

DETEKSI PLAGIARISME TINGKAT KEMIRIPAN JUDUL SKRIPSI PADA FAKULTAS TEKNOLOGI INFORMASI MENGGUNAKAN ALGORITMA WINNOWER

Nur Alamsyah¹⁾, Muhammad Rasyidan²⁾

¹⁾Fakultas Teknologi Informasi, Universitas Islam Kalimantan Muhammad Arsyad Al Banjari Banjarmasin

Email: uniskalam@gmail.com

²⁾Fakultas Teknologi Informasi, Universitas Islam Kalimantan Muhammad Arsyad Al Banjari Banjarmasin

Email: mr.syidan@gmail.com

ABSTRAK

Untuk menentukan kemiripan teks judul skripsi yang di ajukan dengan judul-judul skripsi yang telah ada sebelumnya maka diperlukan penerapan suatu algoritma untuk menentukan kemiripan teks, salah satunya algoritma untuk menentukan kemiripan teks ialah algoritma Winnowing.

Algoritma Winnowing adalah sebuah cara yang digunakan untuk mendeteksi kesamaan kata/kalimat (common subsequence) dalam dua atau lebih teks yang dibandingkan. Dua teks diketahui memiliki kesamaan kata/kalimat apabila di dalam dokumen tersebut dijumpai fingerprint, fingerprint inilah yang akan dijadikan dasar pembandingan antara teks, algoritma ini akan mencari fingerprint (kesamaan di dalam dua teks) dengan cara mengubah n-gram dari sebuah teks ke dalam bentuk nilai angka yang disebut dengan nilai hash, teknik untuk mencari nilai tersebut adalah Hashing.

Dengan sistem ini, Ketua Program Studi atau Koordinator Tugas Akhir nantinya hanya memasukkan judul yang akan diajukan mahasiswa, kemudian sistem akan mengecek secara otomatis dan menampilkan hasilnya. Hasil tersebut bisa dijadikan sebagai pertimbangan dalam pengambilan keputusan dan dapat menentukan apakah diterima atau ditolak judul skripsi tersebut.

Kata kunci : Algoritma Winnowing, fingerprint, n-gram, Hashing

1. PENDAHULUAN

Seiring berjalannya waktu telah banyak judul-judul skripsi yang telah diterima dan tersimpan di akademis. Dalam pengajuan judul tidak sedikit mahasiswa yang berulang kali mengajukan judul skripsi, hal ini dikarenakan mahasiswa tersebut tidak mengetahui judul yang diajukan telah ada sebelumnya karena penentuan diterima atau ditolaknya judul-judul yang diajukan oleh mahasiswa masih dilakukan secara manual yaitu dengan dengan mengecek satu persatu berdasarkan daftar judul-judul skripsi angkatan yang telah lulus yang tersimpan di file Microsoft excel.

Judul skripsi yang sama, tidak menutup kemungkinan isi dari skripsi tersebut juga sama namun tidak menutup kemungkinan

juga isi skripsi berbeda walaupun judul skripsi sama sehingga mahasiswa bingung dalam menentukan judul skripsinya. Dalam menentukan diterima atau tidaknya sebuah judul skripsi yang sekarang ini dilakukan adalah dengan mengecek atau membandingkan judul tersebut dengan daftar judul-judul yang telah ada, tentu saja hal ini akan membutuhkan waktu yang cukup lama belum lagi jika pengaju judul skripsi berjumlah banyak, tidak menutup kemungkinan ada judul yang terulang atau sama dengan judul-judul yang tersimpan di akademis bahkan waktu yang dibutuhkan juga akan semakin lama.

Di samping itu, pengelolaan daftar judul-judul yang telah ada juga masih manual sehingga tidak menutup kemungkinan data akan hilang dan sulit dalam

mendistribusikan informasi tersebut kepada mahasiswa. Oleh karena itu, diperlukan suatu sistem yang mampu mendeteksi persentase kemiripan teks judul secara cepat dan tepat dan mampu menyediakan informasi tersebut kepada seluruh mahasiswa.

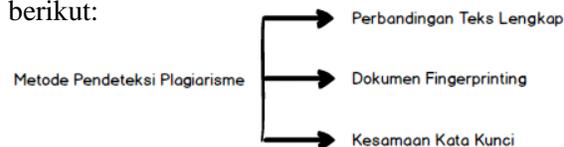
Untuk menentukan kemiripan teks judul skripsi yang di ajukan dengan judul-judul skripsi yang telah ada sebelumnya maka diperlukan penerapan suatu algoritma untuk menentukan kemiripan teks, salah satunya algoritma untuk menentukan kemiripan teks ialah algoritma Winnowing.

Algoritma Winnowing adalah sebuah cara yang digunakan untuk mendeteksi kesamaan kata/kalimat (common subsequence) dalam dua atau lebih teks yang dibandingkan. Dua teks diketahui memiliki kesamaan kata/kalimat apabila di dalam dokumen tersebut dijumpai fingerprint, fingerprint inilah yang akan dijadikan dasar pembanding antara teks, algoritma ini akan mencari fingerprint (kesamaan di dalam dua teks) dengan cara mengubah n-gram dari sebuah teks ke dalam bentuk nilai angka yang disebut dengan nilai hash, teknik untuk mencari nilai tersebut adalah Hashing.

Dengan sistem ini, Koordinator Tugas Akhir atau Ketua Program Studi nantinya hanya memasukkan judul yang akan diajukan ke formulir yang telah disediakan, kemudian sistem akan mengecek secara otomatis dan menampilkan hasilnya. Hasil tersebut bisa dijadikan sebagai pertimbangan dalam menentukan apakah diterima atau ditolak judul tersebut.

2. METODE PENELITIAN

Metode Pendeteksi Plagiarisme di bagi menjadi tiga bagian yaitu metode perbandingan teks lengkap, metode dokumen fingerprinting, dan metode kesamaan kata kunci. Klasifikasi metode pendeteksi plagiarisme dapat digambarkan sebagai berikut:



Gambar 1. Klasifikasi Metode Pendeteksi Plagiarisme

Ketiga metode diatas dapat diuraikan sebagai berikut (Stein, 2006) :

1. Perbandingan Teks Lengkap

Metode ini di terapkan dengan membandingkan semua isi dokumen. Dapat diterapkan untuk dokumen yang besar. Pendekatan ini membutuhkan waktu yang lama tetapi cukup efektif, karena kumpulan dokumen yang diperbandingkan adalah dokumen yang di simpan pada penyimpanan lokal. Metode perbandingan teks lengkap tidak dapat diterapkan untuk kumpulan dokumen yang tidak terdapat pada dokumen lokal.

Algoritma yang digunakan pada metode ini adalah *algoritma brute force, algoritma edit distance, algoritma boyer moore dan algoritma lavenshtein distance*

2. Dokumen Fingerprinting

Dokumen fingerprinting merupakan metode yang digunakan untuk mendeteksi keakuratan salinan antar dokumen, baik semua teks yang terdapat di dalam dokumen atau hanya sebagian teks saja. Prinsip kerja dari metode dokumen fingerprinting ini adalah dengan menggunakan teknik hashing. Teknik hashing adalah sebuah fungsi yang mengkonversi setiap string menjadi bilangan.

3. Kesamaan Kata Kunci.

Prinsip dari metode ini adalah mengekstrak kata kunci dari dokumen dan kemudian di bandingkan dengan kata kunci pada dokumen yang lain. Pendekatan yang digunakan pada metode ini adalah teknik dot.

3. HASIL DAN PEMBAHASAN

Hasil Analisis Algoritma Winnowing

Penelitian ini menggunakan metode eksperimen, dengan tahapan penelitian sebagai berikut:

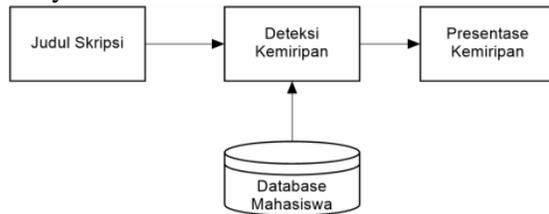
Pengolahan Data

Data yang digunakan berupa dokumen teks yang berasal dari judul-judul skripsi mahasiswa yang telah lulus mulai angkatan Tahun 2011 sd angkatan tahun 2013 dengan total 1092 Judul skripsi, kemudian data yang tersimpan di Microsoft excel nantinya akan

diiputkan ke form input data aplikasi deteksi plagiarisme, berikut contoh data dalam bentuk excel,

Analisis Deteksi Kemiripan dengan Algoritma Winnowing

Proses deteksi kemiripan judul dapat dilakukan dengan menerapkan berbagai metode untuk menghasilkan presentasi kemiripan. Tujuan utama dari deteksi kemiripan judul adalah untuk menentukan presentase kemiripan judul satu dengan judul lainnya.



Gambar 2. Skema deteksi kemiripan judul skripsi

Deteksi kemiripan judul skripsi dapat dideteksi dengan salah satu algoritma yaitu dengan algoritma winnowing dimana hasil presentase kemiripan judul ditampilkan. Proses deteksi kemiripan dilakukan dengan menampilkan presentase dengan menggunakan Persamaan *Jaccard Coeficient* pada hasil Fingerprint.

Adapun langkah-langkah deteksi kemiripan judul skripsi adalah sebagai berikut:

1. Masukkan judul skripsi yang akan dilihat tingkat presentase nya dengan judul skripsi yang sudah ada sebelumnya.
2. Memasukan nilai n-gram, untuk memebentuk rangkaian gram pada judul yang dimasukan dan judul yang dibandingkan.
3. Masukan nilai Window, untuk menentukan pembentukan window dari nilai Hash.
4. Proses deteksi kemiripan judul skripsi dengan algoritma winnowing yang mana tujuanya dari proses deteksi kemiripan judul tersebut ada untuk menampilkan tingkat presentase dengan judul-judul yang sudah ada sebelumnya. Berdasarkan analisa di atas, dapat disimpulkan bahwa proses deteksi kemiripan judul skripsi berdasarkan

algoritma winnowing menghasilkan presentase kemiripan dengan kumpulan judul skripsi yang sudah ada, karena semakin tinggi presentase kemiripan maka akan semakin mirip dengan judul yang dibandingkan.

Implementasi Algoritma Winnowing

Tahapan-tahapan untuk mengimplementasikan Algoritma Winnowing untuk mendeteksi presentase plagiarisme pada judul skripsi adalah sebagai berikut;

1. Ambil 2 judul skripsi yang akan dibandingkan.

Teks Judul 1:

“APLIKASI LAYANAN KEPENDUDUKAN ELEKTRONIK PADA DINAS KEPENDUDUKAN DAN PENCATATAN SIPIL KOTA BANJARMASIN”

Teks Judul 2:

“SISTEM INFORMASI MANAJEMEN LAYANAN KEPENDUDUKAN PADA DISCAPIL KOTA BANJARMASIN”

2. Pembuangan Karakter yang Tidak Relevan

Hapus semua huruf yang bukan A-Z, a-z, 0-9. Kemudian ubah menjadi huruf kecil semua.

Teks Judul 1 akan terbentuk menjadi seperti berikut:

aplikasilayanankependudukanelektronikpada dinaskependudukanpencatatan sipilkotabanjarmasin

dan Teks Judul 2 akan terbentuk menjadi seperti berikut:

sisteminformasimanajemenlayanankependudukanpadadiscapilkotabanjarmasin

3. Pembentukan Rangkaian n-gram

Buat *n-gram* untuk masing-masing judul kalimat. Jumlah data pengelompokan *n-gram* ini bisa dimulai dari 2, 3, 5, 7 dan seterusnya misal ***n-gram=3***.

Pada teks Judul 1 akan terbentuk 87 rangkaian n-gram yaitu:

Pada teks Judul 1 akan terbentuk 87 rangkaian n-gram yaitu:

apl pli lik ika kas asi sil ila lay aya yan ana nan ank nke kep epe pen end ndu dud udu duk uka kan ane nel ele lek ekt ktr tro ron oni nik ikp kpa pad ada dad adi din ina nas ask ske kep epe pen end ndu dud udu duk uka kan and nda dan anp npe pen enc nca cat ata tat ata tan ans nsi sip ipi pil ilk lko kot ota tab aba ban anj nja jar arm rma mas asi sin

Sedangkan pada teks Judul 2 terbentuk 66 rangkaian n-gram yaitu:

sis ist ste tem emi min inf nfo for orm rma mas asi sim ima man ana naj aje jem eme men enl nla lay aya yan ana nan ank nke kep epe pen end ndu dud udu duk uka kan anp npa pad ada dad adi dis isc sca cap api pil ilk lko kot ota tab aba ban anj nja jar arm rma mas asi sin

4. Perhitungan Fungsi Hash untuk Setiap n-gram

Buat Rolling Hash untuk masing-masing N-Gram , Perhitungan nilai hash pada rangkaian n-gram pada teks judul 1 bagian pertama “apl” dengan nilai basis (b) = 2, panjang rangkaian ngram(n) = 3

$$H_{(apl)} = \text{ascii}_{(a)} * 2^3 + \text{ascii}_{(p)} * 2^2 + \text{ascii}_{(l)} * 2^1$$

$$= 97 * 8 + 112 * 4 + 108 * 2$$

$$= 1440$$

5. Pembentukan Window dari Nilai Hash

Kelompokkan (windowing) untuk masing-masing hasil hash, langkahnya mirip seperti n-gram. Pembentukan window dari hasil perhitungan nilai hash dengan ukuran lebar window (w) = 3 pada teks judul 1 yaitu:

W-1 : {1440 1538 1498}	W-61 : {1530 1520 1446}
W-2 : {1538 1498 1462}	W-62 : {1520 1446 1470}
W-3 : {1498 1462 1474}	W-63 : {1446 1470 1412}
W-4 : {1462 1474 1446}	W-64 : {1470 1412 1434}
W-5 : {1474 1446 1556}	W-65 : {1412 1434 1548}
W-6 : {1446 1556 1466}	W-66 : {1434 1548 1434}
W-7 : {1556 1466 1494}	W-67 : {1548 1434 1536}
W-8 : {1466 1494 1454}	W-68 : {1434 1536 1446}
W-9 : {1494 1454 1576}	W-69 : {1536 1446 1550}
W-10 : {1454 1576 1410}	W-70 : {1446 1550 1564}
W-11 : {1576 1410 1488}	W-71 : {1550 1564 1498}
W-12 : {1410 1488 1430}	W-72 : {1564 1498 1532}
W-13 : {1488 1430 1510}	W-73 : {1498 1532 1486}
W-14 : {1430 1510 1484}	W-74 : {1532 1486 1514}
W-15 : {1510 1484 1458}	W-75 : {1486 1514 1532}
W-16 : {1484 1458 1520}	W-76 : {1514 1532 1546}
W-17 : {1458 1520 1448}	W-77 : {1532 1546 1512}
W-18 : {1520 1448 1514}	W-78 : {1546 1512 1362}
W-19 : {1448 1514 1468}	
W-20 : {1514 1468 1570}	
W-21 : {1468 1570 1482}	

Kemudian pada teks judul 2 hasilnya yaitu:

W-1 : {1570 1532 1586}	W-49 : {1498 1510 1404}
W-2 : {1532 1586 1550}	W-50 : {1510 1404 1434}
W-3 : {1586 1550 1454}	W-51 : {1404 1434 1532}
W-4 : {1550 1454 1512}	W-52 : {1434 1532 1486}
W-5 : {1454 1512 1484}	W-53 : {1532 1486 1514}
W-6 : {1512 1484 1510}	W-54 : {1486 1514 1532}
W-7 : {1484 1510 1488}	W-55 : {1514 1532 1546}
W-8 : {1510 1488 1562}	W-56 : {1532 1546 1512}
W-9 : {1488 1562 1542}	W-57 : {1546 1512 1362}
W-10 : {1562 1542 1490}	W-58 : {1512 1362 1392}
W-11 : {1542 1490 1446}	W-59 : {1362 1392 1428}
W-12 : {1490 1446 1558}	W-60 : {1392 1428 1498}
W-13 : {1446 1558 1470}	W-61 : {1428 1498 1464}
W-14 : {1558 1470 1480}	W-62 : {1498 1464 1450}
W-15 : {1470 1480 1410}	W-63 : {1464 1450 1542}
W-16 : {1480 1410 1480}	W-64 : {1450 1542 1490}
W-17 : {1410 1480 1402}	W-65 : {1542 1490 1446}
W-18 : {1480 1402 1470}	W-66 : {1490 1446 1560}

6. Pemilihan Fingerprint dari Setiap Window

Ambil angka terkecil dari masing-masing window. Untuk selanjutnya angka terkecil ini diistilahkan *Fingerprints*. Pemilihan nilai fingerprint dari hasil pembentukan window pada tahap sebelumnya adalah ; nilai fingerprint pada teks judul 1 adalah **78 window**

nilai fingerprint pada teks judul 2 adalah **66 window**

Fingerprints dari kelompok (window) kalimat teks judul 1 dengan fingerprints (window) kalimat teks judul 2 inilah yang nanti akan dihitung tingkat koefisien plagiarisme-nya

7. Persamaan Jaccard Coefficient

Perhitungan kesamaan dengan menggunakan persamaan jaccard coefficient yaitu;

$$\text{Similarity (kemiripan)} = \frac{65}{88} * 100\% = 73.86\%$$

Jumlah Fingerprints pada teks judul 1 = 87
 Jumlah Fingerprints teks judul 2 = 66
 Union (Gabungan) Fingerprints 1 dan 2 = 153
 Intersection (fingerprints yang sama) = 65
 (Union - Intersection) = 88

Prosentase Plagiarisme
Koefisien Jaccard = (Intersection / (Union-Intersection)) * 100 (65/88) * 100 = **73.86 %**

Jadi dapat disimpulkan dari kedua judul diatas memiliki tingkat plagiarisme sebesar **73.86 %**

Pengujian Metode Winnowing

Untuk Fakultas Teknologi Informasi menggunakan ambang batas seperti berikut:

Ambang Batas (Threshold) Plagiarisme

Plagiarisme Ringan : <= 35%

Plagiarisme Sedang: 36% - 70%

Plagiarisme Berat : > 70%

Apabila Plagiarisme ringan maka judul DITERIMA sedangkan untuk plagiarism sedang dan berat maka judul DITOLAK.

4. KESIMPULAN DAN SARAN

Proses pengujian dilakukan mulai dari pegujian jumlah gram, jumlah window dan basis bilangan prima. Dapat dilihat pada tabel dibawah ini; Tabel hasil pengujian deteksi tingkat plagiarisme pada judul skripsi

No	n-gram (n)	Window(w)	Nilai basis	Presentase %
1	3	3	2	73.86 %
2	5	3	2	36.7 %
3	7	3	2	29.46 %
4	3	5	2	73.26 %
5	5	5	2	36.79 %
6	7	5	2	29.36 %
7	3	7	2	55.91 %
8	5	7	2	30.56 %
9	7	7	2	23.42 %
10	3	9	2	50%
11	5	9	2	31.73 %
12	7	9	2	19.82 %

Berdasarkan dari tabel hasil Pengujian diatas dalam menentukan tingkat plagiarisme dengan menggunakan n-gram, window, bilangan prima menghasilkan **73.86 % tingkat plagiarisme tinggi** artinya mendekati plagiarisme dengan n-gram = 3 , window = 3 dan bilangan prima = 2 sedangkan dengan menggunakan n-gram = 3 , window = 3 dan bilangan prima = 2 menghasilkan **19.82 % artinya hanya mempunyai tingkat plagiarisme ringan**

REFERENSI

Astuti, W. (Agustus 2017). Analisis String Matching Pada Judul Skripsi dengan Algoritma Knuth-Morris Pratt(KMP). *ILKOM Jurnal Ilmiah Volume 9 Nomor 2 ISSN Cetak 2087-1716 ISSN Online 2548-7779*, 168 - 172.

Astutik, S., Cahyani, A. D., & Sophan, M. K. (November 2014). Sistem Penilaian Esai Otomatis pada E-Learning dengan Algoritma Winnowing. *Jurnal Informatika, Vol 12, No.2 ISSN 1411-0105*, 47-52.

Christian. (2013). Perancangan Sistem Deteksi Plagiarisme Dokumen Teks

Menggunakan Algoritma Damerau Levenshtein Distance. *Jurnal Mahasiswa PTIIK UB Vol 1*.

Nurdin, & Amin, M. (September 2017). Sistem Pendeteksian Kemiripan Judul Skripsi Menggunakan Algoritma Winnowing. *Jurnal Nasional Informatika dan Teknologi Jaringan (InfoTekJar) Vol 2, No 1 e- ISSN : 2540-7600, p-ISSN : 2540-7597*.

Ridho, M. (2013). *Rancang Bangun Aplikasi Pendeteksi Penjiplakan Dokumen Menggunakan Algoritma Biword Winnowing* . Pekanbaru Riau: UNIVERSITAS ISLAM NEGERI SULTSN SYARIF KASIM .

Setiawan, A. (Januari 2017). Implementasi Algoritma Winnowing Untuk Deteksi Kemiripan Judul Skripsi Studi Kasus STIMIK Budidarma. *Majalah Ilmiah INTI Volume: XII, Nomor : 1 ISSN : 2339-210X*.