Abstract

Lately, information has became very important thing for human. They can answer all kind of questions to solve their problem by information that can be in the form of text, picture, or voice.

They can use all kind of search engine like Google, Yahoo, Altavista, etc. to get that information. There are some process inside that search engine. One of those process is stemmning. Stemming is a process to return original form of word variants. In order to get best return hits, the stemmer process must be good.

There are many stemming algorithm that has been developed. One of them is Nazief & Adriani Algorithm that is Indonesian stemming. This algorithm is the best algorithm for Indonesian language structure. However, there are still some shortcomings in its implementation. One of them is leveling the meaning of the word variants.

Therefore, writer uses corpus-based stemmer which language independent to avoid those problem, including leveling the meaning of the word variants. Using statistics of word variants, this method can enhance accuracy of the Information Retrieval System.

Keywords : *Search Engine, Stemming, Algoritma Nazief Adriani, Corpus-Based Stemmer, Co-occurrence, Information Retrieval*