Abstract

Automatic evaluation for measuring the quality of machine translation output is used automatic metric. A metric will be understood as a measurement. Because of the quality of translation is subjective, therefore, the task for any metric is to assign score in such a way that they correlate with human judgement. The measure of evaluation for metric score is correlation with human judgement.

Machine Translation system evaluation using automatic metric can be faster, easier and cheaper compared with human evaluation. In this final project, implementation can be conducted by making a stemming module for Indonesian text integrated in METEOR metric tool that will be used for evaluating translation result from the machine. The next part of this final project is analyzing the influence of reference number used; the influence of implemented stemming towards score and correlation value resulted from the test. In addition, analysis on METEOR metric performance is also conducted with Indonesian stemming in evaluating several translation machines and comparing the correlation value with BLEU metric.

Testing result shows that reference total number used in evaluation can increase the score metric and the resulted correlation value. Stemming module gives the contribution to increase score metric but the influence towards correlation value is not significant. Compared with BLEU metric, METEOR metric is also able to show a higher correlation value rather than BLEU metric. While for the translation machine evaluation, Google Translator has the highest correlation value with 0.736, while Transtool has the second highest correlation value with 0.478 higher than Rekso Translator with 0.469 for its correlation value.

Keywords: *METEOR metric, stemming, score, correlation value*.