

INTISARI

PERBANDINGAN DISTANCE BASED SIMILARITY MEASURE PADA ALGORITMA RABIN KARP UNTUK MENGHITUNG KEMIRIPAN TEKS BAHASA INDONESIA

Oleh

Siti Hardiyanti
14/363523/PA/15846

Metode untuk mengukur kemiripan teks dapat digunakan untuk pengelompokan dokumen, penilaian tes objektif, dan deteksi plagiarisme. Salah satu metode untuk mendeteksi kemiripan teks adalah string matching. Metode string matching yang digunakan pada penelitian ini adalah Rabin Karp.

Tahap pengukuran similarity teks pada Algoritma Rabin Karp menggunakan distance based similarity measure. Penelitian ini membandingkan beberapa metode distance based similarity measure yang diimplementasikan bersama dengan algoritma Rabin-Karp, yaitu Dice's coefficient, Cosine similarity, dan Jaccard coefficient.

Hasil penelitian menunjukkan bahwa dari ketiga distance based similarity measure yang digunakan, Cosine Similarity memiliki akurasi paling tinggi yaitu sebesar 83,2% dengan k-gram = 2.

Kata Kunci: *Dice's Coefficient, Cosine Similarity, Jaccard Coefficient, Rabin Karp.*

ABSTRACT

THE COMPARISON OF DISTANCE BASED SIMILARITY MEASURE IN RABIN-KARP ALGORITHM TO CALCULATE TEXT SIMILARITY

By

Siti Hardiyanti

14/363523/PA/15846

Methods for measuring text similarity can be use for document clustering, short answer grading, and plagiarism detection. One of method for detecting text similarity is string matching. The current string matching method is Rabin Karp.

The measurement stage of text similarity in Rabin Karp Algorithm uses distance based similarity measure. This research uses some distance based similarity measure which is implemented along with Rabin-Karp algorithm, that is Dice coefficient, Cosine similarity, and Jaccard coefficient.

The results showed that from three different distance based similarity measure, Cosine Similarity was the highest accuracy of 83.2% with k-gram = 2.

Keyword: *Dice's Coefficient, Cosine Similarity, Jaccard Coefficient, Rabin Karp.*