

Abstrak

Stemming merupakan proses mengubah kata berimbuhan menjadi bentuk asalnya dengan menggunakan aturan-aturan tertentu. Ada banyak algoritma yang digunakan untuk melakukan proses *stemming*, khususnya *stemming* bahasa Indonesia. Berdasarkan penelitian yang sudah dilakukan sebelumnya, algoritma *stemming* bahasa Indonesia terdiri dari, algoritma Nazief & Andriani, algoritma Yussof & Sembok, algoritma Idris & Mustofa, algoritma Vega, algoritma Arifin & Setiono, algoritma Porter Bahasa Indonesia, algoritma *Confix Stripping* (CS), dan algoritma *Enhanced Confix Stripping* (ECS). Penelitian ini membandingkan performansi algoritma-algoritma *stemming* untuk menemukan algoritma dengan tingkat akurasi yang paling tinggi. Evaluasi dalam penelitian ini dilakukan secara eksplisit dan implisit. Evaluasi eksplisit dilakukan untuk mengukur dan menganalisis secara langsung perbandingan performansi dari setiap algoritma dengan menggunakan parameter akurasi, *Word Conflation Class*, dan *Index Compression Factor*. Kemudian akan dilakukan evaluasi secara implisit yaitu, dengan mengimplementasikan masing-masing algoritma *stemming* bahasa Indonesia ke dalam proses klasifikasi teks bahasa Indonesia dengan menggunakan *tools* Weka dengan *Support Vector Machine* (SVM) sebagai *classifier*. Hasil dari SVM tersebut berupa nilai akurasi dan *F1-Measure* yang menjadi perbandingan performansi dari masing-masing algoritma. Pengujian juga dilakukan pada dokumen uji yang tidak dilakukan proses *stemming* (*nonstem*) dan dokumen uji yang di-*stemming* secara manual. Hasil penelitian menunjukkan bahwa algoritma Nazief & Andriani memiliki nilai evaluasi eksplisit dan implisit lebih tinggi dari pada algoritma-algoritma *stemming* bahasa Indonesia lainnya.

Kata kunci : *preprocessing*, algoritma *stemming* bahasa Indonesia, klasifikasi teks, evaluasi eksplisit, evaluasi implisit.