

Pengembangan model tanya jawab hukum berbasis LexID knowledge graph dengan model pipeline berbasis BERT (Bidirectional Encoder Representations from Transformers) = Development of a legal question answering model over LexID knowledge graph with pipeline model based on BERT (Bidirectional Encoder Representations from Transformers)

Muhammad Rafif Priyo Waskito, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920553190&lokasi=lokal>

Abstrak

Peraturan perundang-undangan baru di Indonesia dapat mengganti, menghapus, atau menambahkan aturan yang sudah berlaku sehingga setiap warga Indonesia memiliki tanggung jawab untuk mengetahui aturan terbaru saat hendak digunakan. Laman pengumpulan informasi ini sudah ada di laman pemerintah resmi seperti peraturan.go.id. Akan tetapi ada informasi yang tidak mudah dicari seperti isi pasal terbaru pada doku- men yang telah diubah oleh dokumen lain. Sekarang, sudah ada informasi peraturan perundang- undangan yang disimpan dalam knowledge graph (KG)-sebuah basis data terstruktur berbentuk kumpulan entitas dan relasi layaknya sebuah graf- bernama LexID KG (Muninggar & Krisnadhi, 2023). Ditambah lagi, LexID QA (Handi, 2023), yaitu sebuah sistem tanya jawab hukum dengan informasinya dari LexID KG, menangani masalah ini tetapi dengan memaksakan struktur inputnya. Menurut Jain, Kumar, Kota, dan Patel (2018) kebebasan membuat input (free-text) merupakan tur bagus karena seakan-akan sedang berinteraksi dengan manusia. Oleh karena itu, penelitian ini berusaha untuk menyelesaikan permasalahan pemaksaan struktur input pada LexID QA dengan beralih ke model machine learning pipeline. Model pipeline-nya tersusun dari model klasi kasi pertanyaan bebas ke kueri SPARQL, sebuah bahasa untuk mengambil informasi dari KG, dan model NER untuk mengambil informasi entitas penting seperti judul dokumen yang dicari pada pertanyaan bebasnya. Kedua model ini berbasiskan BERT (Bidirectional Encoder Representations from Transformers) (Devlin, Chang, Lee, & Toutanova, 2018). Hasil dari model klasi kasi mencapai nilai sempurna pada dataset baru yang dibuat pada penelitian ini. Kemudian, hasil dari model pipeline-nya juga mencapai nilai makro-akurasi sebesar 0.949 pada dataset yang sama.

.....New laws and regulations in Indonesia can modify, repeal, or adding to existing rules so that every Indonesian citizen has responsibility to know the latest rules when utilizing them. This collection of information is available on of cial government websites such as peraturan.go.id. However, there is information that not easy to nd, such as the contents of latest articles in the documents that have been modi ed by another document. Currently, there is regulatory information stored in a knowledge graph (KG)-a database structured in the form of a collection of entities and relationships like a graph- called LexID KG (Muninggar & Krisnadhi, 2023). Additionally, LexID QA (Handi, 2023), a legal question and answer system with information from LexID KG, addresses this problem but forcing its input structure. According to Jain et al. (2018), free-text input is a good feature because it feels like communication with humans. Therefore, this research trying to solve the problem of forcing input structures in LexID QA by transitioning to a machine learning pipeline model. The pipeline model is composed of free question classi cation model to query SPARQL, a language for retrieving information from KG, and NER model to extracting important entity such as legal title in the free-text input. Both model are based on BERT (Bidirectional Encoder

Representations from Transformers) (Devlin et al., 2018). The classification model achieved a perfect score on the new dataset created in this research. Furthermore, the pipeline model achieved a macro-accuracy score of 0.949 on the same dataset.