

## **INTISARI**

### **PENGUJIAN SIMILARITAS BERITA MENGGUNAKAN METODE RABIN-KARP**

Oleh  
Rinto Achmad Sulchananto  
09/283670/PA/12663

Perkembangan peminat berita berbasis online didorong oleh dua hal, yang pertama adalah semakin mudahnya masyarakat Indonesia untuk memperoleh internet terutama secara *broadband wireless*. Yang kedua adalah peningkatan jumlah pengguna smartphone di Indonesia. Dengan bertambahnya peminat berita *online*, maka bermunculan situs berita dan informasi yang tidak mementingkan kualitas isi berita. Sehingga diperlukan sistem yang dapat digunakan tolok ukur untuk menentukan similaritas berita. Metode yang dapat digunakan adalah menggunakan Rabin-Karp. Rabin-Karp unggul dalam pencarian string dengan pola yang panjang.

Metode Rabin-Karp dalam sistem ini memiliki langkah-langkah preprosesing, penghitungan k-gram menggunakan proses parsing, *hashing* serta menghitung similiaritas menggunakan *dice coefficient*. Dalam preprosesing dilakukan *stemming* menggunakan Algoritma Nazief-Adriani. Pada sistem ini dilakukan modifikasi yakni ketika dokumen akan di parsing akan dilakukan pengurutan kata terlebih dahulu dan sistem memiliki nilai k-gram awal 4. Hasil yang didapatkan dari penelitian ini adalah, sistem dapat menentukan tingkat similaritas berita dengan rata-rata tingkat kesalahan 6,13%.

**Kata Kunci :** algoritma Rabin-Karp, similaritas berita, *stemming*, *dice coefficient*

## **ABSTRACT**

### **NEWS SIMILARITY TEST USING RABIN-KARP METHOD**

by

Rinto Achmad Sulchananto

09/283670/PA/12663

The development of online-based news enthusiasts is driven by two things, the first is the easier the Indonesia people to obtain internet especially using broadband wireless. The second is the increasing number of smartphone users in Indonesia. The presence of smartphones make it easier for users to get news information anywhere. With the addition of online news enthusiasts, the emergence of news and information sites that do not care about the quality of news content. Many news sites often just copy pastes from other sites. So we need a system that can be used benchmarks to determine the news similarity. The method that can be used is using Rabin-Karp. Rabin-Karp excels in the search string with a long pattern.

The Rabin-Karp method in this system has preprocessing, hashing and calculating similarity steps using dice coefficient. In preprocessing is done with stemming using Nazief-Adriani Algorithm. There is modifications in this system parsing process, there are words sorting and the system has an initial k-gram value 4. The results obtained from this research is, the system can determine the level of news similarity with the average error rate of 6,13%.

**Keywords:** Rabin-Karp algorithm, news similarity, stemming, dice coefficient