Abstract

Nowadays, text plagiarism become easier. To solve this problem, each text have to be checked whether the text made with plagiarism. This is a time consuming job if the amount of text is abundance. But this time consumption problem can be solved using a system that can find any similarity between texts, so the output of the system can be used in plagiarism detection.

This system can built by implementing Rabin-Karp algorithm. To find any similarity, two texts compared each others by consider each text as a string containing only 26 letters. The hash value of every substring from each text compared to know whether there are any words similarity.

By using this algorithm, the sistem can find any words that absolutely same and any words that change into others form (from noun to verb, adjective, and etc.) between two texts that containing similarity. Although the system can't find the change of vocabulary, the similarity that found by system is almost similar with the human done. Therefore, with this similarity, text plagiarism detection can be done faster.

Keywords: Rabin-Karp algorithm, string, substring, hash value.