

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil yang didapat setelah melakukan penelitian ini, dengan mengacu pada rumusan masalah yang telah diuraikan pada pendahuluan. Kesimpulan yang diperoleh dari penelitian ini adalah sebagai berikut:

1. Berdasarkan penelitian yang dilakukan, algoritma Rabin-Karp diterapkan pada pendeteksian similarity dokumen dengan menggunakan proses *hashing*. Sebelum memasuki proses *hashing*, dokumen input dan dokumen pada database akan di generate ke dalam bentuk string, lalu dibersihkan pada tahap praproses. Selanjutnya dikelompokkan menjadi kumpulan kata dengan berdasarkan n-gram. N-gram dijadikan parameter algoritma Rabin-Karp untuk membentuk kelompok kata yang akan di *generate* ke dalam nilai hash. Selanjutnya nilai hash dari kedua dokumen dibandingkan dan didapat similarity nya. Dengan melihat perbandingan hasil evaluasi dari berbagai skenario eksperimen yang telah dilakukan, dapat disimpulkan bahwa semakin besar n-gram yang digunakan maka semakin tinggi nilai error persentase similarity nya. Dari hasil evaluasi juga dapat disimpulkan bahwa algoritma Rabin-Karp tidak dapat mengatasi plagiat dengan teknik modifikasi kata dengan kata sinonimnya, karena pada algoritma Rabin-Karp nilai hash akan berbeda jika kata tersebut dimodifikasi.
2. Dengan melihat perbandingan hasil evaluasi dari berbagai skenario eksperimen yang telah dilakukan, nilai error rata-rata pada skenario tanpa stopword removal dan stemming yaitu sebesar 18.36% serta nilai error rata-rata yang dihasilkan oleh skenario menggunakan stopword removal dan

stemming yaitu sebesar 17.97%. Sedangkan pada skenario dengan menggunakan data hapus kata, nilai error rata-rata yang dihasilkan yaitu

10.17%. Nilai error rata-rata pada skenario menggunakan data tambah kata yaitu sebesar 10.01%. Nilai error rata-rata pada skenario menggunakan data modifikasi kata yaitu sebesar 23.83%. Nilai error rata-rata pada skenario menggunakan data hapus modifikasi yaitu sebesar 26.65%. Dan nilai error rata-rata pada skenario menggunakan data tambah modifikasi yaitu sebesar 22.75%.

5.2 Saran

Berdasarkan pembahasan dan hasil penelitian yang telah dilakukan, saran untuk penelitian selanjutnya adalah :

1. Input dokumen skripsi sebaiknya dapat dalam berbagai format seperti .doc, .docx, .txt, tidak hanya dalam format .pdf.
2. Eksperimen dapat dilakukan dengan data dokumen yang lebih banyak agar hasil yang didapat lebih banyak dan agar analisa yang dilakukan dapat lebih baik.
3. Sebaiknya ditambahkan metode yang dapat mengatasi masalah modifikasi dengan kata sinonim agar mendapatkan hasil yang lebih optimal.
4. Agar hasil lebih valid analisis pendeteksian plagiat sebaiknya melibatkan ahli bahasa.
5. Penulis berharap sistem ini dapat dikembangkan jauh lebih baik lagi.