

Penerapan Causal Language Model untuk Tugas Textual Entailment pada Domain Hukum = Utilization of Causal Language Models for Legal Textual Entailment Task

Bryan Tjandra, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920577454&lokasi=lokal>

Abstrak

Penelitian ini menjelaskan partisipasi kami dalam tugas Legal Textual Entailment (Tugas 4) pada kompetisi Competition on Legal Information Extraction/Entailment (COLIEE) 2025. Kami mengeksplorasi tiga metode berbeda untuk mengatasi tantangan ini. Metode pertama kami, UIRunFTune, mengadaptasi penelitian sebelumnya dari tim JNLP dengan menggunakan seleksi prompt dan fine-tuning LLM ringan dengan QLoRA, sebagai model baseline dengan akurasi resmi sebesar 60,27%. Metode kedua, UIRunLang, terbukti menjadi strategi kami yang paling efektif dengan mencapai akurasi resmi sebesar 82,19%. Pendekatan ini memperkenalkan mekanisme seleksi otomatis untuk memilih penggunaan masukan berbahasa Inggris atau Jepang, yang secara signifikan meningkatkan kinerja dengan mengatasi kesalahan penerjemahan kritis pada kumpulan data berbahasa Inggris. Metode ketiga kami, UIRunCoT, memanfaatkan ensemble voting dari teknik prompting advanced, termasuk Chain-of-Thought, Tree-of-Thought, dan Graph-of-Thought. Metode ini mencapai akurasi tertinggi sebesar 82,46% dalam evaluasi internal kami, yang menunjukkan kekuatan dari kerangka penalaran terstruktur. Hasil kami menunjukkan bahwa strategi adaptif-bahasa menawarkan peningkatan substansial dibandingkan fine-tuning standar dan bahwa metode penalaran advanced memiliki potensi yang menjanjikan untuk tugas legal entailment yang kompleks.

.....This paper describes our participation in the Legal Textual Entailment (Task 4) of the Competition on Legal Information Extraction/Entailment (COLIEE) 2025. We explored three distinct methods to tackle the challenge. Our first method, UIRunFTune, adapted the JNLP team's prior work by using prompt selection and a lightweight LLM with QLoRA fine-tuning, which established a baseline performance with an official accuracy of 60.27%. The second method, UIRunLang, proved to be our most effective strategy, achieving an official accuracy of 82.19%. This approach introduced an automatic selection mechanism to use either the English or Japanese input, significantly improving performance by addressing critical mistranslations in the English dataset. Our third method, UIRunCoT, leveraged a voting ensemble of advanced prompting techniques, including Chain-of-Thought, Tree-of-Thought, and Graph-of-Thought. This run achieved the highest accuracy of 82.46% in our internal evaluations, demonstrating the power of structured reasoning frameworks. Our results demonstrate that a language-adaptive strategy offers substantial gains over standard fine-tuning and that advanced reasoning methods hold significant promise for complex legal entailment.