

ABSTRAK

PERBANDINGAN ALGORITMA NAZIEF ADRIANI DAN *PAICE HUSK* UNTUK PROSES *STEMMING* TEKS BAHASA INDONESIA

Oleh:

Linggar Dwi Pratiwi

1127050088

Stemming adalah suatu proses yang terdapat pada tahap *preprocessing* yang mengenali kata dasar dengan cara menggabungkan atau memecahkan setiap varian-varian suatu kata. Algoritma untuk proses *stemming* teks Bahasa Indonesia adalah algoritma Nazief Adriani dan untuk teks bahasa Inggris adalah *Paice Husk*. Saat ini algoritma *Paice Husk* masih digunakan untuk proses *stemming* pada Bahasa Inggris, sehingga untuk menghasilkan proses *stemming* yang baik pada Bahasa Indonesia menggunakan algoritma *Paice Husk* dibutuhkan implementasi aturan-aturan morfologi Bahasa Indonesia pada algoritma. Dari hasil implementasi dan pengujian yang telah dilakukan untuk kompleksitas yang diuji menggunakan metode *cyclomatic complexity* didapatkan hasil nilai 3 untuk Algoritma Nazief Adriani dan 5 untuk Algoritma *Paice Husk*, dalam kategori evaluasi resiko *cyclomatic complexity* rentang 1-10 merupakan *risk evaluation software* bebas resiko error . Akurasi perbandingan antara Algoritma Nazief Adriani dan *Paice Husk* yaitu 91,87% dan 64,43%. Dan waktu yang telah dihitung dan mendapatkan rata-rata hasil waktu proses untuk 200 data kalimat yaitu untuk Algoritma Nazief Adriani 7,3516/detik dan Algoritma *Paice Husk* 7,2510/detik. Dengan demikian, Algoritma yang memiliki tingkat keakuratan yang lebih baik adalah Nazief Adriani tetapi untuk proses waktu lebih cepat *Paice Husk*.

Kata Kunci: *Stemming*, Algoritma Nazief Adriani, Algoritma *Paice Husk*

ABSTRACT

COMPARISON OF NAZIEF ADRIANI AND PAICE HUSK ALGORITHM FOR STEMMING PROCESS INDONESIAN TEXT

By:
Linggar Dwi Pratiwi
1127050088

Stemming is a process that is in the preprocessing stage that recognizes basic words by combining or solving each variant of a word. The algorithm for the Indonesian text stemming process is the Nazief Adriani algorithm and for the English text is Paice Husk. At present the Paice Husk algorithm is still used for the stemming process in English, so to produce a good stemming process in Indonesian using the Paice Husk algorithm requires the implementation of Indonesian morphological rules on the algorithm. From the results of the implementation and testing that has been done for the complexity that was tested using the cyclomatic complexity method, the results obtained value of 3 for the Nazief Adriani Algorithm and 5 for the Paice Husk Algorithm, in the category of risk evaluation cyclomatic complexity range 1-10 is risk evaluation software error free. The accuracy of the comparison between the Nazief Adriani and Paice Husk algorithm is 91.87% and 64.43%. And the time has been calculated and get the average results of the processing time for 200 data sentences, namely for the Nazief Adriani Algorithm 7.3516 / sec and the Paice Husk Algorithm 7.2510 / sec. Therefore, the algorithm which has a better level of accuracy is Nazief Adriani but for faster time processing Paice Husk.

SUNAN GUNUNG DJATI
BANDUNG

Keyword: Stemming, Nazief Adriani Algorithm, Paice Husk Algorithm