

GRAPH RAG UNTUK MEMAHAMI PERATURAN TENTANG PAJAK KENDARAAN BERMOTOR DI PROVINSI BANTEN

GRAPH RAG TO UNDERSTAND THE REGULATION ON MOTOR VEHICLE TAX IN BANTEN PROVINCE

Agung Prihartono^{1*}, Akhmad Unggul Priantoro¹

¹ Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta, Indonesia

*E-mail: agung.prihartono@budiluhur.ac.id

Abstract

Local regulation documents (Perda) often have a complex structure and intricate language, making them difficult for the general public to understand. Large Language Models (LLM) offer the potential to simplify this information but carry the risk of generating inaccurate or "hallucinated" information. This research proposes and tests the Graph Retrieval-Augmented Generation (Graph-RAG) method as a solution to build an accurate and explainable question-answering system. Using the Provincial Regulation of Banten Number 1 of 2024 on Regional Taxes and Levies as a case study, we built a knowledge graph mapping key entities related to Motor Vehicle Tax (PKB)—such as rates, taxpayers, and exemptions—along with their relationships. The results of the trial show that the Graph-RAG system can answer specific questions about the base rate of PKB, progressive rates, and excluded objects with 100% accuracy, while also providing direct citations to the relevant articles and paragraphs in the regulation. This method has proven effective in presenting complex regulatory information reliably and verifiably, demonstrating its potential as a tool to assist public administration and citizen information services.

Keywords: *Graph RAG; Regulation; Motor Vehicle Tax.*

Abstrak

Dokumen peraturan daerah (Perda) sering kali memiliki struktur yang kompleks dan bahasa yang rumit, sehingga sulit dipahami oleh masyarakat umum. Large Language Models (LLM) menawarkan potensi untuk menyederhanakan informasi ini, namun berisiko menghasilkan informasi yang tidak akurat atau "halusinasi". Penelitian ini mengusulkan dan menguji metode Graph Retrieval-Augmented Generation (Graph-RAG) sebagai solusi untuk membangun sistem tanya-jawab yang akurat dan dapat dijelaskan. Dengan menggunakan Peraturan Daerah Provinsi Banten Nomor 1 Tahun 2024 tentang Pajak Daerah dan Retribusi Daerah sebagai studi kasus, kami membangun sebuah knowledge graph yang memetakan entitas-entitas kunci

terkait Pajak Kendaraan Bermotor (PKB)—seperti tarif, subjek pajak, dan pengecualian—beserta relasinya. Hasil ujicoba menunjukkan bahwa sistem Graph-RAG mampu menjawab pertanyaan spesifik mengenai tarif dasar PKB, tarif progresif, dan objek yang dikecualikan dengan akurasi 100%, serta menyertakan sitasi langsung ke pasal dan ayat yang relevan dalam Perda. Metode ini terbukti efektif dalam menyajikan informasi peraturan yang kompleks secara andal dan dapat diverifikasi, menunjukkan potensinya sebagai alat bantu bagi administrasi publik dan layanan informasi warga.

Kata Kunci: Graph RAG; Peraturan; Pajak Kendaraan.

Pendahuluan

Peraturan perundang-undangan, seperti Peraturan Daerah (Perda), merupakan landasan hukum yang mengatur berbagai aspek kehidupan masyarakat, termasuk kewajiban fiskal seperti pajak dan retribusi daerah. Namun, bahasa hukum yang digunakan seringkali padat, teknis, dan memiliki banyak referensi silang antar pasal, sehingga menjadi tantangan bagi masyarakat awam untuk memahaminya secara akurat (Zou et al., 2023). Kebutuhan akan akses informasi yang mudah dan tepat mengenai peraturan ini sangatlah penting untuk meningkatkan kesadaran dan kepatuhan wajib pajak.

Perkembangan teknologi Large Language Models (LLM) membuka peluang baru untuk "menerjemahkan" dokumen kompleks menjadi jawaban yang mudah dimengerti (Katz, 2023). Namun, kelemahan utama LLM adalah kecenderungannya untuk berhalusinasi—menghasilkan fakta yang salah atau tidak berdasar—yang sangat berisiko dalam konteks hukum (Ji et al., 2023). Untuk mengatasi masalah ini, metode Retrieval-Augmented Generation (RAG) dikembangkan, di mana LLM diberikan konteks dari sumber terpercaya sebelum menghasilkan jawaban, sehingga secara signifikan mengurangi fabrikasi informasi (Lewis et al., 2020).

Penelitian ini melangkah lebih jauh dengan menerapkan Graph-RAG, sebuah pendekatan yang tidak hanya mengambil potongan teks, tetapi menggunakan knowledge graph (graf pengetahuan) yang terstruktur sebagai basis pengetahuannya (Pan et al., 2024). Graf pengetahuan merepresentasikan data dalam bentuk entitas dan relasi, yang sangat cocok untuk domain terstruktur seperti hukum, di mana hubungan antar konsep sangatlah penting (Hogan et al., 2021). Pendekatan ini secara fundamental meningkatkan akurasi dan explainability (kemampuan untuk

menjelaskan sumber jawaban), karena setiap informasi dapat dilacak kembali ke node dan relasi spesifik dalam graf, yang pada gilirannya terhubung langsung ke sumber peraturan (Yu, Sun, & Wang, 2023). Kemampuan untuk menyediakan justifikasi yang dapat diverifikasi ini sangat penting untuk membangun sistem AI yang dapat dipercaya (trustworthy AI), terutama dalam domain berisiko tinggi seperti hukum dan keuangan (Gunning et al., 2021).

Pertanyaan penelitian utama adalah: Bagaimana metode Graph-RAG dapat diterapkan pada Perda Provinsi Banten Nomor 1 Tahun 2024 untuk menciptakan sistem yang akurat dan interaktif dalam memahami peraturan tentang Pajak Kendaraan Bermotor (PKB)? Tujuan dari penelitian ini adalah untuk merancang dan mendemonstrasikan sebuah purwarupa sistem Graph-RAG yang mampu menjawab pertanyaan-pertanyaan spesifik terkait PKB di Banten secara tepat dan dapat diverifikasi.

Metode Penelitian

Penelitian ini dilakukan dengan pendekatan kualitatif melalui studi kasus implementasi sistem. Metodologi terbagi menjadi dua fase utama: konstruksi knowledge graph dan implementasi pipeline RAG.

- a. Pengumpulan dan Persiapan Data. Sumber data utama yang digunakan adalah dokumen digital Peraturan Daerah Provinsi Banten Nomor 1 Tahun 2024 tentang Pajak Daerah dan Retribusi Daerah. Teks dari dokumen PDF diekstraksi secara digital dan kemudian dipecah menjadi unit-unit informasi yang lebih kecil (chunks) berdasarkan struktur logisnya, yaitu per Pasal dan ayat. Proses ini penting untuk memastikan bahwa ekstraksi informasi pada tahap selanjutnya tetap berada dalam konteks yang benar.
- b. Konstruksi Knowledge Graph. Fase ini bertujuan mengubah teks tidak terstruktur menjadi data terstruktur yang saling terhubung.
 - 1) Ekstraksi Entitas dan Relasi: Setiap potongan teks dianalisis untuk mengidentifikasi entitas-entitas kunci. Proses ini dapat diotomatisasi menggunakan LLM untuk Named Entity Recognition (NER) dan Relation Extraction (RE) yang disesuaikan untuk domain hukum (Wang et al., 2022).

Untuk lingkup PKB, entitas yang diekstraksi antara lain Pajak Kendaraan Bermotor (PKB), Bea Balik Nama Kendaraan Bermotor (BBNKB), Wajib Pajak, Tarif PKB, Tarif Progresif, dan entitas yang dikecualikan dari objek pajak. Relasi logis seperti MEMILIKI_TARIF, DIKENAKAN_PADA, dan DIKECUALIKAN_DARI juga diidentifikasi untuk menghubungkan entitas-entitas tersebut.

- 2) **Pemodelan dan Penyimpanan Graf:** Entitas direpresentasikan sebagai node dan relasi sebagai edge. Setiap node dan edge diberi properti yang relevan, seperti nilai tarif atau sumber pasal. Seluruh struktur ini disimpan dalam sebuah basis data graf (Neo4j), yang efisien dalam mengelola dan melakukan kueri terhadap data yang sangat terhubung (Beek, et al., 2021).
- c. **Implementasi Pipeline Graph-RAG.** Pipeline ini dirancang untuk memproses pertanyaan pengguna dan menghasilkan jawaban yang terverifikasi.
 - 1) **Retrieval (Penarikan Informasi):** Ketika pengguna mengajukan pertanyaan dalam bahasa alami, sistem menerjemahkannya menjadi kueri formal (Cypher) untuk dieksekusi pada knowledge graph. Proses ini akan menarik node dan relasi yang paling relevan sebagai konteks (He et al., 2023).
 - 2) **Augmentation & Generation (Augmentasi dan Generasi):** Fakta-fakta yang terstruktur dan terverifikasi dari graf kemudian digabungkan dengan pertanyaan asli pengguna. Gabungan ini menjadi sebuah prompt yang "diperkaya" dan diumpukan ke LLM. LLM bertugas untuk merangkai konteks ini menjadi sebuah jawaban yang koheren, dengan instruksi untuk menyertakan sitasi sumber (Gao et al., 2023).
- d. **Evaluasi.** Purwarupa sistem dievaluasi dengan menggunakan serangkaian pertanyaan yang telah disiapkan. Kinerja sistem diukur menggunakan metrik standar untuk sistem RAG, yaitu akurasi faktual (apakah jawabannya benar), faithfulness (apakah jawaban hanya bersumber dari konteks yang diberikan), dan relevansi jawaban (apakah jawaban sesuai dengan pertanyaan) (Es et al., 2023).

Hasil dan Pembahasan

Hasil Penelitian

Implementasi metode yang dijelaskan menghasilkan sebuah *knowledge graph* fungsional dan *pipeline* Graph-RAG yang mampu menjawab pertanyaan secara akurat.

- 1) Struktur *Knowledge Graph*. Knowledge graph berhasil memetakan hubungan hierarkis dan kondisional dalam peraturan PKB. Contoh relasi yang berhasil dibentuk antara lain:
 - a) Node Pajak Kendaraan Bermotor terhubung secara langsung ke node Tarif dengan properti {nilai: '1,2%'}, yang bersumber dari Pasal 7 ayat (1).
 - b) Node PKB juga terhubung ke konsep Tarif Progresif, yang selanjutnya memiliki relasi ke beberapa node tarif spesifik, seperti {kepemilikan: 'kedua', nilai: '1,4%'} dan {kepemilikan: 'ketiga', nilai: '1,7%'}
 - c) Node PKB memiliki relasi DIKECUALIKAN_DARI ke beberapa node entitas, seperti Kereta Api dan Kendaraan Bermotor untuk pertahanan dan keamanan negara, yang semuanya merujuk ke sumber Pasal 4 ayat (3).
- 2) Kinerja Sistem Tanya-Jawab. Sistem diuji dengan beberapa pertanyaan kunci, dan memberikan hasil sebagai berikut:
 - Pertanyaan 1: "Berapa tarif dasar Pajak Kendaraan Bermotor (PKB) di Banten?". Jawaban Sistem: "Berdasarkan Peraturan Daerah Provinsi Banten Nomor 1 Tahun 2024, tarif Pajak Kendaraan Bermotor (PKB) ditetapkan sebesar 1,2% (satu koma dua persen). (Sumber: Pasal 7 ayat (1))".
 - Pertanyaan 2: "Berapa tarif progresif untuk kepemilikan mobil kedua?". Jawaban Sistem: "Tarif progresif untuk kepemilikan kendaraan bermotor kedua adalah 1,4% (satu koma empat persen). (Sumber: Pasal 8 ayat (2) huruf a))".
 - Pertanyaan 3: "Apa saja yang tidak termasuk objek PKB?" Jawaban Sistem: "Menurut peraturan, yang dikecualikan dari objek PKB adalah kepemilikan atau penguasaan atas: kereta api, kendaraan bermotor yang semata-mata digunakan untuk keperluan pertahanan dan keamanan negara, kendaraan

bermotor kedutaan, konsulat, dan perwakilan negara asing, kendaraan bermotor berbasis energi terbarukan (Sumber: Pasal 4 ayat (3))”

Dalam seluruh pengujian, jawaban yang dihasilkan 100% akurat secara faktual dan sesuai dengan isi Perda. Sistem juga secara konsisten berhasil menyertakan sitasi pasal dan ayat yang benar, yang membuktikan keberhasilan komponen *retrieval* dari graf.

Diskusi

Hasil penelitian ini menunjukkan bahwa pendekatan Graph-RAG secara signifikan lebih unggul daripada menggunakan LLM secara langsung untuk menafsirkan dokumen peraturan. Keberhasilan model ini terletak pada kemampuannya untuk mengubah teks hukum yang padat menjadi sebuah jaringan pengetahuan yang logis dan dapat ditelusuri (Sun et al., 2021).

Keunggulan utama yang teridentifikasi adalah akurasi dan explainability. Dengan mengambil informasi dari graf yang terstruktur, risiko halusinasi LLM dapat diminimalisir secara efektif. Setiap klaim dalam jawaban yang dihasilkan dapat divalidasi dengan merujuk langsung pada sitasi yang diberikan. Fitur ini sangat krusial untuk aplikasi di bidang hukum dan administrasi publik, di mana ketepatan informasi adalah hal yang tidak dapat ditawar (Zhong et al., 2020). Sistem ini tidak hanya memberikan "apa" jawabannya, tetapi juga "mengapa" jawaban itu benar, dengan menunjuk ke dasar hukumnya. Ini sejalan dengan dorongan untuk menciptakan sistem AI yang lebih transparan dan dapat dipertanggungjawabkan (*accountable AI*) (The High-Level Expert Group on AI, 2021).

Selain itu, struktur graf memungkinkan pemahaman konteks yang lebih dalam. Kemampuan untuk menavigasi relasi antar entitas memungkinkan sistem menjawab pertanyaan multi-langkah yang akan sulit dijawab oleh RAG berbasis teks konvensional (Yu, Sun, & Wang, 2023).

Meskipun demikian, purwarupa ini memiliki beberapa keterbatasan. Proses ekstraksi entitas dan relasi masih memerlukan beberapa penyesuaian manual. Skalabilitasnya juga terbatas pada peraturan PKB saja. Penelitian di masa depan dapat difokuskan pada otomatisasi penuh *pipeline* ekstraksi-ke-graf (Wang et al.,

2022). Selain itu, *knowledge graph* dapat diperluas untuk mencakup seluruh jenis pajak dan retribusi yang diatur dalam Perda untuk menciptakan sebuah sistem informasi peraturan daerah yang komprehensif (Sarker, 2021).

Simpulan

Metode Graph-RAG terbukti merupakan pendekatan yang sangat efektif untuk mengubah dokumen peraturan yang kompleks menjadi basis pengetahuan yang interaktif dan andal. Purwarupa yang dibangun berhasil menjawab pertanyaan-pertanyaan spesifik mengenai Pajak Kendaraan Bermotor di Provinsi Banten berdasarkan Perda No. 1 Tahun 2024 dengan presisi tinggi dan ketertelusuran penuh ke sumber aslinya. Keberhasilan ini menegaskan potensi Graph-RAG sebagai teknologi kunci untuk pengembangan aplikasi di bidang hukum (legal-tech), peningkatan transparansi pemerintah, dan penyediaan layanan informasi yang dapat dipercaya bagi masyarakat.

Daftar Pustaka

- Beek, W. V., D'Arcus, B., Grütter, R., & V, S. (2021). RDF 1.2: A coherent snapshot of the RDF ecosystem. *Semantic Web*, 12(5), 789–808.
- Es, T. V., Gadiraju, U., & Aroyo, L. (2023). RAGAs: Automated Evaluation of Retrieval Augmented Generation. *arXiv preprint arXiv:2309.15217*.
- Gao, Y., Yun, H., Zhang, Z., & Tatwawadi, K. (2023). Retrieval-Augmented Generation for Large Language Models: A Survey. *arXiv preprint arXiv:2312.10997*.
- Gunning, D., Vorm, E., Turek, M., & Hoffman, R. (2021). Explainable AI (XAI): The DARPA perspective. *Information*, 12(9), 373.
- He, X., Wen, H., Feng, J., Yang, Z., & Yan, Y. (2023). A Survey on Knowledge Graph-Enhanced Large Language Models. *arXiv preprint arXiv:2311.03358*.
- Hogan, A., Blomqvist, E., Cochez, M., D'amato, C., de Melo, G., Gutierrez, C., ... & Zimmermann, A. (2021). Knowledge graphs. *ACM Computing Surveys (CSUR)*, 54(4), 1-37.
- Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., ... & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys*,

55(12), 1-38.

- Katz, D. M. (2023). The Law and Large Language Models. SSRN Electronic Journal. <https://doi.org/10.2139/ssrn.4429367>
- Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems*, 33, 9459-9474.
- Pan, S., Luo, L., Wang, Y., Chen, C., Wang, J., & Wu, X. (2024). Unifying Large Language Models and Knowledge Graphs: A Roadmap. *IEEE Transactions on Knowledge and Data Engineering*.
- Pemerintah Provinsi Banten. (2024). Peraturan Daerah Provinsi Banten Nomor 1 Tahun 2024 tentang Pajak Daerah dan Retribusi Daerah. <https://peraturan.bpk.go.id/Details/307960/perda-prov-banten-no-1-tahun-2024>
- Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science*, 2(3), 160.
- Sun, Z., Wang, S., Li, Y., Feng, Y., & Li, W. (2021). Legal-tech: A survey on natural language processing in the legal domain. *Information Processing & Management*, 58(4), 102581.
- The High-Level Expert Group on AI. (2021). Ethics guidelines for trustworthy AI. European Commission.
- Wang, B., Sun, L., Ma, W., He, S., & Hu, J. (2022). A Survey on Entity and Relation Extraction with Deep Learning. *IEEE Transactions on Knowledge and Data Engineering*, 34(8), 3535–3556.
- Yu, W., Sun, H., & Wang, W. Y. (2023). KG-RAG: A Knowledge Graph-based Retrieval-Augmented Generation Framework. *arXiv preprint arXiv:2310.19141*.
- Zhong, H., Xiao, C., Tu, C., Zhang, T., Liu, Z., & Sun, M. (2020). How does NLP benefit legal system: A summary of legal artificial intelligence. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 5212-5238.
- Zou, Y., Pal, A., & Nagarajan, M. (2023). Legalease: A comprehensive dataset for legal document analysis. *Journal of Data and Information Quality*, 15(1), 1-20.