

Abstract

Document similarity detection and measurement which commonly used nowadays is term based similarity and measurement. Two documents that talk about different topic in the same environment can use same words. By using term based detection and measurement, the two documents can be judged as having same information. But when we consider the phrases used by each document we can see the difference and similarity of information between the two documents.

In this final task, an application has been developed to detect and measure the similarity among documents by comparing two documents at a time. Documents are represented using graph concept so the similarity can be detected and measured. In this final task, also done the combination similarity (combination of term based and phrase based similarity). Input documents are text document (*.txt) in bahasa Indonesia.

There are four steps in document similarity, they are preprocessing, graph construction, similarity detection, and similarity measurement.

The output of the application is similarity measurement between each input document.

Kata kunci: *Similarity*, Document, Graph, Bahasa Indonesia, txt