

ABSTRAK

Bahasa Jawa menjadi salah satu bahasa daerah di Indonesia. Kini para ahli, guru serta pemerhati bahasa dan budaya Jawa mulai melestarikan dan mengembangkan bahasa serta budaya Jawa. Salah satunya dengan mengembangkan dalam bidang keilmuan, seperti stemming atau pencarian kata dasar. Algoritma *stemming* untuk bahasa yang satu berbeda dengan algoritma stemming untuk bahasa lainnya. Efektifitas *stemming* diukur berdasarkan tingkat akurasi ketepatan dan kecepatan waktu proses. Selain itu efektifitas *stemming* terhambat dengan adanya *overstemming* dan *understemming*. *Overstemming* dan *understemming* adalah kata yang terlalu banyak atau terlalu sedikit dihilangkan pada waktu proses *stemming*. Penelitian ini melakukan *stemming* pada kata-kata berbahasa Jawa untuk mengetahui akurasi dan kecepatan waktu prosesnya yang diterapkan terhadap Algoritma ECS.

Penelitian ini menerapkan metode Enhanced Confix Stripping (ECS) yang disesuaikan dengan morfologi Bahasa Jawa. Dalam morfologi Bahasa Jawa proses pembentukan kata terdapat pada 5 jenis kata, yaitu kata kerja, kata benda, kata sifat, kata keterangan, dan kata bilangan. Dalam metode stemming pencarian kata dasar proses pemenggalan kata dibagi berdasarkan jenis afiks, seperti prefiks, sufiks, infiks, dan konfiks. Algoritma yang digunakan adalah ECS karena memiliki performa yang lebih baik karena telah memodifikasi beberapa aturan dengan adanya *loopPengembalianAkhiran* dan proses *recoding* untuk mengatasi kesalahan pemenggalan.

Penerapan penelitian ini pada perangkat lunak berbasis website menggunakan bahasa pemrograman PHP didukung database MySQL dan *control panel* XAMPP pada sebuah perangkat keras laptop atau komputer. Data stopword yang digunakan untuk *filtering* berjumlah 284 kata. Data kamus berjumlah 28.391 kata yang diperoleh dari Sastra Bausastra karya W. J. S. Poerwadarminta melalui www.sastra.org dan beberapa dari kamus pepak Bahasa Jawa. Data uji dari penelitian menggunakan kata Bahasa Jawa dan data peneliti sebelumnya. Hasil penelitian metode ECS berbasis kamus memiliki akurasi lebih tinggi dan kecepatan waktu proses lebih lama jika dibandingkan dengan Metode Enhanced Confix Stripping (ECS) non kamus. Metode ECS berbasis kamus juga dapat menangani *overstemming* dan *understemming*. Hasil akurasi ECS untuk kata kerja 94,6%, kata benda 95,57%, kata sifat 96,9%, kata keterangan 96,1%, kata bilangan 85,7%, dan rata-rata akurasi sebesar 93,7%.

Kata kunci : stemming, afiks, overstemming, understemming, ECS