

ABSTRAK

Direktori Putusan Mahkamah Agung Republik Indonesia telah memuat lebih dari 10,5 juta dokumen putusan hingga akhir April 2026, menjadikan temu kembali informasi hukum yang spesifik dan akurat sebagai tantangan yang signifikan bagi hakim, pengacara, dan akademisi hukum. Sistem pencarian yang tersedia saat ini berbasis kata kunci dan belum mampu menangkap kesamaan makna maupun hubungan struktural antar entitas. Akibatnya, sistem ini berpotensi melewatkan kasus-kasus yang sebenarnya serupa tetapi disampaikan dengan pilihan kata yang berbeda. Penelitian ini mengusulkan pendekatan *Structural Semantic Retrieval* yang mengintegrasikan representasi semantik berbasis Indo-LegalBERT dengan representasi struktural berbasis *Knowledge Graph* yang merepresentasikan relasi subjek, predikat, objek (SPO) dalam kerangka *Case-Based Reasoning* (CBR). Penelitian mengikuti kerangka kerja CRISP-DM dengan mengonversi *Knowledge Graph* menjadi vektor menggunakan Node2Vec, kemudian menggabungkan skor semantik dan struktural melalui fungsi hibrida berbobot (α). Evaluasi dilakukan pada 5.022 sampel dari dataset Indo-Law yang memuat 20.326 dokumen putusan pidana umum menggunakan teknik *stratified sampling* dan metrik *Precision@3*, *Recall@3*, dan *MRR@3*. Hasil eksperimen menunjukkan bahwa konfigurasi optimal diperoleh pada $\alpha = 0,85$, dengan *Precision@3* sebesar 0,6547 dan *MRR@3* sebesar 0,7370, lebih tinggi dibandingkan representasi semantik murni ($\alpha = 1,00$) yang menghasilkan *Precision* 0,6427 dan *MRR* 0,7307. Ditemukan pula 843 kasus di mana representasi struktural berhasil memperbaiki peringkat yang tidak tepat dari pendekatan semantik. Temuan ini mengindikasikan bahwa informasi struktural berperan sebagai sinyal penguat yang efektif, meskipun peningkatannya masih terbatas akibat keterbatasan pendekatan *triples* sederhana dalam membedakan konteks hukum yang kompleks.

Kata Kunci: *Case-Based Reasoning*, Dokumen Hukum, Indo-LegalBERT, *Knowledge Graph*, *Structural Semantic Retrieval*.

ABSTRACT

The Supreme Court Decision Directory of the Republic of Indonesia has accumulated over 10.5 million decision documents as of late April 2026, rendering specific and accurate legal information retrieval a significant challenge for judges, lawyers, and legal scholars. Current keyword-based retrieval systems are unable to capture semantic similarities and structural relationships between entities. Consequently, these systems risk overlooking cases that are substantially similar but expressed using different wording. This study proposes a Structural Semantic Retrieval approach that integrates Indo-LegalBERT-based semantic representation with Knowledge Graph-based structural representation depicting Subject-Predicate-Object (SPO) relations within a Case-Based Reasoning (CBR) framework. Following the CRISP-DM framework, the Knowledge Graph is converted into vectors using Node2Vec, and the semantic and structural scores are subsequently combined through a weighted hybrid function (α). Evaluation was conducted on 5,022 samples from the Indo-Law dataset, which comprises 20,326 general criminal decision documents, employing stratified sampling alongside Precision@3, Recall@3, and MRR@3 metrics. Experimental results demonstrate that the optimal configuration is achieved at $\alpha = 0.85$, yielding a Precision@3 of 0.6547 and an MRR@3 of 0.7370. These figures outperform the pure semantic representation ($\alpha = 1.00$), which yields a Precision of 0.6427 and an MRR of 0.7307. Furthermore, the structural representation successfully corrected inaccurate rankings from the purely semantic approach in 843 cases. These findings indicate that structural information acts as an effective reinforcing signal, even though its improvement remains constrained by the limitations of a simple Triples approach in distinguishing complex legal contexts

Keywords: *Case-Based Reasoning, Indo-LegalBERT, Knowledge Graph, Legal Documents, Structural Semantic Retrieval.*