

ABSTRACT

Plagiarism comes from the Latin language that is plagiarus, which means taking other people's work, form of writing, paper, or opinion, and so on, to be recognized as the work of the offender. This practice is particularly common in academic circles. To anticipate, we need a way to analyze the techniques of plagiarism committed. There are several approaches or techniques that can be taken, one of which is to matching string / term. For matching string / term there are several existing algorithms, one of which is the Rabin-Karp string matching algorithm. Rabin-Karp algorithm use hash function in string matching process. This hash function is the Rabin-Karp algorithm's key. Rabin-Karp algorithm is suitable to be applied in a document plagiarism detection, since the algorithm is suitable for multiple search patterns. To perform plagiarism detection, particularly documents written in Indonesia, the need for preprocessing process before a document is ready to enter into the core of the string matching process. In the preprocessing there are contains several steps, case folding, tokenization, number removal, filtering and the last is stemming or return to the original form (root). To conduct stemming steps there are several algorithms one of which is Nazief-Adriani stemming algorithm. Nazief-Adriani stemming algorithms have percentage accuracy (precision) is greater than the other stemming algorithms. Accuracy of the results stemming algorithm Nazief-Adriani strongly influenced the completeness of the root words dictionaries. The more complete dictionary of root words are the more accurate the stemming results. After preprocessing and string matching process, the last step is measure the similarity between documents.

Keywords : plagiarism, string matching, Rabin-Karp, hash function , preprocessing, case folding, tokenization, number removal, filtering, stemming, Nazief-Adriani, hashing, similarity, documents.