Abstract

This study discusses about the analysis and implementation of a system that is used to answer the questions automatically the user needs based on the existing FAQ by measuring the similarity between one another. Then the answers of the questions in FAQ data that have the highest match value is assumed to answer user questions. The test is using the dataset SemEval-2016 Task 3. In the process, user queries and FAQ data will go through a preprocessing process that generates tokens and then a similarity measurement is taken using an overall similarity. Overall similarity is the result of a linear combination of matching values of semantic similarity and statistic similarity.

The test results of several scenarios indicate that the highest MAP value is 59.65%, with the other test results of each precision obtained by 44%, 100% recall value, F1-measure value of 61.11%, and Accuracy value of 71 , 43%.

Key words : *similarit, FAQ, overall similarity, semantic, accuracy, precision and recall.*