

Abstract

Currently, there are many different types of methods used for keyphrase extraction. There are two categories for keyphrase extraction; keyphrase extraction for single document and keyphrase extraction from a set of document.

Keyphrase extraction categories discussed in this final project is keyphrase extraction from a set of document. The method is CorePhrase. It is expected that by using this CorePhrase can get keyphrase which accordance with the topic and has a good performance.

Corephrase works by intersecting document using graph-based model then Features of the extracted candidate keyphrases are calculated and phrases are ranked based on their feature and then will get the top rank.

In this final project the performance of each keyphrase can be measured based on the overlap and precission. Based on the experiments conducted, the performance of the highest overlap is 1 and the performance of the highest precission is 1. It could be argued that the performance is good. The resulting keyphrase is in accordance with the topic of the document.

Keywords: *CorePhrase , ekstraksi, phrase feature, overlap, precission*