

PERBANDINGAN ALGORITMA NAZIEF & ADRIANI, PORTER DAN ARIFIN SETIONO UNTUK DOKUMEN TEKS BAHASA INDONESIA

*Oppie Rezalina¹, Deni Arifianto, M.Kom²,
Jurusan Teknik Informatika, Fakultas Teknik,
Universitas Muhammadiyah Jember
Jln. Karimata No. 49, Telp (0331) 336728, Jember
E-mail : oppierezalina@gmail.com*

ABSTRAK

Pencarian informasi berupa dokumen teks atau yang dikenal dengan istilah *Information Retrieval* (IR) merupakan proses pemisahan dokumen-dokumen yang dianggap relevan dari sekumpulan dokumen yang tersedia. Salah satu bagian penting dari *Information Retrieval* adalah proses *stemming*. *Stemming* adalah proses mereduksi kata berimbuhan menjadi kata dasar. Dengan proses *stemming*, kata yang dimasukkan ke dalam *index* adalah dalam bentuk umum, sehingga dapat menghasilkan dokumen yang lebih relevan. Terdapat beberapa algoritma *stemming* yang memiliki kelebihan dan kekurangan masing-masing. Pada penelitian ini penulis ingin mengetahui performansi dari algoritma *stemming* Nazief & Adriani, Porter dan Arifin Setiono pada pencarian kata dasar yang terdapat dalam abstrak sebuah jurnal dengan implementasi yang dilakukan yaitu memberikan suatu kemudahan dalam hal *stemming* dokumen teks berbahasa Indonesia serta analisis algoritma yang tingkat akurasi yang besar dan membutuhkan waktu yang cepat. Parameter yang diuji berupa kecepatan dan ketepatan dari masing-masing algoritma. Dari hasil pengujian sistem dapat disimpulkan bahwa algoritma Nazief & Adriani memiliki tingkat keakuratan paling tinggi dengan prosentase 0,1% lebih akurat dibandingkan algoritma Arifin Setiono dan 0,9% lebih akurat dibandingkan dengan algoritma Porter. Begitu juga dalam hal kecepatan, algoritma Nazief & Adriani masih lebih cepat menyelesaikan proses *stemming* dibandingkan dengan dua algoritma lainnya. Sebagai saran untuk pengembang berikutnya kamus kata dasar diharapkan lebih lengkap dan melakukan pengembangan terhadap morfologi pada algoritma porter stemmer untuk memperoleh akurasi yang lebih besar.

Kata Kunci: Stemming, Nazief & Adriani, Porter, Arifin Setiono.

THE COMPARISON NAZIEF & ADRIANI, PORTER AND ARIFIN SETIONO STEMMING ALGORITHM FOR INDONESIAN DOCUMENT TEXT

*Oppie Rezalina¹, Deni Arifianto, M.Kom²,
Jurusan Teknik Informatika, Fakultas Teknik,
Universitas Muhammadiyah Jember
Jln. Karimata No. 49, Telp (0331) 336728, Jember
E-mail : oppierezalina@gmail.com*

ABSTRACT

A search for information in the form of a text document or known as Information Retrieval (IR) is a separation process documents that are considered relevant from a set of available documents. One important part of the Information Retrieval is the process of stemming. Stemming is the process of reducing words berimbunan be a basis. With the process of stemming, words entered into the index is in a general form, so it can produce more relevant documents. There are several algorithms stemming which has advantages and disadvantages of each. In this study the authors wanted to determine the performance of the algorithm stemming Nazief & Adriani, Porter and Arifin Setiono on finding basic words contained in the abstract of a journal with implementation done is provide a convenience in terms of stemming document Indonesian-language text as well as analysis algorithm's accuracy is great and requires a fast time. The parameters tested in the form of speed and accuracy of each algorithm. From the test results it can be concluded that the algorithm system Nazief & Adriani has the highest level of accuracy with a percentage of 0.1% is more accurate than algorithms Arifin Setiono and 0.9% more accurate than the Porter algorithm. Likewise, in terms of speed, the algorithm Nazief & Adriani still faster to complete the process stemming compared with two other algorithms. As a suggestion for the next developer basic word dictionary is expected to be more complete and carry out the development of the morphology of the algoirtma porter stemmer to obtain greater accuracy.

Keywords: Stemming, Nazief & Adriani, Porter, Arifin Setiono.