

Deteksi Kemiripan Halaman pada Al-Qur'an dengan Menggunakan Algoritma Rabin Karp dan Jaccard Similarity

Winda Eka Samodra¹, Moch Arif Bijaksana².

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

¹samoded@student.telkomuniversity.ac.id, ²arifbijaksana@telkomuniversity.ac.id,

Abstrak

Plagiarisme merupakan suatu tindakan yang dilakukan untuk mendapatkan pengakuan secara ilmiah tanpa memberikan sumber asli. Di lain sisi plagiarisme dapat digunakan untuk mencari *similarity* antara satu dokumen dengan dokumen yang lain. Cukup sulit mengukur kesamaan dokumen hanya dengan membaca atau mengukurnya secara manual. Namun konsep plagiarisme ini dapat digunakan untuk mendeteksi kemiripan antar ayat pada Al-Qur'an. Dataset yang digunakan merupakan sekumpulan ayat pada halaman Al-Qur'an. Oleh karena itu, dibutuhkan suatu sistem yang dapat melakukan deteksi kemiripan. Untuk membangun sistem tersebut, pada tugas akhir ini menggunakan metode *Rabin Karp* dengan diolah menggunakan parsing *N-gram* pada setiap dokumen inputnya. Data akan dicari nilai *similarity* diantara ayat kesatu pada halaman pertama dengan semua ayat pada halaman kedua. Selanjutnya akan diambil nilai kemiripan yang bernilai 1 dan 0. Sistem yang dibangun menggunakan salah satu metode kemiripan yaitu, *Jaccard Similarity*. Dengan menggunakan teknik hashing dan menggunakan Algoritma *Rabin Karp* sistem dapat menghasilkan nilai *precision* terbaik sebesar 32,76 % dan menghasilkan *f-Measure* sebesar 39,81%.

Kata kunci : Plagiarisme, Al-Quran, Rabin Karp, Jaccard Similarity