

ABSTRAK

Adi Sparta, I Wayan (2019), *Penerapan Metode Levenshtein Distance Berbasis N-gram untuk Meningkatkan Akurasi pada Pengembangan Rest Web Service Deteksi Plagiarisme*. Tesis, Ilmu Komputer, Program Pascasarjana, Universitas Pendidikan Ganesha.

Tesis ini sudah disetujui dan diperiksa oleh Pembimbing I : Dr. Drs. I Nyoman Sukajaya, M.T dan Pembimbing II : Dr. I Gede Aris Gunadi, S.Si. M.Kom.

Kata-kata kunci : persentase deteksi plagiat, *Levenshtein Distance*, *jaro-winkler distance*, *n-gram*.

Metode *Levenshtein Distance* merupakan salah satu metode yang dapat dipergunakan untuk melakukan pendeteksian plagiarisme. Metode ini memiliki kekurangan pada waktu pemrosesannya yang tergolong lama. Penelitian ini bertujuan untuk mempercepat proses yang dilakukan pada metode *Levenshtein Distance* untuk mendeteksi plagiarisme. Ide yang digunakan sebagai upaya dalam mempercepat proses yang dilakukan pada metode *Levenshtein Distance* adalah memadukannya dengan teknik *n-gram*. Keberhasilan dari penerapan algoritma diuji melalui artikel yang diambil pada E-Journal Universitas Pendidikan Ganesha (Undiksha). Tiap artikel dipisah antara abstrak dan isinya. Abstrak tiap artikel akan dipergunakan untuk pengujian teks dengan ukuran sedang, sedangkan isi artikel digunakan untuk pengujian teks ukuran panjang. Selain itu, penelitian ini juga menggunakan kata-kata kerja dalam bahasa Indonesia untuk uji teks pendek. Metode yang digunakan dalam pengembangan sistem ini adalah dengan menggunakan metode *Research and Development* dengan model *Rapid Application Development* (RAD). Hasil penelitian menunjukkan bahwa akurasi *Levenshtein Distance* berbasis *n-gram* yang dihitung dengan metode *Mean Absolute Percentage Error* (MAPE), ternyata memberikan nilai *error* yang lebih tinggi dibandingkan *Levenshtein Distance*. Namun implementasi *n-gram* pada *Levenshtein Distance* berhasil mempercepat proses *Levenshtein Distance*. Rata-rata kecepatan pada pengujian abstrak artikel menggunakan *Levenshtein Distance* membutuhkan waktu 0.0258764 detik sedangkan pada *Levenshtein 2* dan *3 gram* berturut-turut menghasilkan nilai 0.022878 dan 0.019559 detik. Perbedaan yang cukup besar diperoleh pada pengujian pada isi artikel. Pada *Levenshtein Distance* membutuhkan rata-rata waktu 19.642312 detik sedangkan pada *Levenshtein 2 gram* membutuhkan 15.21176 detik diikuti *Levenshtein 3 gram* dengan waktu 14.64673 detik. Berdasarkan hasil penelitian ini, dapat menjadi referensi untuk pengembangan metode untuk pendeteksian plagiarisme selanjutnya.

ABSTRACT

Adi Sparta, I Wayan (2019), *Application of Levenshtein Distance Method Based on N-gram to Increase Accuracy in Development of Rest Web Service Detection of Plagiarism*. Thesis. Post Graduate Program. Ganesha University of Education.

This thesis has been checked and approved by Supervisor I: Dr. Drs. I Nyoman Sukajaya, M.T and supervisor II: Dr. I Gede Aris Gunadi, S.Si. M.Kom.

Keywords : the percentage of plagiarisme detected, Levenshtein *Distance*, *jaro-winkler distance*, *n-gram*.

The Levenshtein Distance method is one method that can be used to detect plagiarism. This method has shortcomings when processing is quite old. This study aims to accelerate the process carried out on the method of Levenshtein Distance to detect plagiarism. The idea used as an effort to speed up the process carried out by the Levenshtein Distance method is to combine it with the n-gram technique. The success of the implementation of the algorithm was tested through articles taken in the E-Journal of the University of Education Ganesha (Undiksha). Each article is separated between the abstract and its contents. Abstract each article will be used for testing medium-sized text, while article content is used for testing long-size text. In addition, this study also uses verbs in Indonesian to test short texts. The method used in developing this system is to use the Research and Development method with the Rapid Application Development (RAD) model. The results showed that the accuracy of Levenshtein Distance based on n-gram calculated using the Mean Absolute Percentage Error (MAPE) method, turned out to provide a higher error value than Levenshtein Distance. But the n-gram implementation on Levenshtein Distance managed to speed up the process of Levenshtein Distance. The average speed in testing the abstract of the article using Levenshtein distance requires 0.0258764 seconds while in Levenshtein 2 and 3 grams it produces 0.022878 and 0.019559 seconds respectively. Considerable differences were obtained in testing the contents of the article. In Levenshtein Distance requires an average time of 19.642312 seconds while in Levenshtein 2 grams requires 15.21176 seconds followed by Levenshtein 3 grams with a time of 14.64673 seconds. Based on the results of this study, it can be a reference for the development of methods for subsequent plagiarism detection.