ABSTRACT

PENCARIAN KATA TIDAK BAKU PADA DOKUMEN DI BALAI BAHASA BANDUNG MENGGUNAKAN *STEMMING ENGINE* BAHASA INDONESIA

By

JAJANG APRIANSYAH

10108188

Balai Bahasa Bandung *is an organization which works in surface of language and literature and it in shelther of* Pusat Bahasa. *One of the* Balai Bahasa *work program is collect scientific and print it to a book. To check scientific and distribution scientific, therefore need application which can helping checking document which contain not standard word. This problem can be solved by stemming engine*

Stemming is a process for separating affixes from words containing affixes and generate a basic word, there are two processes that must be done before to stemming there are tokenizing and filtering. Tokenizing is a process to cut the sentences into words, while filtering is a process to eliminate unimportant words.

Results of the study and after checking the stemming engine, the conclution is stemming engine could help checking process of the document contain not-standard word in Balai Bahasa Bandung, *it could help faster checking document and it could help distribution work of scientific to* Balai Bahasa Bandung.

Keywords: stemming, tokenizing, filtering, bahasa, Indonesia