

Abstrak

Peringkasan teks otomatis (*Automated Text Summarization*) merupakan proses pengidentifikasian hal penting dari sebuah teks dan mengekspresikannya dalam sebuah dokumen yang lebih singkat. Salah satu teknik peringkasan teks adalah teknik *sentence compression* yang meringkas teks pada level kalimat, menggunakan data set yang diambil dari Ziff-Davis corpus yang berisi mengenai penjualan produk komputer. Kompresi kalimat menggunakan penghapusan, penyisipan dan penggantian kata dalam rangka menghasilkan ringkasan. Tugas akhir ini digunakan pendekatan statistik dengan metode *Lexicalized Markov Grammar* yang didasari oleh *Markov Chain Rule* yang memodelkan tiga buah model yaitu *source model* $P(c)$, *channel model* $P(l/c)$ dan decoder $\text{argmax } P(c/l)$, dimana c adalah kalimat ringkasan dan l adalah kalimat panjang. Permasalahan pada metode ini adalah mengidentifikasi noise, yaitu kata yang tidak penting dan harus dibuang agar kalimat ringkasan dapat terbentuk. Hasil eksperimen menunjukkan bahwa eksperimen ini berhasil mengimplementasikan metode *lexicalized markov grammar* untuk sentence compression memberikan tingkat *similarity* hingga 81.75% dengan ringkasan standar manusia, nilai *grammaticality* sebesar 93.2% dan nilai *importance* sebesar 92.7% yang artinya sudah mendekati hasil ringkasan manusia.

Kata Kunci : *Text Summarization, Lexicalized Markov Grammar, Markov Chain, Sentence Compression.*