

INTISARI

NAMED ENTITY RECOGNITION DAN DEPENDENCY PARSING UNTUK EKSTRAKSI KONSEP YANG LEBIH BAIK DALAM DETEKSI PLAGIARISME IDE

Oleh

UMAR TAUFIQ

17/420344/SPA/00613

Semakin banyak penelitian dikhususkan untuk mendeteksi plagiarisme ide atau gagasan dalam dokumen teks. Pada plagiarisme ide, ringkasan suatu dokumen teks disisipkan ke dalam dokumen teks lain sehingga lebih sulit untuk dideteksi dengan metode pendeteksian plagiarisme biasa. Metode yang dikembangkan dalam disertasi ini bertujuan untuk mengatasi masalah tersebut.

Disertasi ini mengembangkan pendekatan baru untuk deteksi plagiarisme ide berdasarkan *Named Entity Recognition* dan *Dependency Parsing*, yang lebih baik dibandingkan dengan metode berbasis algoritma genetika. Deteksi plagiarisme ide berbasis algoritme genetik masih dapat ditingkatkan performanya. *Named Entity Recognition* dan *Dependency Parsing* digunakan untuk mengekstraksi kemiripan konteks antar dokumen teks.

Metode yang dikembangkan berhasil mendeteksi plagiarisme ide di tingkat dokumen dengan akurasi yang lebih tinggi daripada metode berbasis algoritma genetika yang ada. Metode deteksi plagiarisme ide berbasis *Named Entity Recognition* dan *Dependency Parsing* juga dapat menghasilkan akurasi pendeteksian plagiarisme pada tingkat kalimat hingga lebih dari 84% untuk kasus *benchmark* dan *threshold* tertentu. Metode tersebut juga telah diuji pada jenis plagiarisme lain, dan menghasilkan performa yang sangat baik. Penelitian dalam disertasi ini telah menghasilkan metode deteksi plagiarisme ide yang lebih akurat dibandingkan dengan metode berbasis algoritme genetika.

Kata-kata kunci : Deteksi plagiarisme ide, Named Entity Recognition, Dependency Parsing, jangkak informasi, Cosine Similarity.

ABSTRACT

NAMED ENTITY RECOGNITION AND DEPENDENCY PARSING FOR BETTER CONCEPT EXTRACTION IN IDEA PLAGIARISM DETECTION

By

UMAR TAUFIQ

17/420344/SPA/00613

More and more research is devoted to detecting plagiarism of ideas or ideas in text documents. In idea plagiarism, a summary of a text document is inserted into another text document making it more difficult to detect with ordinary plagiarism detection methods. The method developed in this dissertation aims to overcome this problem.

This dissertation develops a new approach to detecting plagiarism of ideas based on *Named Entity Recognition* and *Dependency Parsing*, which is better than methods based on genetic algorithms. Genetic algorithm-based idea plagiarism detection can still be improved. *Named Entity Recognition* and *Dependency Parsing* are used to extract contextual similarities between text documents.

The developed method successfully detects idea plagiarism at the document level with higher accuracy than existing genetic algorithm based methods. Idea plagiarism detection methods based on *Named Entity Recognition* and *Dependency Parsing* can also produce plagiarism detection accuracy at the sentence level of up to more than 84% for certain benchmark and threshold cases. The method has also been tested against other types of plagiarism, and it performs very well. The research in this dissertation has produced a method for detecting plagiarism of ideas that is more accurate than methods based on genetic algorithms.

Keywords : Idea plagiarism detection, Named Entity Recognition, Dependency Parsing, information anchors, Cosine Similarity.