

Abstrak

Di zaman globalisasi belakangan ini, informasi tentunya menjadi hal yang sangat penting bagi manusia. Dengan informasi yang dapat berupa teks, gambar, ataupun suara, manusia dapat menjawab segala macam bentuk pertanyaan yang muncul untuk memecahkan masalah yang dihadapinya.

Untuk mendapatkan informasi tersebut, manusia dapat menggunakan berbagai macam *search engine* seperti Google, Yahoo, Altavista, ataupun yang lainnya. Di dalam *search engine* tersebut terdapat suatu proses pengembalian kata dasar yang disebut *stemming*. Untuk menghasilkan pencarian dokumen yang akurat, tentunya proses *stemming* tersebut harus baik.

Banyak terdapat algoritma *stemming* yang telah dikembangkan, salah satunya adalah Algoritma Nazief Adriani yang merupakan algoritma *stemming* untuk Bahasa Indonesia. Algoritma ini adalah algoritma terbaik untuk struktur morfologi Bahasa Indonesia. Namun, di dalam pengimplementasiannya masih terdapat beberapa kekurangan, salah satunya adalah penyamarataan makna variasi kata.

Oleh karena itulah digunakan metode *corpus-based stemmer* yang tidak tergantung bahasa untuk menghindari penyamarataan makna variasi kata tersebut. Pada pengimplementasiannya, metode yang menggunakan statistik *co-occurrence* dari variasi kata ini dapat meningkatkan akurasi dari sistem *Information Retrieval*.

Kata kunci : *Search Engine, Stemming, Algoritma Nazief Adriani, Corpus-Based Stemmer, Co-occurrence, Information Retrieval*