

ABSTRACT

One way to help to understand a concept is with finding a relation between one concept to another. Latent semantic analysis is one of the application of that where LSA can find a similarity value between concept. LSA is a way to extract text and create a statistical representation of the text with low dimensional approximation, where a small dimensions are used to determine a whole meaning of the document.

This final project tested the effectiveness of using latent semantic analysis method which is a process of semantic analysis on a short text like al-Quran to find the similarity value between verse. Beside that, LSA is dynamic, where there is no certain way to determine a proper dimension to use for every case, then it is necessary to know the impact of dimension choice to the similarity value.

Based on the test result, term-document matrix without LSA has a high accuracy of 85%, but based on the measurement of F-measure, its effectiveness is very low at 0% because it can not detect the verse with similar meaning. Compared to the measurement with the highest accuracy of LSA is 71% but the value of the F-measure reached 40% by LSA with another variable. Beside that, the dimensions that are used affect the results. If it is too low then the value of the similarity would be too uniform, biased and the results will not be optimal.

Keyword: *latent semantic analysis, similarity, al-Quran, matrix, singular value decomposition, cosine similarity*