

ABSTRAK

Stemming adalah proses dalam pemrosesan bahasa alami yang bertujuan untuk mengubah kata-kata menjadi bentuk dasarnya atau kata dasar. Hal ini dilakukan dengan menghilangkan awalan atau akhiran tertentu dari kata agar konteks dan makna tetap terjaga. Namun, dalam bahasa Indonesia, permasalahan muncul ketika variasi awalan dan akhiran sangat beragam, menyebabkan kesulitan dalam menentukan aturan pemotongan yang tepat. Penelitian ini membahas mengapa penggunaan *stemming* penting dalam bahasa Indonesia dan bagaimana algoritma *Conflix Stripping* (CS) dan *Enhanced Conflix Stripping* (ECS) digunakan untuk mengatasi permasalahan ini. CS dan ECS adalah dua metode yang umum digunakan dalam penentuan kata dasar dalam proses *stemming*. *Algoritma Conflix Stripping* adalah metode *stemming* berbasis aturan linguistik untuk bahasa Indonesia. Sedangkan, *Enhanced Conflix Stripping* merupakan pengembangan dari algoritma sebelumnya dengan aturan atau *rule* yang ditingkatkan. Terakhir, hasil eksperimen menunjukkan bahwa akurasi *Conflix Stripping* mencapai 65,62%, sementara ECS menghasilkan akurasi sebesar 66,58%. Penelitian ini memiliki tujuan utama sebagai referensi bagi penelitian berikutnya dalam bidang pengolahan teks. Dengan merangkum penjelasan mengenai *Stemming*, permasalahan variasi bahasa Indonesia, konsep ECS dan *Conflix Stripping*, serta hasil akurat dari kedua algoritma tersebut, penelitian ini berharap dapat memberikan panduan bagi penelitian selanjutnya dalam upaya mengatasi tantangan yang muncul dari perkembangan bahasa dan pengolahan teks.

Kata Kunci - *Stemming*, Bahasa Indonesia, *Conflix Stripping*, *Enhanced Conflix Stripping*, *Text Preprocessing*

ABSTRACT

Stemming constitutes a procedure within the realm of natural language processing that aims to transform words into their foundational forms or root words. This operation involves the removal of specific prefixes or suffixes from words while upholding their contextual significance. Nevertheless, when dealing with the Indonesian language, complexities arise due to the extensive assortment of prefixes and suffixes, leading to challenges in determining precise truncation rules. This study delves into the importance of employing stemming within the Indonesian context and elucidates how the Confix Stripping (CS) and Enhanced Confix Stripping (ECS) algorithms are harnessed to address this intricate issue. Both CS and ECS serve as prevalent techniques for establishing fundamental words during the stemming process. The Confix Stripping algorithm operates on rule-based principles for the Indonesian language, while Enhanced Confix Stripping represents a refined iteration of the former algorithm, incorporating an elevated approach. A succinct exposition of both algorithms will be presented within the context of their application in stemming. Ultimately, empirical findings reveal that Confix Stripping attains an accuracy rate of 65.62%, whereas ECS achieves a precision level of 66.58%. The central objective of this research is to function as a valuable point of reference for subsequent inquiries within the realm of text processing. By encapsulating explanations pertaining to Stemming, the intricacies presented by the diversified nature of the Indonesian language, the underpinnings of ECS and Confix Stripping, and the accurate results emanating from these two algorithms, this study seeks to provide a compass for forthcoming research endeavors that confront the emerging complexities stemming from language evolution and text manipulation.

Keywords – Stemming, Indonesian Language, Confix Stripping, Enhanced Confix Stripping, Text Preprocessing