Abstract

Acts of plagiarism could have been done accidentally, chances are it may happen due to lack of information a person against another person writings which had been published in advance. Therefore, the author is responsible for checking whether or not plagiarism indicated his work before publication. The problem arises, do not allow for someone checked one by one to all the papers there. Therefore needed a system capable of automatically detecting indications of plagiarism. In this study, a system built using machine learning approach, using Bayesian network. To achieve this approach, it takes the form of feature extraction unigram, POS gram, and N-gram. The new features that handle cases of lexical and syntactic, which has an important role as a marker indicative of plagiarism. To measure the performance of the system is built, used F1-score. The system built is able to detect indications of plagiarism in the English language document, with the highest F1-score reached 0.987.

Keywords: Plagiarisme; Bayesian Network; Unigram; POS Gram; N-gram.