

Abstract

Comparative Text Mining (CTM) is one of many techniques in text mining which peculiarly have the function: finding common theme from all collection and finding the special theme from a document. The use of CTM, for example, is for summarizing reviews. Summarization is an automatic process yielding shorter document version (50% or less) but remains useful for user. By summarization, user is expected to be able to catch the document's content without having to see the overall of document.

Clustering method is a method having ability to analyze and to group documents automatically. Generally, clustering technique using word and document is usually considered as word sets without the existence of sequence, called bag of word. Suffix Tree Clustering (STC) is the first algorithm that use phrase (multi-word terms) so that its process is simpler than other algorithm. STC is an incremental algorithm, the complexity of the algorithm is linear and fulfill the criterion for clustering web documents.

This final assignment aims to study and to prove the performance of STC's algorithm by applying it to CTM case. This experiment is done to see how parameters, which are background theme and bestRank range, influence the optimalization of the result.

Keywords: Comparative Text Mining (CTM), method of clustering, Suffix Tree Clustering (STC), and theme.