ABSTRACT

In text plagiarism, the size and number of the texts can affect the performance of detection methods. Stemming from this problem, it is then proposed to use keyword extraction to produce a key text from each text. The key text will then be used to replace its source text in plagiarism detection.

In this thesis, implementation of keyword extraction in a plagiarism detection method and the design of a system based on said concept will be discussed. Keyword extraction concept that will be implemented is centrality measure, with LCS (longest common subsequence) as basic method. Three kind of centrality that will be discussed and implemented are degree centrality, closeness centrality, and betweenness centrality.

System is tested to find out whether the proposed system can produce better performance than basic method. Performace is measured by system accuracy, precision, and running time.

Testing result is analyzed to find the performance comparation between system and basic method. A conclusion will also be reached about which centrality concept that fit best for the detection system.

Keyword: Plagiarism indication, detection of similarity between texts, text comparison, keyword extraction, centrality, centrality measure, longest common subsequence (LCS).