Abstract

Plagiarism is the act of copying someone's creation and claim it as their own creation. Plagiarism of documents is hard to be avoid. Therefore, automatic detectors are very helpful to find the practices of plagiarism in documents such as MOSS, Tessy, JPlag, COpyCatch and others.

To make a document plagiarism detection, the point is with do some method which consists of two sequential processes that convert documents into a linear representation called a token string, then compare each token string with one another. The algorithm that used in this thesis is the Running Karp-Rabin Greedy String tiling (RKR-GST). This algorithm is able to find the parts that are identical in the two strings without affected the sequence and position of the substring. For each comparison of two documents, the similarity is calculated and parts of the document which allegedly is the result of the practice of plagiarism is marked.

Applications RKR-GST algorithm builds upon the basic principles using Java programming language. This application is able to find the practice of plagiarism and has been tested using a document collection of electronic sites, in addition to the tests was conducted using the training documents that have been modified.

For testing, similarity values should not be the final decision to determine cases of plagiarism because there is the possibility of incorrect detection results. Therefore, for the dubious plagiarism examiner should examine the contents of the document pairs before making a decision.

Keywords: : *document plagiarism, preprocessing, RKR-GST, string matching, scanpattern.*