Y. Huang, J. Xu, J. Lai, Z. Jiang, T. Chen, Z. Li, Y. Yao, X. Ma, L. Yang, H. Chen, S. Li, and P. Zhao, "Advancing Transformer Architecture in Long-Context Large Language Models: A Comprehensive Survey," arXiv:2311.12351, Nov. 2023.

[34] S. Raschka, "Understanding and Implementing Qwen3 From Scratch," *Ahead of AI*, Sep. 2025. [Online]. Available: https://magazine.sebastianraschka.com/p/qwen3-from-scratch

[35] J. Su, Y. Lu, S. Pan, A. Murtadha, B. Wen, and Y. Liu, "RoFormer: Enhanced Transformer with Rotary Position Embedding," *Neurocomputing*, vol. 568, Art. no. 127063, Feb. 2024, arXiv:2104.09864.

[36] Y. Gao, Y. Xiong, X. Gao, K. Jia, J. Pan, Y. Bi, Y. Dai, J. Sun, M. Wang, and H. Wang, "Retrieval-Augmented Generation for Large Language Models: A Survey," arXiv:2312.10997, Dec. 2023.

[37] M. Iratni, M. Boughanem, and T. Dkaki, "Dynamic Context Selection for Retrieval-Augmented Generation: Mitigating Distractors and Positional Bias," arXiv:2512.14313, Dec. 2025.

[38] N. S. Pillai and R. R. Hemamalini, "Hybrid User Acceptance Test Procedure to Improve the Software Quality," *International Arab Journal of Information Technology*, vol. 19, no. 6, pp. 956–964, Nov. 2022, doi: 10.34028/iajit/19/6/14.

[39] International Organization for Standardization, ISO/IEC 25010:2023 Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Product quality model, 2nd ed. Geneva, Switzerland: ISO, 2023.

[40] International Organization for Standardization, ISO/IEC 25019:2023 Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Quality-in-use model, 1st ed. Geneva, Switzerland: ISO, 2023.

[41] Aliyah, N. Hartono, and A. A. Muin, "Penggunaan User Acceptance Testing (UAT) pada Pengujian Sistem Informasi Pengelolaan Keuangan dan Inventaris Barang," *Switch: Jurnal Sains dan Teknologi Informasi*, vol. 3, no. 2, pp. 42–58, Mar. 2025, doi: 10.62951/switch.v3i1.330.

UNIVERSITAS
GADJAH MADA
Rancang Bangun Sistem Manajemen Dokumen dengan Fitur Pencarian Cerdas Berbasis Retrieval-Augmented Generation Menggunakan Large Language Model pada PT Parama Data Unit
Muhammad Hanif Sauqi, Dr.Eng. Ir. Ganjar Alfian, S.T., M.Eng.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

[42] G. Faggioli, L. Dietz, C. Clarke, G. Demartini, M. Hagen, C. Hauff, N. Kando, E. Kanoulas, M. Potthast, B. Stein, and H. Wachsmuth, "Perspectives on Large Language Models for Relevance Judgment," in *Proc. 2023 ACM SIGIR Int. Conf. Theory of Information Retrieval (ICTIR)*, Taipei, Taiwan, Jul. 2023, pp. 39–50, doi: 10.1145/3578337.3605136.

[43] P. Thomas, S. Spielman, N. Craswell, and B. Mitra, "Large Language Models can Accurately Predict Searcher Preferences," in *Proc. 47th Int. ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR)*, Washington, DC, USA, Jul. 2024, pp. 1930–1940, doi: 10.1145/3626772.3657707.

[44] Qwen, Qwen3-4B-Instruct-2507. [Online]. Available: https://huggingface.co/Qwen/Qwen3-4B-Instruct-2507. Accessed: Sep. 10, 2026.

[45] Ollama, "Modelfile Reference." [Online]. Available: https://docs.ollama.com/modelfile. Accessed: Oct. 7, 2026.

[46] Ollama, "Context length." [Online]. Available: https://docs.ollama.com/context-length. Accessed: Oct. 7, 2026.