

ABSTRAKSI

Stemming merupakan sebuah proses untuk menemukan kata dasar dari bentuk kompleksnya dengan menghilangkan semua afiks yang melekat padanya. Stemming telah diterapkan dalam aplikasi text atau document clustering, classification, summarization, information retrieval dan word-based text compression.

Stemmer berbagai bahasa telah dibuat, termasuk Bahasa Indonesia, tetapi Bahasa Indonesia merupakan salah satu bahasa yang paling rumit diantara bahasa – bahasa lainnya. Bahasa Indonesia mempunyai bentuk - bentuk afiks yang rumit. Yaitu berupa prefix, infix, suffix, konfiks, dan perulangan.

Stemmer Bahasa Indonesia pertama kali dibuat oleh Nazief – Adriani kemudian algoritmanya diperbaiki oleh Jelita Asian yang disebut dengan CS Stemmer. Peningkatan yang sangat besar telah dicapai oleh CS Stemmer sehingga saat ini merupakan algoritma stemmer yang mempunyai akurasi paling tinggi, akan tetapi masih terdapat kesalahan – kesalahan.

Sebuah algoritma baru akan diusulkan untuk memperbaiki algoritma stemmer sebelumnya dengan melakukan modifikasi algoritma Nazief-Adriani khususnya dengan mengatur ulang urutan langkah – langkah proses stemming. Untuk membandingkan akurasi diantara ketiga algoritma stemmer tersebut dilakukan eksperimen dengan 3 sumber dokumen, yaitu sebuah buku novel, hadits dan berita online. Proses stemming menggunakan kamus kata dasar yang diambil dari Kamus Besar Bahasa Indonesia 2008. Hasil dari eksperimen menunjukkan bahwa algoritma baru mempunyai akurasi yang lebih baik dari kedua stemmer Nazief – Adriani dan CS.

Kata kunci: Stemming, Indonesia, Nazief – adriani, CS stemmer, algoritma baru