

Konstruksi Knowledge Graph pada Dokumen Peraturan Perundang-undangan Indonesia Menggunakan Large Language Model = Knowledge Graph Construction on Indonesian Legal Documents using Large Language Model

Muhammad Haddad, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920567789&lokasi=lokal>

Abstrak

Indonesia merupakan negara hukum yang mengadopsi asas Fictie Hukum. Asas tersebut memandang setiap orang tahu hukum tanpa pengecualian. Penyediaan akses yang mudah terhadap produk hukum merupakan konsekuensi dari hal tersebut. Meski telah tersedia beberapa layanan daring pencarian hukum, baik oleh pemerintah maupun swasta, layanan tersebut belum mampu menangkap relasi intradokumen dan antardokumen dengan baik. Dalam meningkatkan sistem pencarian hukum, terdapat knowledge graph (KG) bernama LexID yang menghadirkan representasi peraturan perundang-undangan Indonesia dalam sebuah graf. KG tersebut dikonstruksi dengan pendekatan rule-based. Namun, pendekatan rule-based tidak mudah beradaptasi dengan perubahan dalam format atau konten dokumen dan memerlukan pemeliharaan berkelanjutan. Penelitian ini mengusulkan pendekatan lain dalam konstruksi LexID. Proses konstruksi LexID dilakukan menggunakan pre-trained large language model (LLM) berupa CodeGemma parameter 7B, Code Llama parameter 7B, dan Phi-3 parameter 7B. Jenis prompt yang digunakan, yaitu prompt kode dan teks dengan variasi 1-shot dan 2-shot, sehingga total terdapat dua belas skenario percobaan. Hasil konstruksi KG kemudian dievaluasi terhadap KG LexID dan diukur menggunakan metrik precision, recall, dan skor F1. Dari hasil konstruksi, didapatkan skor F1 hasil dari prompt teks 1-shot: CodeGemma 0.405, CodeLlama 0.452, dan Phi 0.362; skor F1 hasil dari prompt kode 1-shot: CodeGemma 0.645, CodeLlama 0.567, dan Phi 0.526; skor F1 hasil dari prompt teks 2-shot: CodeGemma 0.572, CodeLlama 0.502, dan Phi 0.386; skor F1 hasil dari prompt kode 2-shot: CodeGemma 0.687, CodeLlama 0.583, dan Phi 0.539.

.....Indonesia operates under a legal system that adopts the principle of Legal Fiction, which posits that every individual is presumed to be aware of the law without exception. Consequently, providing easy access to legal documents is imperative. Despite the availability of several online legal search services offered by both government and private entities, these services have yet to effectively capture intra-document and inter-document relationships. To enhance the legal search system, a knowledge graph (KG) named LexID has been developed to represent Indonesian legislation in a graph format. This KG has traditionally been constructed using a rule-based approach. However, this approach struggles to adapt to changes in document format or content and necessitates continuous maintenance. This study proposes an alternative approach for the construction of LexID utilizing pre-trained large language models (LLMs), specifically CodeGemma with 7 billion parameters, Code Llama with 7 billion parameters, and Phi-3 with 7 billion parameters. The study employs both code and text prompts, with variations of 1-shot and 2-shot, resulting in a total of twelve experimental scenarios. The constructed KG is then evaluated against the existing LexID KG, using precision, recall, and F1 score metrics. The results of the construction indicate the following F1 scores: for 1-shot text prompts, CodeGemma achieved 0.405, Code Llama 0.452, and Phi 0.362; for 1-shot code prompts, CodeGemma achieved 0.645, Code Llama 0.567, and Phi 0.526; for 2-shot text prompts, CodeGemma achieved 0.572, Code Llama 0.502, and Phi 0.386; and for 2-shot code prompts, CodeGemma

achieved 0.687, Code Llama 0.583, and Phi 0.539.