SIMILARITY ANALYSIS OF LEARNING JOURNAL USING TF-IDF ALGORITHM AND COSINE SIMILARITY (CASE STUDY D4-INFORMATION ENGINEERING, POLIJE)

Supervisor (1 People)

Gabriela Caroline Runturambi

Study Program of Informatics Engineering

Majoring in Information Technology

ABSTRACT

Plagiarism means plagiarizing another person's work and claiming that the work is his own. The government has given special attention to the problem of plagiarism. There is Regulation of the Minister of National Education of the Republic of Indonesia number 17 regarding the prevention and control of plagiarism cases in universities in 2010, and the Law of the Republic of Indonesia number 19 regarding copyright 2002. However, plagiarism cases are still widely carried out in Indonesia, especially what happens in the student group is doing the assignments given by the lecturers related to learning journals. One way to minimize the existence of plagiarizing is to compare learning journals among students by calculating the level of similarity.

This study uses the Term Frequency-Inverse Document Frequency (TF-IDF) and Cosine Similarity algorithm by using a dataset that opens 27 learning journal data, so that 1 document is compared to 26 other documents (1 : 26). The results obtained from these data are that the comparison of document 1 has a high level of similarity value to 3 other documents with similarity results above 50%, and a lowlevel similarity value with results below 50% in 23 other documents. The accuracy results obtained from the algorithm are 96.15%.

Keywords: *Plagiarism, Similarity, Learning Journal, Term Frequency-Inverse Document Frequency (TF-IDF), Cosine Similarity*