Abtract

Stemming is the process of removal of affixes on like prefixes, suffixes, infixes, and confixes performed using a specific algorithm to restore a word into its basic form. Stemming is part of information retrieval and use on search engines. By applying the stemming on the search engine, then it can obtain more specific information than the amount of information available. Stemming is the core of natural language processing techniques to obtain information retrieval effective and efficient.

In Indonesian, a word can be combined with a prefix, suffix, infixess, or confixes. This causes the difficulty of matching related words. Some of the merger may cause changes in the root. In the Indonesian language text processing, can be used stemming technique. Stemming in information retrieval systems are used to limit the different variants of the word into its basic form, so that later can increase the system's ability to find relevant documents according to the existing query. In this final paper, would implement stemming technique using confix-stripping algorithm in information retrieval systems.

Confix-stripping algorithm (CS) based on morphology Indonesian rule, which are grouped together and encapsulated in affixes, including prefixes, suffixes, infixes, and confixes. Confix-stripping algorithm (CS) using three components: a collection of affixes, rules, and a dictionary. In this final paper analysis will be conducted confix-stripping algorithm (CS) is the Information Retrieval system.

Keywords: information retrieval, stemming, confix-stripping