Abstract

Plagiarism is an irresponsible act of misusing by citing part or all of the contents of other parties' documents that are recognized as their own documents, without stating the source correctly. To overcome this, each document must be examined whether the results of plagiarism. However, manual checks take a very long time if the number of documents examined too much. Therefore we need a system that can find similarities in these documents.

This system is built using the Rabin-Karp algorithm. The process of detecting document text similarity will go through the case folding, tokenizing, k-gram and hashing processes then using the Rabin-Karp Algorithm look for similarities and using Dice's Similarity Coeficient to calculate the similarity percentage. From the research conducted, the results obtained with F1-score were 62.29%.

Keywords: plagiarism, Rabin-Karp, string, hashing , k-gram