

RESULT DOCUMENT

The similarity between the translations performed by this translation engine and the actual translations are compared using two coefficients: Jaccard Coefficient and cosine similarity. The translation is performed on the "input.txt" file and the output is written in the "output.txt". The results in the output file are compared with the text in the file "actual.txt" (which is provided by the user as well). The lines in the two files are taken one at a time and the similarities are calculated between the corresponding lines. The similarities over all the lines are averaged and the result is recorded. The stop words are removed while calculating similarities to get a more accurate sense of the similarity scores.

The sample similarity scores mentioned here are for the following parameters:

Number of iterations = 15

Number of words considered in a sentence = 11

Training is performed on the first 1000 lines of the dataset.

Jaccard Coefficient = 0.149

Cosine Similarity = 0.222