Abstrak

Tugas akhir ini menganalisis metode *n-gram* dalam mengenali bahasa suku bangsa Indonesia berbasis teks. Untuk menganalisis akurasi dari metode *n-gram* dilakukan pengujian dengan menggunakan panjang *n* yang berbeda dari *n-gram*, panjang rangking berbeda dari model *n-gram* bahasa, dan pengujian untuk menganalisis pengaruh banyaknya kata di dalam dokumen yang ingin dikenali bahasa daerahnya. Proses pelatihan dilakukan guna membentuk model *n-gram* bahasa dari masing-masing bahasa daerah. Sistem yang dibuat menggunakan metode *n-gram* untuk pemodelan bahasa daerah dan teknik rank-order-statistic untuk pengklasifikasian bahasa daerahnya. Dari seluruh pengujian yang dilakukan didapatkan bahwa untuk akurasi pengenalan bahasa daerah Sunda dan Jawa dapat digunakan panjang minimum *rank=100* dan panjang *n* dari *n-gram* yaitu *n=3, n=4, n=5*, dengan akurasi pengenalan pada penggunaan panjang *rank=100* sebesar 100% untuk *n=3*, 98,75% untuk *n=4*, 97,50% untuk *n=5*. Sedangkan rasio antara panjang *rank* dengan banyaknya kata di dalam dokumen yang ingin dikenali bahasanya yaitu *pjg_rank* : *jml_kata* = 100 : 40, dengan penggunaan panjang minimum *rank=100* dan banyak kata minimum di dalam dokumen yang ingin dikenali = 40 kata.

Kata kunci: n-gram, performansi, akurasi, rank-order-statistic.