

IMPLEMENTASI ALGORITMA WINNOWER PADA DETEKSI PENJIPLAKAN

FAWAIDIL ABROR, 12 1065 1093
PROGRAM STUDI TEKNIK INFORMATIKA
UNIVERSITAS MUHAMMADIYAH JEMBER, 2016

Abstrak

Dari penelitian ada sekitar 89% mahasiswa yang setuju dan mengerti arti dari plagiarisme dalam dunia pendidikan, dan menyarankan agar materi tentang hal itu dijelaskan pada awal perkuliahan. Tetapi sekitar 65% mengaku bahwa mereka merasa bingung dengan pengertian plagiat, 59% diantaranya menyatakan bahwa mereka tidak diberi tutorial yang cukup untuk menghindari tindak plagiat dalam menyelesaikan tugas mereka.

Algoritma *Winnower* merupakan algoritma yang digunakan untuk deteksi plagiarisme. *Input* dari algoritma ini adalah dokumen teks yang diproses sehingga menghasilkan *output* berupa kumpulan nilai-nilai *hash*, nilai *hash* merupakan nilai numerik yang terbentuk dari perhitungan *ASCII* tiap karakter. Kumpulan-kumpulan nilai *hash* tersebut selanjutnya disebut *fingerprint*.

Nilai basis, *gram* dan panjang *window* sangat berpengaruh dalam menghitung *similarity* kesamaan dokumen, jika semakin besar *gram* maka semakin kecil kesamaan dokumen.

Kata Kunci :Plagiarisme, Algoritma *Winnower*, nilai-nilai *hash*, *fingerprint*, Nilai basis.

Abstract

From research there is about 89% of students who agree to and understand the meaning of plagiarism in education, and suggested that the material on it was described at the beginning of the term. But about 65% admitted that they are confused by the notion of plagiarism, 59% of them stated that they were not given enough tutorial to avoid acts of plagiarism in completing their tasks.

Winnower algorithm is an algorithm used to detect plagiarism. The input of this algorithm is a text document that is processed to produce output in the form of a collection of hash values, hash value is formed from the numerical value of each character ASCII calculation. Multitudes hash value is subsequently called a fingerprint.

Value basis, grams and length of window is very influential in calculating the document similarity similarity, if greater, the smaller gram similarity document.

Keywords: Plagiarism, Winnower algorithm, hash values, fingerprint, Value basis.

1. Latar Belakang

Penjiplakan atau plagiarisme berarti mencontoh atau meniru atau mencuri tulisan dan karya orang lain yang kemudian diakui sebagai karangannya sendiri dengan atau pun tanpa seizin penulisnya. Penjiplakan dokumen digital bukanlah hal yang susah, cukup dengan menggunakan teknik *copy paste-modify* pada sebagian isi dokumen dan bahkan keseluruhan isi dokumen sudah bisa dikatakan bahwa dokumen tersebut merupakan hasil duplikasi dari dokumen lain (Agus, 2015).

Winnowing menggunakan *window* sebagai metodenya, yaitu pembentukan *window* setelah proses *hashing*. Nilai *hash* sendiri merupakan nilai numerik yang terbentuk dari perhitungan *ASCII* tiap karakter. Setelah pembentukan *window* yang berisi nilai-nilai *hash*, maka dipilih nilai *hash* yang paling kecil dari setiap *window*. Jika ada lebih dari satu nilai yang paling kecil di proses *Winnowing*, maka nilai yang terkecil pada *window* sebelumnya itu dipilih untuk dijadikan *document's fingerprints*. *Fingerprints* inilah yang akan dijadikan dasar pembandingan antar file-file teks yang telah dimasukkan, sehingga dapat diketahui berapa persentasi kemiripan dua dokumen tersebut.

Algoritma *winnowing* lebih cepat waktu komputasinya dari pada algoritma *rabin-karp* karena *fingerprint* dari algoritma *rabin-karp* lebih banyak sedangkan *winnowing* dipilih nilai minimum dari *window*-nya, algoritma *winnowing* dipilih karena algoritma *winnowing* dapat memberikan informasi posisi *fingerprint*. Penggunaan nilai *window* yang semakin besar dapat mempengaruhi dalam waktu proses, karena semakin besar nilai dari suatu *window* mempengaruhi proses pembentukan nilai-nilai *hashing* yang semakin kecil. Selain itu, nilai *k-gram* dan basis juga mempengaruhi dalam memberikan persentasi kemiripan yang diperoleh sistem itu sendiri (Kusmawan, 2010).

2. Algoritma Winnowing

Algoritma *Winnowing* merupakan algoritma yang digunakan untuk deteksi

plagiarisme. *Input* dari algoritma ini adalah dokumen teks yang diproses sehingga menghasilkan *output* berupa kumpulan nilai-nilai *hash*, nilai *hash* merupakan nilai numerik yang terbentuk dari perhitungan *ASCII* tiap karakter. Kumpulan-kumpulan nilai *hash* tersebut selanjutnya disebut *fingerprint* (Adi, dkk, 2011). *Fingerprint* inilah yang digunakan dalam deteksi plagiarisme. Langkah awal dalam penerapan algoritma *Winnowing* adalah membuang karakter-karakter dari isi dokumen yang tidak relevan misal tanda baca spasi dan simbol lain. Langkah kedua isi dokumen yang telah dilakukan pembersihan selanjutnya dilakukan pembentukan rangkaian *gram*, pada $n=5$. Langkah ketiga dari rangkaian *gram* yang telah terbentuk dibentuk nilai *hash* dengan nilai *ASCII* tiap karakter. Pembentukan nilai *hash* menggunakan persamaan *rolling hash*. Langkah keempat nilai-nilai *hash* yang telah terbentuk dibentuk *window-window*. Dari *window-window* yang telah terbentuk dilakukan pemilihan nilai *hash* terkecil pada tiap *window* untuk dijadikan *fingerprint* tiap dokumen. Nilai-nilai *fingerprint* inilah yang digunakan untuk menemukan tingkat persentase kesamaan sebuah dokumen dengan dokumen lain. *Input* dari algoritma ini adalah *string* dari dokumen tersebut, dan *outputnya* berupa nilai-nilai *hash* yang dinamakan *fingerprints* dari dokumen tersebut.

Banyak algoritma atau metode yang dapat digunakan untuk mendeteksi kemiripan dokumen teks. Ada beberapa persyaratan yang harus dipenuhi oleh algoritma pendeteksi kemiripan dokumen teks (Pratama, 2012), *Whitespace Insensitivity* artinya dalam melakukan pendeteksian terhadap dokumen teks, algoritma tidak boleh dipengaruhi oleh spasi, jenis huruf (kapital atau normal), tanda baca dan sebagainya. Oleh sebab itu, dilakukan penghapusan terhadap karakter yang tidak relevan tersebut sehingga nantinya hanya akan tersisa satu rangkaian yang terdiri atas huruf dan angka.

$$C_1 * b^{(k-1)} + C_2 * b^{(k-2)} + \dots + C_{(k-1)} * b^k + C_k \dots \dots \dots$$

Keterangan :

- c : nilai ASCII karakter
- b : basis (bilangan prima)
- k : banyak karakter

Algoritma *winning* lebih cepat waktu komputasinya dari pada algoritma *rabin-karp* karena *fingerprint* dari algoritma *rabin-karp* lebih banyak sedangkan *winning* dipilih nilai minimum dari *window*-nya. Keunggulan algoritma *winning* bisa memberikan informasi posisi *fingerprint*. Penggunaan nilai *window* yang semakin besar dapat mempengaruhi dalam waktu proses, karena semakin besar nilai dari suatu *window* mempengaruhi proses pembentukan nilai-nilai *hashing* yang semakin kecil. Selain itu, nilai *k-gram* dan basis juga mempengaruhi dalam memberikan persentasi kemiripan yang diperoleh sistem itu sendiri (Kusmawan, 2010).

3. Deteksi Penjiplakan

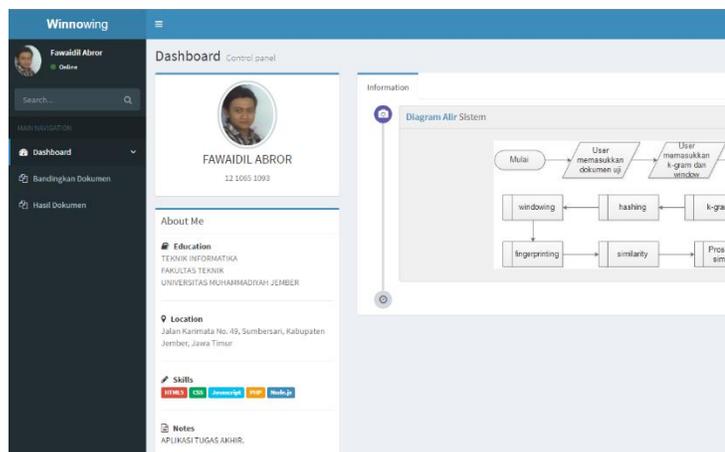
Penjiplakan menurut Kamus Besar Bahasa Indonesia berarti menggambar atau menulis garis-garis gambaran atau tulisan yang telah tersedia (dengan menempelkan kertas kosong pada gambar atau tulisan yang akan ditiru), mencontoh atau meniru tulisan atau pekerjaan orang lain, mencuri karangan orang lain dan mengakui sebagai karangan sendiri, mengutip karangan orang lain tanpa seizin penulisnya (www.pusatbahasa.diknas.go.id diakses april 2016).

4. Implementasi Sistem

Implementasi sistem merupakan tahap lanjutan dari perancangan sistem. Dalam tahap ini, mengimplementasikan hasil analisis dan perancangan sistem kedalam suatu bahasa pemrograman. Dengan adanya implementasi sistem, akan lebih terlihatnya hasil akhir dari perancangan sistem yang sudah di analisis sebelumnya.

4.1. Tampilan halaman form dashboard (awal)

Tampilan form dashboard merupakan tampilan form yang pertama kali muncul saat sistem dijalankan. Tampilan form dashboard dapat dilihat pada Gambar 4.1.

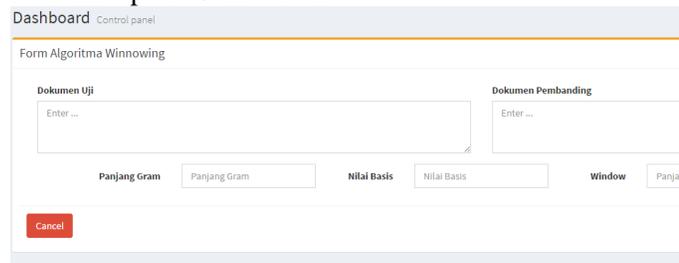


Gambar 4.1. Form Dashboard

Padahalaman form ini, menampilkan data-data dari nama, nomor induk, serta judul dari penelitian ini.

4.2. Tampilan Form Bandingkan Dokumen

Pada tampilan ini menampilkan form inputan yang akan dihitung tingkat kemiripan dengan algoritma *winning*, dapat di lihat pada Gambar 4.2.



Gambar 4.2. Form Banding Dokumen

Pada tampilan form banding dokumen terdapat text dokumen uji, dokumen pembanding, panjang gram, nilai basis, window, tombol cancel dan tombol proses yang memuat isi dokumen yang akan di proses.

5. Kesimpulan

Berdasarkan pembahasan dan evaluasi dari bab terdahulu, maka dapat ditarik kesimpulan sebagai berikut:

1. Algoritma *winning* dapat digunakan dalam menghitung tingkat kemiripan suatu dokumen dengan yang lain.
2. Nilai basis, gram dan panjang window sangat berpengaruh dalam menghitung similarity kesamaan dokumen, jika

semakin besar gram maka semakin kecil kesamaan dokumen.

6. Saran

Berikut adalah beberapa saran untuk pengembangan lebih lanjut terhadap penelitian skripsi ini:

1. Dapat ditambahkan stemmer untuk memperoleh kata dasar sebelum diproses ketahap *white space insensitivity* sehingga dapat memberi tingkat kemiripan yang lebih akurat.
2. Sistem ini dapat dikembangkan lebih lanjut dengan menfokuskan pada mengembangkan fitur-fitur yang ada dikarenakan sistem yang dibuat lebih menfokuskan pada penghitungan kemiripan satu file dengan yang lain menggunakan dua algoritma.

DAFTAR PUSTAKA

- Adi, P,A,P.Kurniawan, U, U. Brotoharsono, T. 2011. Aplikasi Wireless User Positioning Dengan Metode Location Fingerprinting Dan Inverse Distance Weighted. Bandung. Universitas Telkom.
- Hernawati, K. 2010. Implementasi Cipher Viginere pada kode ASCII dengan Memanfaatkan Digit Desimal Bilangan Phi. Jurnal FMIPA Universitas Negeri. Yogyakarta
- Kusmawan, P. Y., Yuhana, U. L., & Purwitasari, D. (2010). Aplikasi Pendeteksi Penjiplakan pada File Teks dengan Algoritma Winnowing. ITS Under graduate- 3100010038856.
- Schleimer, S., Wilkerson, D., dan Aiken, A. 2003. Winnowing: Local algorithms for document fingerprinting. In Proceedings of the ACM SIG MOD international conference on management of data.
- Syahputra, R, A. 2015. Implementasi Algoritma Winnowing Untuk Deteksi Kemiripan Text. Medan. Teknik Informatika STMIK Budi darma Medan.
- Thompsett, A., Ahluwalia, J. 2010. Students Turned Off by Turnitin? Perception of Plagiarism and Collusion by Under graduate Bioscience Students. School of Health and Biosciences, University of East London. Volume 16: December 2010.
- Yuwono, Putu., Yuhana, UmiLaili., dan Purwitasari, Diana. 2010. Aplikasi deteksi penjiplakan pada file teks dengan algoritma winnowing, Elex Media, Jakarta.
- _____,_____, Kamus Besar Bahasa Indonesia [Online], (<http://www.pusatbahasa.diknas.go.id/kbbi>, diakses tanggal 1 11 april 2016).
- Stein, B., &Eissen, S. M.Z., 2006, Near Similarity Search and Plagiarism Analysis. Dalam Spiliopoulou et.al. (Eds). Data and Information Analysis to Knowledge Engineering Selected Papers from the 295th Annual Conference Of German Classification Society (GFKI), pp. 430-437. Magdeburg : Springer.
- Setiawan, Guntur. 2004. Implementasi Dalam Birokrasi Pembangunan. Jakarta. Cipta Dunia.
- Pratama, M.R. & Cahyono, E.K. & Marthasari, G.I. 2012. Aplikasi

Pendeteksi Duplikasi Dokumen
Teks Bahasa Indonesia
Menggunakan Algoritma
Winnowing Dengan Metode K-
Gram Dan Synonym
Recognition. Skripsi. Jurusan
Teknik Informatika Universitas
Muhammadiyah Malang.