



Hak cipta dan penggunaan kembali:

Lisensi ini mengizinkan setiap orang untuk menggubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

Copyright and reuse:

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

BAB V

SIMPULAN DAN SARAN

5.1 Simpulan

Berdasarkan hasil uji coba yang sudah dilakukan, didapatkan beberapa simpulan. Berikut ini adalah simpulan-simpulan tersebut.

1. Aplikasi yang dapat mendeteksi plagiarisme pada dokumen teks berbahasa Indonesia dapat diimplementasikan.
2. Algoritma Rabin-Karp dapat diimplementasikan untuk mendeteksi plagiarisme dengan memakai konsep *text pre-processing*, *n-grams* dan *Dice's coefficient*.
3. Modifikasi terhadap algoritma Rabin-Karp memiliki pengaruh terhadap waktu proses dan nilai *similarity*. Waktu proses yang dibutuhkan lebih cepat. Ini disebabkan karena algoritma Rabin-Karp modifikasi tidak lagi melakukan pengecekan karakter pada teks. Selain itu, nilai *similarity* yang dihasilkan lebih tinggi dari algoritma Rabin-Karp tanpa modifikasi dan tidak terlalu signifikan. Namun, nilai *similarity* menjauh dari akurat. Adanya perbedaan nilai *similarity* disebabkan oleh dua faktor, yaitu panjang *n-grams* dan pemilihan *hash function*. Panjang *n-grams* akan menghasilkan banyak kombinasi karakter dari suatu *substring*. Kombinasi karakter yang banyak ini memungkinkan adanya beberapa *substring* yang memiliki nilai *hash* yang sama. Oleh karena itu, pemilihan *hash function* harus baik sehingga nilai *hash* dari setiap *substring* tidak ada yang sama dan tidak memengaruhi nilai *similarity* yang dihasilkan algoritma Rabin-Karp modifikasi.

4. Nilai *n-grams* memengaruhi waktu proses dan nilai *similarity*. Semakin besar nilai *n-grams*, waktu proses yang dibutuhkan semakin cepat dan nilai *similarity* yang dihasilkan semakin akurat. Ini disebabkan karena semakin kecil nilai *n-grams*, *substring* yang terbentuk semakin banyak. Dengan banyaknya jumlah *substring*, waktu yang dibutuhkan untuk membandingkan masing-masing *substring* akan lebih lama. Di samping itu, penggunaan rumus *Dice's coefficient* menyebabkan beberapa *substring* yang sama persis akan dianggap menjadi satu *substring* sehingga jumlah *substring* akan berkurang. Semakin kecil nilai *n-grams*, semakin tinggi kemungkinan munculnya *substring* yang sama sehingga pengurangan jumlah anggota *substring* menjadi lebih signifikan. Ini akan menghasilkan nilai *similarity* yang tidak akurat.

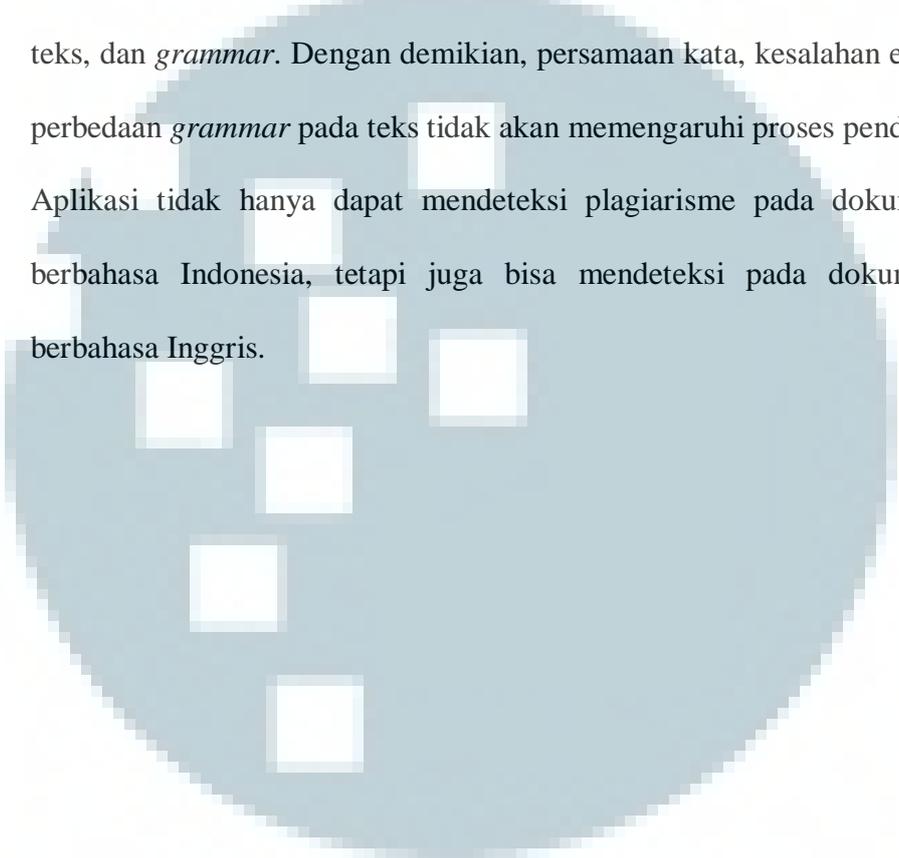
5.2 Saran

Berikut ini adalah beberapa saran yang diajukan oleh penulis terhadap penelitian yang telah dilakukan untuk penelitian dan pengembangan aplikasi selanjutnya.

1. Pada saat menghitung nilai *similarity*, seharusnya *substring* yang sama tidak dihilangkan sehingga hasil perhitungan dapat lebih akurat. Terlihat pada saat *n-grams* bernilai satu, nilai *similarity* yang dihasilkan sama sekali tidak akurat.
2. *Hash function* yang digunakan harus lebih kompleks dan cepat dieksekusi agar nilai *similarity* yang dihasilkan algoritma Rabin-Karp modifikasi tidak

berbeda dari algoritma Rabin-Karp tanpa modifikasi dan waktu perhitungan *hash value* tidak terlalu lama.

3. Aplikasi dapat memperhatikan sinonim, kesalahan ejaan yang ada dalam teks, dan *grammar*. Dengan demikian, persamaan kata, kesalahan ejaan, dan perbedaan *grammar* pada teks tidak akan memengaruhi proses pendeteksian.
4. Aplikasi tidak hanya dapat mendeteksi plagiarisme pada dokumen teks berbahasa Indonesia, tetapi juga bisa mendeteksi pada dokumen teks berbahasa Inggris.



UMN