

Named Entity Recognition pada Dokumen Legal Berbahasa Indonesia = Named Entity Recognition on Indonesian Legal Documents

Jafar Abdurrohman, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920541914&lokasi=lokal>

Abstrak

Sebagai negara hukum, Indonesia memiliki sistem peradilan yang aktif. Setiap bulannya, terdapat sekitar 100.000 dokumen putusan yang dihasilkan oleh lembaga kehakiman di Indonesia. Volume dokumen yang banyak tersebut menimbulkan suatu tantangan bagi insan hukum di Indonesia. Guna mendukung dan mengakomodasi institusi publik dari aspek teknologi dan informasi, serta mendorong pelaksanaan keterbukaan informasi bagi masyarakat umum, maka dibutuhkan suatu metode yang dapat membantu untuk mencari dan mengumpulkan informasi penting dari suatu dokumen putusan pengadilan. Secara khusus, metode ini juga ditujukan untuk membantu para praktisi hukum untuk kepentingan penegakan hukum dan para akademisi hukum untuk kepentingan pendidikan dan pengembangan di bidang hukum. Salah satu teknik untuk mengumpulkan informasi penting dari suatu dokumen adalah Named Entity Recognition (NER). Teknik NER bekerja dengan cara menandai kata-kata yang merupakan informasi penting seperti orang, tempat, lokasi, waktu, kejadian, dan lainnya. NER dapat diterapkan untuk berbagai bidang permasalahan, seperti medis, hukum, dan pertanian. NER yang secara spesifik bekerja untuk menandai entitas hukum disebut dengan Legal Entity Recognition (LER). Penelitian sebelumnya telah menerapkan LER untuk dokumen legal berbahasa Indonesia dengan pendekatan BiLSTM dan CRF sehingga diperlukan penelitian lebih lanjut untuk mengetahui bagaimana performa model-model lain terhadap kasus tersebut. Penelitian ini menguji performa language model, yaitu model berbasis RoBERTa dan model berbasis BERT serta membandingkannya dengan deep learning model, yaitu BiLSTM dan BiLSTM-CRF sebagai model baseline penelitian sebelumnya untuk task LER bahasa Indonesia. Hasil penelitian menunjukkan model berbasis RoBERTa memiliki performa terbaik untuk task LER pada dataset penulis, XLM-R large dengan skor F1 sebesar 0,9295, XLM-R base dengan skor F1 sebesar 0,9281 dan Indonesian RoBERTa dengan skor F1 sebesar 0,9246.

As a rule of law country, Indonesia has an active justice system. Every month, there are around 100,000 decision documents produced by the judiciary in Indonesia. The large volume of documents poses a challenge for legal people in Indonesia. In order to support and accommodate public institutions from the aspects of technology and information, as well as encourage the implementation of information disclosure for the general public, a method is needed that can help find and collect important information from a court decision document. In particular, this method is also intended to assist legal practitioners for the benefit of law enforcement and legal academics for the benefit of education and development in the field of law. One technique for collecting important information from a document is Named Entity Recognition (NER). The NER technique works by marking words that are important information such as people, places, locations, times, events, and so on. NER can be applied to various problem areas, such as medical, legal, and agriculture. NER which specifically works to mark legal entities is called Legal Entity Recognition (LER). Previous studies have applied LER to legal documents in Indonesian using the BiLSTM and CRF approaches, so further research is needed to find out how other models perform in this case. This study examines the performance of language models, namely the

RoBERTa-based model and the BERT-based model and compares them with deep learning models, namely BiLSTM and BiLSTM-CRF as the baseline models for previous studies for the Indonesian language LER task. The results showed that the RoBERTa-based model had the best performance for the LER task in the author's dataset, XLM-R large with an F1 score of 0.9295, XLM-R base with an F1 score of 0.9281 and Indonesian RoBERTa with an F1 score of 0.9246.</p>