

Abstrak

Stemming merupakan proses pembentukan kata dasar dari kata-kata yang telah mendapatkan modifikasi dalam penggunaannya. Penggunaan kata yang terdapat pada kalimat terstruktur diantaranya sudah mendapat imbuhan yang terdiri dari awalan, akhiran ataupun sisipan. *Stemming* merupakan bagian dari *Preprocessing*, yaitu fase terakhir setelah *Tokenization* dan *Stoplist Removal*. Proses *stemming* berbeda dalam tiap bahasa karena dalam pembentukan kata memiliki perbedaan pada tiap bahasa. Pada bahasa Indonesia, ada beberapa algoritma yang dapat dipakai dalam proses *stemming*, diantaranya Algoritma Arifin-Setiono, Algoritma Nazief-Adriani dan Algoritma Enhanced Confix Stripping (ECS) Stemmer. Algoritma ECS adalah algoritma perbaikan dari algoritma Confix Stripping (CS) Stemmer.

Pada tugas akhir ini akan diajukan implementasi dan modifikasi algoritma Enhanced Confix Stripping Stemmer pada teks berbahasa Indonesia. Algoritma Enhanced Confix Stripping Stemmer memiliki kekurangan dan keterbatasan dalam menangani kata yang memiliki sisipan dan kata yang memiliki huruf akhir seperti akhiran. Modifikasi algoritma Enhanced Confix Stripping Stemmer dirancang untuk melakukan perbaikan terhadap kelemahan tersebut. Dari hasil pengujian akan terlihat perbedaan tingkat akurasi antara algoritma Enhanced Confix Stripping Stemmer dan modifikasi algoritma Enhanced Confix Stripping Stemmer, akan dibuktikan bahwa skema yang telah dimodifikasi dapat menghasilkan nilai akurasi yang lebih tinggi.

Kata Kunci : Stemming, Enhanced Confix Stripping Stemmer, Imbuhan, Preprocessing.