

Abstract

Making written work from a text document is so easy, that make plagiarism, which is should be avoided, is easier to conduct. Plagiarism in text document covers a very large scope. One of the indications is content similarity, to be more specific, the similarity of the words constructing text documents. In bioinformatics field, Smith-Waterman algorithm is used to identify significant similarity (local alignment) among gene sequences of living creature, by using sequence alignment method which is a way of arranging hydrogen base pair chain that construct the sequence with another sequence based on structure-similarity [14].

By applying the concept of Smith-Waterman algorithm, a sentence in text documents will be likened to a gene sequence that was formed by a series of words and sentences, and compared to the sentences in the other text documents. Previously, the text document will go through the stages of noise elimination. And then processed using Smith-Waterman algorithm resulting similarity percentage, indicating how similar the document to the other documents.

Keywords: plagiarism, text document, sequence alignment, Smith-Waterman algorithm.