

Abstrak

Semakin besarnya volume berita elektronik berbahasa Indonesia mengakibatkan informasi tersedia dalam jumlah yang besar, beraneka ragam, dan umumnya tidak terstruktur. Hal ini mendorong terjadinya peningkatan kebutuhan untuk mencari dan mengelola informasi dengan baik dari dokumen berita berbahasa Indonesia sehingga dihasilkan pengetahuan yang bermanfaat. *Text mining* merupakan upaya penambangan data yang berupa teks dimana sumber data biasanya diperoleh dari dokumen, dengan tujuan mencari kata-kata yang dapat mewakili isi dokumen sehingga dapat dilakukan analisis keterhubungan antar dokumen.

NLP digunakan sebagai *preprocessing* dalam *text mining* karena *user* menentukan relevansi dokumen dengan membaca dan menganalisisnya. *Preprocessing* pada penelitian ini dibatasi hanya pada tiga tahapan, yaitu: *POS tagging*, *syntax parsing*, dan *semantic role labeling (SRL)*.

Dalam Tugas Akhir ini, implementasi *SRL* didasarkan pada teori *case grammar*. Label yang digunakan mengacu pada label PropBank. *SRL* dibagi menjadi dua *task* utama, yaitu indentifikasi *frame* dan pelabelan elemen-elemen *frame* yang sudah teridentifikasi, sesuai peran semantisnya.

Hasil implementasi *semantic role labeling* dapat memecahkan permasalahan *SRL* kalimat bahasa Indonesia untuk kalimat tunggal dan deklaratif secara terkomputerisasi. Pengujian dengan menggunakan 52 kalimat yang diambil dari 300 artikel yang diambil dari Harian Kompas memberikan hasil pelabelan yang akurat untuk semua kata dalam setiap kalimat, jika dibandingkan dengan pelabelan secara manual.

Kata Kunci: *text mining, preprocessing, natural language processing, semantic role labeling, case grammar.*