Volume 12 Number 2 July 2022 Page. 151-156
Journal Homepage : http://teknois.stikombinaniaga.ac.id/index.php/JBS
DOI Link : http://doi.org/10.36350/jbs.v12i2

Article DOI: 10.36350/jbs.v12i2.141
Received: May; Accepted: June; Published: July

Penerapan Algoritma *N-GRAM* Dan *WINNOWING* Untuk Deteksi Plagiarisme Usulan Dokumen Perguruan Tinggi

Adiat Pariddudin[1*], Muhammad Jihad Abdul Fatah[2]
[1] Sistem Informasi/Universitas Binaniga Indonesia.
Email: adiat@stikombinaniaga.ac.id
[2] Teknik Informatika/Universitas Binaniga Indonesia.
Email: muhammadjihadaf@gmail.com

ABSTRACT

Plagiarism which is a disgraceful act can occur in any series of documents, including one of them is a feasibility study book document as is often made by universities, foundations of organizing bodies, to consultants related to the interests of establishing universities as well as proposals for opening new study programs submitted to the Ministry of Education and Culture and Higher Education (Kemendikbud Dikti). Consultants, especially those with legal entities, will determine the smoothness of proposals to rejection of proposals due to flaws in the text, the most common of which is plagiarism, which requires consultants to remain skilled in ensuring the authentication of documents made to clients as well as to the institution or institution that is the evaluator. This plagiarism can be prevented by building a plagiarism detection system with a programming algorithm approach such as; The second N-Gram Algorithm and Winnowing Algorithm were successfully built by researchers using the Research and Development (R&D) method to detect plagiarism in the proposed college documents, in this case training documents and test documents in the form of systems and applications. The application is made web-based by utilizing an interface to make it easier for leaders to ensure the similarity and authentication of documents made by consultants. The n-gram algorithm is a document processing method that is usually used in spelling correction, word prediction and other processing and Winnowing Algorithm is one of the algorithms in the document fingerprinting method. Both are used to detect Plagiarism of Proposed College Documents with the Feasibility Study Book document object resulting in a level of accuracy and effectiveness with the parameters of similarity between documents as stated in the percentage of similarity (similarity percentage) which can be concluded the value of plagiarism between documents.

Keywords*: Plagiarism; N-Gram; Winnowing; Training Document; Test Documents.*

ABSTRAK

Plagiarisme yang merupakan perbuatan tercela dapat terjadi dalam rangkaian dokumen apapun termasuk salah satunya adalah dokumen buku studi kelayakan sebagaimana yang sering dibuat oleh perguruan tinggi, Yayasan badan penyelenggara, hingga konsultan terkait kepentingan pendirian perguruan tinggi maupun usulan pembukaan program studi baru yang disampaikan ke Kementerian Pendidikan dan Kebudayaan serta Pendidikan Tinggi (Kemendikbud Dikti). Konsultan, terlebih yang berbadan hukum akan menentukan kelancaran usulan hingga penolakan usulan karena cela dalam naskah yang paling umum terjadi yaitu plagiarisme sehingga menuntut konsultan untuk tetap terampil dalam memastikan otentikasi dokumen yang dibuat terhadap klien maupun terhadap institusi atau Lembaga yang menjadi evaluatornya. Plagiarisme ini dapat dicegah

Volume 12 Number 2 July 2022 Page. 151-156
Journal Homepage : http://teknois.stikombinaniaga.ac.id/index.php/JBS
DOI Link : http://doi.org/10.36350/jbs.v12i2

dengan cara membangun sebuah sistem deteksi plagiarisme dengan pendekatan algoritma pemrograman seperti diantaranya; Algoritma N-Gram dan Algoritma Winnowing yang kedua berhasil dibangun oleh peneliti dengan metode Research and Development (R&D) untuk mendeteksi plagiarisme dalam naskah usulan dokumen perguruan tinggi dalam hal ini dokumen latih dan dokumen uji dalam bentuk sistem maupun aplikasi. Aplikasi dibuat berbasis web dengan memanfaatkan interface untuk memudahkan para pimpinan untuk memastikan kemiripan dan otentikasi dokumen yang dibuat oleh para konsultan. Algoritma n-gram adalah suatu metode pengolahan dokumen yang biasanya digunakan dalam spelling correction, word prediction dan pengolahan lainnya dan Algoritma Winnowing merupakan salah satu algoritma pada metode document fingerprinting. Keduanya dimanfaatkan untuk mendeteksi Plagiarisme Usulan Dokumen Perguruan Tinggi dengan objek dokumen Buku Studi Kelayakan menghasilkan tingkat akurasi dan efektivitas dengan parameter kemiripan antar dokumen yang dituangkan dalam persentase kemiripan (similarity percentage) yang dapat disimpulkan nilai plagiarisme antar dokumen.

***Keywords**: Plagiarisme; N-Gram; Winnowing; Dokumen Latih; Dokumen Uji.*

A. PENDAHULUAN

1. latar Belakang

Kemajuan dunia Information Technology (IT) saat ini telah berkembang pesat terlebih dengan keberadaan Internet of Things (IoT), seiring dengan ditemukannya berbagai macam fasilitas yang memudahkan kegiatan manusia. Hal ini semakin memanjakan kehidupan manusia di era modern ini. Jarak tak lagi menjadi penghalang untuk dapat melakukan sesuatu. Efisien terhadap waktu. Dengan sedikit waktu, beberapa permasalahan bisa terselesaikan dengan cepat. Begitu pula dengan ruang dan biaya, semua bisa dihemat dengan adanya penemuan-penemuan di bidang IT pada saat ini.

Hampir setiap informasi dapat dirujuk sumbernya di internet melalui literasi digital baik berupa teks, audio, video, dan format lainnya yang memudahkan setiap orang yang membutuhkannya seperti pencarian untuk bahan tugas akhir, bahan kurikulum, bahan kajian tertentu dan lainnya. Banyaknya informasi yang tersedia memudahkan seseorang untuk membuat kajian dengan tipe, pola, bahan, gaya Bahasa, gaya penulisan, dan setiap apapun dari informasi yang tersedia. Konsumsi informasi ini dapat dinikmati tidak hanya oleh perorangan, namun juga badan hukum seperti penyedia jasa pembuatan kajian tertentu seperti kajian Studi Kelayakan, Proposal, Instrumen dan dokumen lainnya. Hal ini seperti yang dilakukan oleh beberapa pemberi jasa dokumen-dokumen tersebut, terdapat sebuah perusahaan konsultan di bidang jasa kajian kelayakan untuk institusi Pendidikan tinggi baik untuk akreditasi, pendirian, perubahan bentuk dan layanan lainnya yang menuntut kajian dokumen tertentu.

Aspek keamanan (Security) dan kepercayaan (Trust) dalam layanan ini menjadi 2 (dua) hal yang harus dipenuhi karena data harus aman dari kata yang seharusnya tidak boleh sama sekali tertuang dan tercantum serta kepercayaan harus tetap dipertahankan sebagai pertaruhan dalam bisnis jasa layanan kajian dokumen. Diantara pertaruhan ini adalah bahwa dokumen harus terbebas dari unsur plagiarisme.

Secara peraturan yang berlaku dalam lintas instansi terhadap kliennya, kemiripan data dengan usulan lain dalam dokumen tidak boleh lebih dari 20% kesamaan dari total tubuh setiap dokumen. Hal ini sudah banyak diberlakukan baik di BAN PT, Kemendikbud, Kemenristek Dikti dan instansi lainnya. Menurut pengakuan salah satu konsultan bahwa selama ini setiap staf yang ada sudah memenuhi dan menjaga setiap kajiannya agar tidak terduga, tersindikat, diketahui, ditemukan adanya Plagiarisme dalam dokumen yang dibuat. Sampai hari ini, masih melakukan hal tersebut secara manual, dimana setiap kajian yang dibuat oleh para staf dan tim ahli, harus dievaluasi dan dianalisis oleh Direktur melalui (1) keumuman keahlian dan kepakaran serta keahlian Direktur berdasarkan persepsi dan daya ingatnya, (2)

membandingkan dokumen satu dengan dokumen lainnya utuk menemukan kemiripan melalui aplikasi umum seperti MS Word dengan fitur "Find" atau fitur menemukan kata dan kalimat.

Hal tersebut diakui kurang efektif karena dianggap tidak efisien dan memiliki peluang untuk tidak terdeteksinya plagiarisme dalam kondisi tertentu atau dapat disebut "Tidak Terukur". Tak hanya, perusahaan ini, diluar sana tentu banyak sekali perorangan, badan hukum yang memberikan jasa sejenis seperti proposal, tugas, laporan, buku, artikel dan sejenisnya yang tentu harus dapat terbebas dari plagiarisme. Penelitian ini mengkaji untuk menawarkan sebuah aplikasi yang akan dibangun oleh peneliti untuk memberikan solusi yang dapat digunakan untuk analisis sekaligus memeriksa keberadaan plagiarisme dalam dokumen.

Pendeteksian kesamaan kata pada dokumen merupakan salah satu langkah untuk mencegah plagiarisme, tetapi membutuhkan waktu lama ketika pengecekan secara manual. Deteksi kesamaan kata dapat dilakukan menggunakan algoritma yang memperhatikan akurasi sistem saat mendeteksi, kecepatan dan efisiensi waktu adalah kelebihan dari penggunaan sistem deteksi plagiarisme dalam membantu pengecekan dokumen. Deteksi kesamaan kata memiliki kriteria tanda spasi, huruf kapital tidak berpengaruh, menghilangkan kata yang tidak relevan, dan tidak terpengaruh pada letak kata.

Terdapat beberapa algoritma yang digunakan dalam mendeteksi kemiripan kata, diantaranya algoritma n-gram yang menggunakan nilai n. Karakter yang terdapat pada teks digunakan untuk mengetahui tingkat kemiripan dengan panjang sesuai n, Algortima winnowing adalah tahap selanjutnya setelah menggunakan algoritma n-gram.

Berdasarkan pemaparan tentang beberapa hal di atas yang menjadi landasan untuk dilakukannya penelitian ini. Karena hal tersebut berinisiatif untuk membangun aplikasi deteksi plagiarisme yang menerapkan metode algoritma n-gram dan winnowing dengan objek dokumen studi kelayakan perguruan tinggi. Aplikasi yang di harapkan dapat menampilkan hasil dari Penerapan Algoritma N-Gram Dan Winnowing Untuk Deteksi Plagiarisme Usulan Dokumen Perguruan Tinggi.

2. Permasalahan

Dalam pemeriksaan dokumen studi kelayakan dengan membandingkan antara dokumen satu dengan lainnya dan menggunakan fasilitas "Find" pada MS Word menjadi hambatan dalam mengetahui seberapa besar kemiripan dokumen yang diperiksa dengan dokumen yang lainnya.

3. Tujuan

Menerapkan algoritma algoritma n-gram dan algoritma winnowing untuk meningkatkan akurasi dan efektivitas pendeteksi plagiarisme dokumen studi kelayakan.

4. Tinjauan Pustaka

a. Pengertian Algoritma

Algoritma adalah langkah-langkah logis penyelesaian masalah yang disusun secara sistematis dan logis untuk menghasilkan solusi yang tepat. Awalnya algoritma digunakan untuk penghitungan dalam ilmu matematika namun dalam perkembangannya, kata algoritma justru banyak dipakai pada bidang pemrograman komputer (Rosa A.S., 2010:1).

b. Algoritma N-Gram dan Winnowing

Algortima N-gram merupakan salah satu proses yang secara luas digunakan dalam text mining (pengolahan teks) dan pengolahan bahasa. Dalam hal ini n-gram digunakan untuk memisahkan teks menjadi rangkaian kata dengan panjang atau n yang ditentukan adapun Algoritma winnowing digunakan untuk merubah nilai n-gram menjadi nilai angka menggunakan metode hashing.

B. METODE

Metode yang digunakan yaitu Algoritma N-Gram dan Winnowing merupakan Metode konseptual atau teori yang digunakan adalah Algoritma N-gram dan Algoritma Winnowing, Algortima N-gram merupakan salah satu proses yang secara luas digunakan dalam text mining (pengolahan teks) dan pengolahan bahasa. Dalam hal ini n-gram digunakan untuk memisahkan teks menjadi

Volume 12 Number 2 July 2022 Page. 151-156
Journal Homepage : http://teknois.stikombinaniaga.ac.id/index.php/JBS
DOI Link : http://doi.org/10.36350/jbs.v12i2

rangkaian kata dengan panjang atau n yang ditentukan. Panjang n dimulai dari n=1 sampai tak terhinga. Contoh rangkaian n-gram dari kata “pulangpergi”, dengan n=4 akan menjadi “pula, ulan, lang, angp, ngpe, gper, perg, ergi”.

Rangkaian n-gram dibentuk dengan diawali dari menghilangkan spasi pada rangkaian kata dasar, kemudian pembentukan n-gram menurut panjang n. Sistem pada penelitian ini menggunakan n=7 untuk mendapatkan nominal n yang paling efektif digunakan dalam pengecekan indikasi plagiarisme dokumen studi kelayakan.

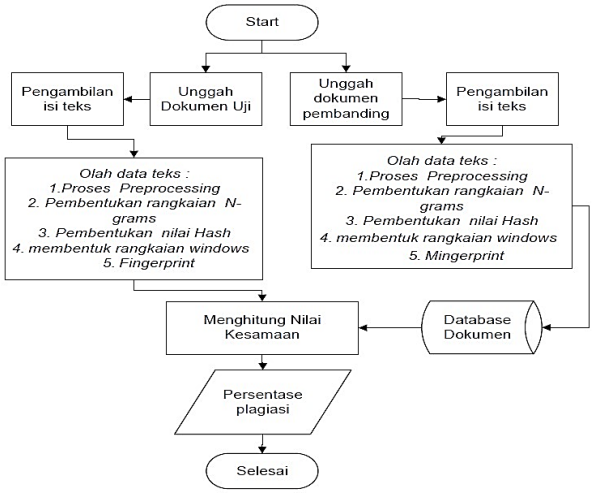
Algoritma winnowing digunakan untuk merubah nilai n-gram menjadi nilai angka menggunakan metode hashing. Rumus yang digunakan dalam perhitungan hashing pada penelitian ini adalah Rolling Hash.

$$H(C1...Cn) = C1*b(n-1)+C2*b(n-2)+\cdots+C(n-1)*b(n)+Cn$$

Berdasarkan rumus tersebut, diketahui c yaitu bilangan ASCII dari karakter, b = bilangan prima yang ditentukan, dan n = panjang n-gram. Untuk rumus kedua digunakan pada n-gram yang sama dan seterusnya, dikarenakan tidak memerlukan penghitungan karakter pertama.

Algortima winnowing digunakan untuk menentukan Fingerprint diawali dengan merubah nilai hash menjadi rangkaian hash berdasarkan nilai window (w). Maka dipilih nilai hash terkecil dari setiap window. Selanjutnya dilakukan pengecekan apabila ada nilai hash yang sama maka hanya di ambil salah satu.

Secara umum alur dari sistem dalam memeriksa plagiarisme ini terdiri dari beberapa proses, yang disajikan pada Gambar 1. Proses Pemeriksaan Plagiarisme



Gambar 1. Proses Pemeriksaan Plagiarisme

Berdasarkan alur proses pada Gambar 1. Proses Pemeriksaan Plagiarisme diawali dengan unggah file dokumen untuk pembanding lalu didapatkan isi teks, proses preprocessing yaitu dilakukan tokenizing (mengubah kalimat menjadi rangkaian kata), menghilangkan stopword atau menghapus kata tidak relevan kemudian stemming untuk mendapatkan kata dasar, selanjutnya pembentukan n-gram dari hasil preprocessing, mendapatkan nilai hash dari hasil proses sebelumnya, membentuk windowing dan mencari nilai fingerprint lalu disimpan dalam basis data dan digunakan untuk membandingkan dengan hasil dokumen uji.

Langkah selanjutnya mengunggah dokumen uji, dan melakukan proses preprocessing sampai mendapatkan nilai fingerprint. kemudian dilakukan pengecekan dengan membandingkan fingerprint dokumen pembanding dan uji untuk mendapatkan prosentase kesamaan kata antar dokumen.

C. HASIL DAN PEMBAHASAN

1. HASIL

Volume 12 Number 2 July 2022 Page. 151-156
Journal Homepage : http://teknois.stikombinaniaga.ac.id/index.php/JBS
DOI Link : http://doi.org/10.36350/jbs.v12i2

Berdasarkan Tabel 2. Hasil Pengecekan Dokumen Latih Dan Dokumen Uji. Dokumen uji yang telah di extract menghasilkan 8603 fingerprint, kemudian fingerprint tersebut dibandingkan dengan data latih yang sudah tersipan di database. Perhitungan similarity menggunakan rumus jaccard similarity. Intersection adalah jumlah fingerprint yang sama antara dokumen latih dan dokumen uji, sedangkan Union adalah jumlah gabungan fingerprint dokumen latih dan dokumen uji.

$$Similarity\ (\%) = \frac{Intersection}{Union}\ X\ 100\%$$

Hasil similarity yang diperoleh dari pengujian disajikan pada Gambar 2. Hasil Similarity Antara Dokumen Latih Dan Dokumen Uji

Gambar 2. Hasil Similarity Antara Dokumen Latih Dan Dokumen Uji

Nilai similarity menunjukan persentase kemiripan dokumen latih dan dokumen uji. Semakin besar nilai similarity, maka semakin mirip dengan dokumen lainnya.

2. PEMBAHASAN

a. Implementasi Metode Algoritma N-Gram dan Winnowing pada Web

Implementasi metode algoritma n-gram dan winnowing diterapkan pada php dan dikemas dengan tampilan web dilakukan pada saat proses pengecekan terkait plagiarisme. Hal ini dilakukan guna mengetahui indikasi dugaan plagiarisme suatu dokumen dalam bentuk hasil akhir prosentase. Hasil implementasi metode algoritma n-gram dan winnowing disajikan Tabel 1. Hasil dari Perhitungan Algoritma N-Gram Dan Winnowing

Tabel 1. Hasil dari Perhitungan Algoritma N-Gram Dan Winnowing

No	Nama Dokumen	Praproses		N-Gram	Hashing	Winnowing	Fingerprint
		Sebelum	Sesudah				
1	Doc1.docx	181.006	104.871	104.865	104.865	104.861	8.301
2	Doc2.docx	59.700	31.416	31.410	31.410	31.406	4.103
3	Doc3.docx	204.377	88.711	88.705	88.705	88.701	7.363
4	Doc4.docx	165.777	93.364	93.358	93.358	93.354	8.458
5	Doc5.docx	56.037	29.755	29.749	29.749	29.745	3.965
6	Doc6.docx	204.883	88.925	88.919	88.919	88.915	7.365
7	Doc7.docx	194.202	106.365	106.359	106.359	106.355	8.090
8	Doc8.docx	262.685	142.715	142.709	142.709	142.705	9.596
9	Doc9.docx	65.208	40.315	40.309	40.309	40.305	4.796
10	Doc10.docx	127.368	75.212	75.206	75.206	75.202	6.725

Sampel diambil dari objek penelitian yaitu dokumen Buku Studi Kelayakan sebanyak 10 (sepuluh) dokumen, yang kemudian diambil fingerprint terkecil dari setiap dokumen dan di simpan untuk dilakukan pengecekan dengan dokumen uji. Kemudian dilakukan pengujian dokumen menggunakan dokument uji yang telah disiapkan sebelumnya. Hasil dari extract dokument uji disajikan pada Tabel 2. hasil pengecekan dokumen latih dan dokumen uji.

Tabel 2. Hasil Pengecekan Dokumen Latih Dan Dokumen Uji

No	Nama Dokumen	Praproses		N-Gram	Hashing	Winnowing	Fingerprint
		Sebelum	Sesudah				
1	Doc-uji.docx	208.615	105.684	105.678	105.678	105.674	8.603

D. KESIMPULAN

Berdasarkan hasil penelitian yang telah diuraikan, maka kesimpulan yaitu:

1. Implementasi algoritma n-gram dan winnowing berhasil diterapkan dalam php dan dikemas dalam tampilan web untuk pengecekan plagiarisme. Algoritma n-gram adalah suatu metode pengolahan dokumen yang biasanya digunakan dalam spelling correction, word prediction dan pengolahan lainnya dan Algoritma Winnowing merupakan salah satu algoritma pada metode document fingerprinting.
2. Berdasarkan hasil penelitian dengan perhitungan Algoritma n-gram dan winnowing yang telah dilakukan antara dokumen latih dan dokumen uji berhasil meningkatkan akurasi plagiarisme sehingga dapat diketahui tingkat akurasi dan efektivitas dengan parameter kemiripan antar dokumen yang dituangkan dalam persentase kemiripan (similarity percentage) yang dapat disimpulkan nilai plagiarisme antar dokumen.

E. DAFTAR PUSTAKA

[1] Alamsyah, Nur, “Deteksi Plagiarisme Tingkat Kemiripan Judul Skripsi Dengan Algoritma Winnowing,” Technologia,Vol.8, No.4, pp.205-213, Desember 2017.
[2] Hargyo Tri Nugroho I. Pengaruh Algoritma Stemming Nazief-Adriani Terhadap Kinerja Algoritma Winnowing Untuk Mendeteksi Plagiarisme Bahasa Indonesia
[3] Harya Chandra, dkk Plagiarisme Abstrak Menggunakan Algoritma Winnowing Dan SynsetsJurnal (Jiksi) Ilmu Komputer dan Sistem Informasi, Vol. 6 No. 2 Tahun 2018
[4] Ilham, Penerapan Algoritma Winnowing Untuk Mendeteksi Kemiripan Pada Karya Tulis Mahasiswa Jurnal Teknologi Informasi dan Komunikasi Vol. 7 No. 2 tahun 2017
[5] Nawawi, dkk (2019) “Deteksi Plagiarisme Pada Dokumen Skripsi Berdasarkan Tingkat Kesamaan Dengan Menggunakan Metode Longest Common Subsequence". JANAPATI Jurnal Nasional Pendidikan Teknik Informatika ISSN 2089-8673 (Print) | ISSN 2548-4265 (Online) Volume 8, Nomor 3, Desember 2019"
[6] Pratama, dkk Analisa Perbandingan Jenis N-Gram Dalam Penentuan Similarity Text pada Deteksi Plagiat. Jurnal "Citec" Creative Information Technology Journal, Vol. 4 No. 4 Tahun 2017.
[7] Prof. Dr. Burhan Nurgiyantoro, Dr. Widyastuti Purbani, Dr. Sutiyono, 2015, “Panduan Anti Plagiarisme”, UNY, Yogyakarta.
[8] Purwitasari, dkk Deteksi Keberadaan Kalimat Sama sebagai Indikasi Penjiplakan dengan Algoritma Hashing Berbasis N-Gram, Jurnal Ilmiah Kursor, Vol.6 No.1 Januari tahun 2011.
[9] Rosa A.S., 2010, Logika Algoritma dan Pemrograman Dasar, Modula, Bandung.
[10] Setiawan A. Implementasi Algoritma Winnowing Untuk Deteksi Kemiripan Judul Skripsi Studi Kasus Stmik Budidarma Jurnal INTI (Informasi dan Teknologi Ilmiah Volume : XII, Nomor : 1, Januari 2017 ISSN: 2339-210X Tahun 2017.
[11] Stevson, Agung, Mulia, “Aplikasi Pendeteksi Plagiarisme Tugas Dan Makalah Pada Sekolah Menggunakan Algoritma Rabin Karp” Jurnal Algoritma dan Komputasi, Vol. 1, No. 1: 12 – 17 th. 2018.
[12] Suryadi H.S., Agus Sumin, 1994, Pengantar Algoritma dan Pemograman Teknik Diagram Alur dan Bahasa Basic Dasar, Gunadarma, Depok.
[13] Utari, Lis, and Imam T. Agustianto. "Menghitung Ketepatan Jawaban Soal Ujian Essay dengan Penerapan Algoritma Boyer-Moore." Teknois, vol. 8, no. 2, Nov. 2018, pp. 57-66, doi:10.36350/jbs.v8i2.15.
[14] Yuwono, Pratomo, Cahyana, Fauziah, Fachrurradjie, “Peningkatan Layanan Skripsi Mahasiswa Bebas Pagiarisme pada Program Studi Teknik Informatika UPN “Veteran” Yogyakarta,” Telematika, Vol. 15, No. 02, pp. 125-131, Oktober 2018.