

ABSTRAKSI

Stemming adalah proses untuk menemukan kata dasar dari sebuah kata. Dengan menghilangkan semua imbuhan baik yang terdiri dari awalan (*affixes*), akhiran (*suffixes*) dan *confixes* (kombinasi dari awalan dan akhiran) pada kata turunan. Teknik *stemming* berbeda untuk masing-masing bahasa. Hal ini dikarenakan struktur kata pada setiap bahasa memiliki aturan pembentukan yang berbeda. Seperti *stemming* untuk teks berbahasa Inggris akan berbeda dengan teks berbahasa Indonesia.

Dari hasil implementasi dan pengujian didapatkan bahwa algoritma *stemming* paice/husk dapat diimplementasikan dalam Bahasa Indonesia dengan akurasi dan kekuatan yang cukup baik.

Pada analisa pengaruh stemming dengan menggunakan algoritma paice/husk dalam proses *text mining*, dilakukan proses kategorisasi dengan menggunakan multinominal naïve bayes sebagai *Classifiernya*, didapatkan bahwa proses stemming memperbaiki proses kategorisasi baik dalam hal akurasi, waktu proses, maupun *F-measure* jika dibandingkan dengan dokumen yang tidak terstemming. Hal ini dikarenakan proses *stemming* mengurangi jumlah *term* unik pada dokumen uji.

Kata Kunci: *Stemming, Preprocessing Text Mining, morfologi* Bahasa Indonesia, kategorisasi teks.