

ABSTRAK

Stemming adalah proses di mana berbagai bentuk kata dipetakan ke bentuk dasarnya. Ini merupakan salah satu pendekatan dasar dalam pra-pemrosesan teks yang digunakan dalam Pemodelan Bahasa, Pemrosesan Bahasa Alami, dan aplikasi Pengambilan Informasi. Teknik ini sangat penting untuk meningkatkan akurasi dan efisiensi dalam berbagai aplikasi pengolahan teks.

Penelitian ini bertujuan untuk mengimplementasikan stemming pada bahasa Indonesia informal menggunakan algoritma Arifin dan Setiono, serta melakukan evaluasi terhadap tingkat akurasi hasil stemming tersebut. Algoritma Arifin dan Setiono dipilih karena relevansinya dalam konteks bahasa Indonesia, terutama dalam menangani kata-kata informal yang sering ditemukan dalam komunikasi sehari-hari, seperti di media sosial.

Dalam penelitian ini, digunakan 300 data tweet yang diambil dari *platform* media sosial Twitter yang mengandung kata-kata informal dalam bahasa Indonesia. Data ini akan diuji secara manual dengan memeriksa setiap kata untuk mendapatkan bentuk kata dasarnya. Hasil stemming kemudian akan dibandingkan dengan kata dasar yang diperoleh secara manual untuk menghitung tingkat akurasi algoritma yang digunakan.

Penelitian ini diharapkan dapat memberikan kontribusi signifikan dalam bidang pemrosesan bahasa alami, khususnya dalam pengembangan teknik stemming untuk bahasa Indonesia. Selain itu, hasil penelitian ini diharapkan dapat membantu para peneliti dan praktisi dalam memilih dan mengembangkan algoritma stemming yang lebih efektif untuk bahasa Indonesia informal.

Kata Kunci: *Stemming*, Algoritma Arifin dan Setiono

ABSTRACT

Stemming is a process in which various word forms are mapped to their base form. It is one of the fundamental approaches in text preprocessing used in Language Modeling, Natural Language Processing, and Information Retrieval applications. This technique is crucial for enhancing the accuracy and efficiency of various text processing applications.

This study aims to implement stemming for informal Indonesian language using the Arifin and Setiono algorithm, and to evaluate the accuracy of the stemming results. The Arifin and Setiono algorithm was chosen for its relevance in the context of the Indonesian language, especially in handling informal words commonly found in everyday communication, such as on social media.

In this research, 300 tweets containing informal Indonesian words were collected from the social media platform Twitter. The data will be manually examined by checking each word to obtain its base form. The stemming results will then be compared to the manually obtained base forms to calculate the accuracy of the algorithm used.

This research is expected to make a significant contribution to the field of natural language processing, particularly in the development of stemming techniques for the Indonesian language. Additionally, the results of this study are anticipated to assist researchers and practitioners in selecting and developing more effective stemming algorithms for informal Indonesian.

Keywords: Stemming, Arifin and Setiono Algorithm