

ABSTRAK

Plagiarisme adalah tindakan menyerahkan (*submitting*) atau menyajikan (*presenting*) ide atau kata/kalimat orang lain tanpa menyebutkan sumbernya. Di Indonesia terdapat banyak sekali karya ilmiah yang beredar dari tahun ke tahun. Segala macam jenis karya ilmiah yang bisa dibaca atau dijadikan sebagai acuan untuk mengerjakan sebuah tugas yang beredar banyak di internet. Hal ini mengakibatkan jika semakin banyak web atau situs yang menyediakan karya ilmiah secara bebas maka menyebabkan peluang orang lain untuk melakukan plagiat. Untuk itu diperlukan cara-cara untuk mengatasi permasalahan tersebut. Penelitian ini bertujuan untuk membangun sistem yang secara otomatis mampu menghitung tingkat kemiripan antara jurnal berdasarkan isi dari jurnal tersebut. Data yang digunakan adalah jurnal berbahasa Indonesia dan akan diolah menggunakan metode *case folding*, *stopword*, *stemming*, *TF-IDF*, dan *cosine similarity*. Hasil penelitian ini adalah berupa akurasi tingkat kemiripan dari setiap jurnal yang dibandingkan. Dengan adanya penelitian ini diharapkan dapat membantu untuk menentukan nilai tingkat kemiripan dari sebuah jurnal dengan baik dan diharapkan bisa mengurangi kegiatan plagiarisme.

Kata kunci : *Cosine similarity, Information Retrieval, Plagiarisme, TF-IDF.*

ABSTRACT

Plagiarism is the act of submitting or presenting other people's ideas or words/sentences without citing the source. In Indonesia, there are a lot of scientific works circulating from year to year. All kinds of scientific works that can be read or used as a reference for doing an assignment are widely circulated on the internet. This results in more and more websites or sites that provide scientific work freely, it creates opportunities for others to do plagiarism. For that we need ways to overcome these problems. This study aims to build a system that is automatically able to calculate the level of similarity between journals based on the contents of the journal. The data used are Indonesian language journals and will be processed using case folding, stopword, stemming, TF-IDF, and cosine similarity methods. The results of this study are in the form of the accuracy of the similarity level of each journal being compared. With this research, it is hoped that it can help to determine the value of the similarity level of a journal properly and is expected to reduce plagiarism activities.

Keywords: *Cosine Similarity, Information Retrieval, Plagiarism, TF-IDF.*