

ABSTRAK

Dokumen perundang-undangan di Indonesia memiliki struktur kalimat majemuk yang sangat kompleks dan kaku, sehingga menyulitkan masyarakat awam memahami substansi hukum secara utuh. Penelitian ini bertujuan mengembangkan sistem penyederhanaan teks hukum otomatis yang mampu menurunkan tingkat kerumitan bahasa tanpa mendistorsi makna esensial pasal terkait. Pendekatan utama yang digunakan adalah rekayasa komputasional melalui adaptasi domain pada arsitektur model bahasa IndoT5. Pemodelan dilatih menggunakan 1.623 pasang korpus paralel teks hukum asli dan penyederhanaannya. Algoritma pembangkit teks dimodifikasi dengan parameter ambang batas panjang dinamis 60 persen dan penalti teks pendek untuk mencegah kompresi berlebihan dan hilangnya makna materiil. Evaluasi kinerja sistem dilakukan secara kuantitatif menggunakan metrik komputasi dan secara kualitatif melalui triangulasi kepakaran. Pengujian otomatis menghasilkan skor SARI 47.64 dan skor BLEU 0.0898. Penurunan presisi susunan kata wajar terjadi karena model proaktif mengeksekusi teknik eksplisitasi atau penambahan frasa penjelas guna mengeliminasi ambiguitas. Secara kualitatif, pakar linguistik menyimpulkan sistem beroperasi sangat baik dan layak digunakan tanpa perbaikan sebagai agen parafrase natural. Sementara itu, pakar hukum menyatakan sistem layak digunakan dengan perbaikan, mengingat sensitivitas yurisprudensi terhadap pergeseran frasa mutlak. Intervensi ambang batas terbukti berhasil merestorasi frasa krusial pidana dan menyisipkan contoh kasus konkret secara otonom. Purwarupa kecerdasan buatan ini difungsikan murni sebagai instrumen pendukung literasi publik dan bukan substitusi dokumen otentik negara.

Kata Kunci: IndoT5, Penyederhanaan Teks, Hukum, *SARI*, Eksplisitasi, Pemrosesan Bahasa Alami, Literasi Hukum.

ABSTRACT

Legal documents in Indonesia possess complex and rigid compound sentence structures, making it difficult for the general public to comprehend the legal substance. This research aims to develop an automatic legal text simplification system capable of reducing language complexity without distorting the essential meaning of the articles. The primary approach uses computational engineering through domain adaptation on the IndoT5 language model architecture. The model was trained using 1,623 pairs of parallel corpora containing original and simplified legal texts. The generation algorithm was modified using a 60 percent dynamic length threshold and a short text penalty to prevent excessive sentence compression and meaning loss. System performance was evaluated quantitatively using computational metrics and qualitatively through expert triangulation. Automatic testing yielded a SARI score of 47.64 and a BLEU score of 0.0898. The decline in exact word precision occurred naturally because the model proactively executed explicitation techniques by adding explanatory phrases to eliminate ambiguity. Qualitatively, linguistic experts concluded the system operates excellently and is suitable for use without improvements as a natural paraphrasing agent. Meanwhile, legal experts stated the system is suitable with improvements, considering jurisprudential sensitivity toward shifting absolute phrases. The threshold intervention successfully restored crucial criminal phrases and autonomously inserted concrete case examples. This AI prototype functions purely as a supporting literacy instrument for the public and is not a substitute for authentic state documents.

Keywords: *IndoT5, Text Simplification, Legal, SARI, Explicitation, Natural Language Processing, Legal Literacy.*