

## **Abstract**

This paper examines the calculation of the similarity between words in English using word representation techniques. Word2Vec is a model used in this paper to represent words into vector form. The model in this study was formed using the 320,000 articles in the English Wikipedia as the corpus and then Cosine Similarity calculation method is used to determine the similarity value. This model then tested by the test set gold standard WordSim-353 as many as 353 pairs of words and SimLex-999 as many as 999 pairs of words, which have been labelled with similarity values according to human judgment. Pearson Correlation was used to find out the accuracy of the correlation. The results of the correlation from this study are 0.665 for WordSim-353 and 0.284 for SimLex-999 using the Windows size 9 and 300 vector dimension configurations.