

IMPLEMENTASI SISTEM PENDETEKSI PLAGIAT PADA DOKUMEN BAHASA INDONESIA MENGGUNAKAN ALGORITMA WINNOWING

Ratih Ratna Hutami¹, Suyanto^{2, 3}

¹Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

Abstrak

Plagiarisme merupakan tindakan menjiplak atau mengakui hasil karya orang lain sebagai hasil karya sendiri tanpa adanya ijin. Dengan melakukan penyalinan dan modifikasi dengan sempurna, maka dapat menyamarkan plagiarisme karena karya yang dihasilkan hampir seperti karya milik sendiri sehingga sulit untuk mengetahui keorisinalitasan dari sebuah dokumen. Oleh karena itu, perlu adanya suatu sistem yang mampu mendeteksi plagiarisme.

Algoritma WInnowing merupakan salah satu algoritma pada metode document fingerprinting. Metode ini akurat dalam mengidentifikasi penyalinan teks termasuk bagian kecil yang mirip dalam sekumpulan dokumen melalui fingerprint yang dihasilkan. Melalui pencocokan fingerprint akan diperoleh nilai similarity antar dokumen.

Berdasarkan hasil pengujian, parameter k-gram, basis hash, dan ukuran window berpengaruh dalam nilai similarity yang dihasilkan antar dokumen. Pada tugas akhir ini, sistem pendeteksi plagiarisme mampu mendeteksi plagiarisme dengan tipe plagiat aktif pasif, penambahan kata, carbon copy, dan ubah susunan kalimat.

Kata Kunci : plagiarisme, algoritma WInnowing, fingerprint

Abstract

Plagiarism is an action that copying or confessing someone's document as his or her document without permission. By doing perfect duplication and modification, then it can camouflage plagiarism because the produced document almost like different document from the source so it can be hard to know the originality of the document. Because of that, it is needed a system that can detect plagiarism.

WInnowing algorithm is one of the algorithm in document fingerprinting method. This method is accurate when identifying text duplication include small part that similar from group of document by produced fingerprint. by adjusting the fingerprint will be obtained similarity value among document.

Based on testing result, the parameter k-gram, basis hash and window size influence to the similarity value that produced among document. In this thesis, plagiarism detection system can detect plagiarism with kind of plagiarism active passive, word increasing, carbon copy, and sentence scheme changing.

Keywords : plagiarism, WInnowing algorithm, fingerprint

Bab I Pendahuluan

I.1. Latar Belakang

Plagiarisme merupakan salah satu tindakan yang dilarang, karena tindakan tersebut termasuk pelanggaran terhadap hasil karya seseorang yaitu dengan menjiplak atau mengakui hasil karya orang lain sebagai hasil karya sendiri tanpa adanya ijin. Plagiarisme dapat memberikan dampak buruk bagi para plagiat dan juga orang yang dijiplak hasil karyanya. Selain itu, plagiat dapat menurunkan kreativitas seseorang dalam menciptakan hasil karya.

Plagiarisme sering terjadi di berbagai kalangan. Plagiarisme dapat dilakukan dengan cara menulis ulang secara lengkap teks yang sudah ada, menghilangkan atau menambahkan kata, parafrase, mengganti kata dengan sinonimnya, mengubah kalimat aktif menjadi pasif atau sebaliknya dan lain lain. Dengan melakukan penyalinan dan modifikasi dengan sempurna, maka dapat menyamarkan plagiarisme karena karya yang dihasilkan hampir seperti karya milik sendiri sehingga sulit untuk mengetahui keorisinalitasan dari sebuah dokumen. Kemungkinan untuk mendeteksi plagiarisme terhadap sekumpulan dokumen yang sangat banyak akan membutuhkan waktu yang sangat lama dan akan mengalami kesulitan. Oleh karena itu dibutuhkan sebuah sistem yang dapat mendeteksi plagiat.

Pada tugas akhir ini akan dibangun sebuah sistem yang dapat melakukan pendeteksian kemiripan (*similarity*) dengan membandingkan dokumen asli dengan dokumen uji yang keduanya merupakan dokumen berbahasa Indonesia menggunakan algoritma *Winnowing*. *Winnowing* merupakan algoritma *document fingerprinting* yang akurat dalam mengidentifikasi penyalinan teks termasuk bagian kecil yang mirip dalam sekumpulan dokumen melalui *fingerprint* yang dihasilkan.

I.2. Rumusan Masalah

Berdasarkan latar belakang permasalahan di atas, maka dapat dirumuskan masalah sebagai berikut :

1. Bagaimana mengimplementasikan sistem pendeteksi plagiat pada dokumen berbahasa Indonesia ?
2. Bagaimana mendapatkan *fingerprint* dari sebuah teks dengan menggunakan algoritma *Winnowing* ?
3. Bagaimana nilai kemiripan (*similarity*) yang dihasilkan dengan menggunakan algoritma *Winnowing* ?
4. Bagaimana pengaruh parameter (*k-gram*, basis *hash*, dan ukuran *window*) dalam algoritma *Winnowing* terhadap nilai kemiripan (*similarity*) yang dihasilkan?

Adapun batasan masalah atau lingkup dari tugas akhir ini meliputi :

1. Dokumen teks yang diuji mempunyai ekstensi .txt dan berbahasa Indonesia
2. Dokumen teks yang diidentifikasi minimal sepasang dokumen
3. Sistem tidak menangani apabila terdapat kesalahan ejaan penulisan pada dokumen
4. Sistem tidak mendeteksi plagiarisme berdasarkan adanya sinonim pada dokumen
5. Proses *stemming* yang ada dalam tahap *preprocessing* tidak menjadi fokus dalam Tugas Akhir ini
6. Sistem fokus pada pengukuran nilai kemiripan (*similarity*) dalam teks
7. Tidak memperhitungkan kompleksitas waktu proses penggunaan algoritma *Winnowing*.

I.3. Tujuan

Tujuan yang ingin dicapai dalam pembuatan tugas akhir ini, antara lain :

1. Mengimplementasikan sistem pendeteksi plagiat pada dokumen berbahasa Indonesia dengan menggunakan algoritma *Winnowing*
2. Mendapatkan *fingerprint* sebuah teks menggunakan algoritma *Winnowing*
3. Mengetahui nilai kemiripan (*similarity*) yang dihasilkan dengan menggunakan algoritma *Winnowing*
4. Menganalisis pengaruh parameter (*k-gram*, basis *hash*, dan ukuran *window*) dalam algoritma *Winnowing* terhadap nilai kemiripan (*similarity*) yang dihasilkan.

I.4. Metodologi Penyelesaian Masalah

Metodologi penyelesaian masalah yang digunakan dalam penyusunan Tugas Akhir ini adalah sebagai berikut :

1. Studi literatur
Melakukan pembelajaran mengenai metode *document fingerprinting* dan algoritma *Winnowing* melalui berbagai macam media, antara lain melalui internet, jurnal, dan buku.
2. Perancangan sistem
Melakukan perancangan model terhadap sistem yang akan dibangun yaitu merancang model penerapan metode *document fingerprinting* dengan menggunakan algoritma *Winnowing* untuk mendapatkan *fingerprint* dan posisinya dari sebuah dokumen. Kemudian membandingkan *fingerprint* yang dimiliki oleh satu file dengan file yang lainnya.
3. Implementasi
Pengimplementasian sistem pendeteksi plagiat berdasarkan perancangan yang telah dibuat sebelumnya dalam sebuah *source code* ke dalam program komputer.

4. Pengujian dan analisis
Melakukan uji coba terhadap program yang telah dibuat dengan membandingkan antara dokumen asli dan dokumen uji. Kemudian dilakukan analisis dan evaluasi terhadap kekurangan program dan memperbaikinya.
5. Pembuatan laporan
Melakukan pembuatan laporan tugas akhir berdasarkan pada apa yang telah dilakukan mulai dari analisis, perancangan, implementasi hingga pengujian.



Bab V Kesimpulan dan saran

V.1. Kesimpulan

Berdasarkan analisis dan pengujian yang telah dilakukan dapat disimpulkan bahwa:

1. Algoritma *Winnowing* dapat diimplementasikan untuk melakukan pendeteksian tindakan plagiarisme pada sebuah dokumen teks berdasarkan pencocokan *fingerprint-fingerprint* yang dihasilkan.
2. Dari parameter-parameter pada algoritma *Winnowing*, nilai k dalam pembentukan K -gram sangat berpengaruh terhadap nilai *similarity* yang dihasilkan.
3. Semakin besar nilai k pada pembentukan k -gram akan menghasilkan nilai *similarity* yang rendah.
4. Nilai *similarity* yang dihasilkan algoritma *Winnowing* mempunyai akurasi di atas 80%
5. Dengan menggunakan konfigurasi nilai k -gram, basis *hash*, dan ukuran *window* yang tepat dapat menghasilkan akurasi sistem yang tinggi.
6. Sistem ini dapat digunakan untuk mendeteksi tipe-tipe plagiat seperti aktif pasif, *carbon copy*, ubah susunan kalimat, dan tambah kata dengan akurasi yang hampir mendekati 100%.

V.2. Saran

Untuk penelitian lebih lanjut, disarankan menggunakan data uji yang lebih variatif seperti pengubahan isi teks berdasarkan tipe-tipe plagiat yang ada sehingga sistem dapat mendeteksi semua jenis tipe plagiat. Selain itu melakukan modifikasi pada algoritma *Winnowing* untuk menghasilkan akurasi sistem yang lebih tinggi.

Telkom
University

Daftar Pustaka

- [2] _____. _____. Algoritma Pencarian *String*. [Online]. Tersedia: http://id.wikipedia.org/wiki/Algoritma_pencarian_string. [12 Januari 2012].
- [2] _____. _____. Kamus Besar Bahasa Indonesia. [Online]. Tersedia: <http://www.pusatbahasa.diknas.go.id/kbbi>. [21 April 2012].
- [3] Asian, Jelita (dkk.). 2005. "Stemming Indonesian". *Australian Computer Science Conference (ACSC 2005)*. 38.
- [4] Du Zou (dkk.). 2010. "A Cluster-Based Plagiarism Detection Method". *Lab Report for PAN at CLEF*.
- [5] Elbegbayan, Norzima. 2005. "Winnowing, a Document Fingerprinting Algorithm". TDDC03 Projects, Spring 2005.
- [6] Firdaus, Hari Bagus. 2008. "Deteksi Plagiat Dokumen Menggunakan Algoritma Rabin-Karp". Program Studi Teknik Informatika Sekolah Teknik Elektro dan Informatika, Institut Teknologi Bandung (ITB). Bandung.
- [7] Hoad, Timothy C dan Justin Zobel. 2003. "Methods for Identifying Versioned and Plagiarized Documents". *Journal of the American Society for Information Science and Technology*. 54 (3), 203-215.
- [8] Iyer, Parvati dan Abhispita Singh. 2005. "Document Similarity Analysis for a Plagiarism Detection System". In *Proceedings of the 2nd Indian International Conference on Artificial Intelligence (IICAI-05)*, pp. 2534-2544.
- [9] Kurniawati, Ana (dkk.). 2008. "Perbandingan Pendekatan Deteksi Plagiarism Dokumen Dalam Bahasa Inggris". *Proceeding, Seminar Ilmiah Nasional Komputer dan Sistem Intelijen (KOMMIT 2008)*, 286.
- [10] Munir, Rinaldi. 2006. *Strategik Algoritmik*. Bandung.
- [11] Nugroho, Eko. 2011. Perancangan Sistem Deteksi Plagiarisme Dokumen Teks dengan Menggunakan Algoritma Rabin-Karp. Skripsi pada FMIPA Universitas Brawijaya Malang: <http://female.store.co.id/images/media/ilmu%20komputer%20-%20full.pdf> [23 November 2011].
- [12] Purwitasari, Diana (dkk.). 2011. "Deteksi Keberadaan Kalimat Sama Sebagai Indikasi Penjiplakan Dengan Algoritma Hashing Berbasis *N-Gram*". *Jurnal Ilmiah KURSOR*. 6 (1), 37-44.
- [13] Rachmad, Aeri dan Nonot Wisnu K. 2008. "Aplikasi Web Browser dengan URL Text Filtering Menggunakan Metode Brute Force". *Prosiding Seminar Nasional Teknoin 2008 Bidang Teknik Informatika*.

- [14] Schleimer, Saul (dkk.). 2003. “*Winnowing: Local Algorithms For Document Fingerprinting*”. In *Proceedings of the ACM SIGMOD international conference on management of data*, pp.76-85.

