

**LAPORAN AKHIR
PENELITIAN DOSEN MUDA
BOPTN 2013**



APLIKASI PENDETEKSI PLAGIARISM PADA DOKUMEN TEKS

TIM PENELITI

- 1. Ernawati, ST, MCs (Ketua)**
- 2. Desi Andreswari, ST, MCs (Anggota)**

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS BENGKULU
Desember 2013**

Created with

HALAMAN PENGESAHAN

Judul : Aplikasi Pendeteksi Plagiarisme Pada Dokumen Teks

Peneliti/Pelaksana
Nama Lengkap : Ernawati, ST, MCs
NIP : 197308142006042001
NIDN : 14087304
Jabatan Fungsional : Lektor
Program Studi : Teknik Informatika
Nomor HP : 081373057873
Alamat Surel (e-mail) : w_ier_na@yahoo.com
Anggota
Nama Lengkap : Desi Andreswari, ST,MCs
NIDN : 7127803
Program Studi : Teknik Informatika
Institusi Mitra (jika ada) : -
Penanggung Jawab :
Biaya Penelitian : Rp 5.000.000,- (Lima Juta Rupiah)

Bengkulu, 9 Desember 2013

Mengetahui,
Dekan Fakultas Teknik

Ketua

Khairul Amri, ST, MT
NIP 197202121998021002

Ernawati, S.T.,M.Cs
NIP 197308142006042001

ABSTRAK

Teks mining pada penelitian ini digunakan untuk mendeteksi plagiarisme pada dokumen teks. Metode teks mining yang digunakan adalah Knuth Morris Pratt. Penelitian dilakukan dengan membangun aplikasi dengan tools IDE Netbeans 7.3, Jdk 1.7 dan MySQL 5.0.2.6. Penelitian merupakan penelitian terapan dengan desain menggunakan UML. Hasil yang diperoleh dari penelitian adalah bahwa aplikasi yang dibangun mampu mendeteksi plagiarisme pada dokumen teks.

Key word : plagiarisme, knuth morris pratt, text mining

KATA PENGANTAR

Puji syukur penulis ucapkan kepada Allah SWT karena berkat Rahmat dan Karunia-Nya penulis dapat menyelesaikan laporan penelitian ini dengan lancar tanpa ada kendala yang berarti. Tak lupa juga penulis ucapkan terimakasih kepada :

1. Bapak Khairul Amri, ST, MT selaku Dekan Fakultas Teknik Universitas Bengkulu.
2. Ibu Desi Andreswari selaku Kaprodi Teknik Informatika dan anggota peneliti yang telah meluangkan waktu untuk berdiskusi dalam mengerjakan penelitian ini.
3. Azhar Firdaus selaku programmer yang ikut membantu dalam membangun aplikasi.
4. Suami dan anak-anakku tercinta yang telah rela memberi ruang dan waktu bagi penulis dalam menyelesaikan penelitian ini.

Dan semua pihak yang telah membantu yang tidak dapat penulis sebutkan satu persatu.

Semoga laporan penelitian ini dapat berguna bagi institusi dan pihak-pihak yang membacanya. Tidak lupa penulis mohon maaf jika ada kata-kata yang kurang berkenan dan mohon kritik dan saran yang membangun untuk perbaikan di waktu yang akan datang.

Bengkulu Desember 2013

Penulis

Created with

DAFTAR ISI

	Hal
Halaman Pengesahan.....	i
Ringkasan	ii
Kata Pengantar.....	iii
Daftar Isi.....	iv
BAB I. PENDAHULUAN	1
BAB II. TINJAUAN PUSTAKA	3
BAB III. TUJUAN DAN MANFAAT PENELITIAN	10
BAB IV. METODE PENELITIAN	11
BAB V. HASIL DAN PEMBAHASAN	15
BAB VI. RENCANA TAHAPAN BERIKUTNYA	21
BAB VII. KESIMPULAN DAN SARAN	22
DAFTAR PUSTAKA	23
LAMPIRAN	

BAB I

PENDAHULUAN

Mudahnya masyarakat memperoleh informasi dewasa ini tak lepas dari kemajuan teknologi informasi yang berkembang. Berbagai bentuk informasi dapat diperoleh diantaranya informasi berupa dokumen teks, gambar, film dan musik. Kemudahan yang ditawarkan oleh perkembangan teknologi sering sekali disalahgunakan oleh seseorang atau beberapa orang dalam menyelesaikan pekerjaan. Salah satu tindakan penyalahgunaan yang terjadi adalah melakukan penjiplakan terhadap seluruh atau beberapa teks dari satu atau beberapa berkas dokumen teks sumber ke berkas dokumen teks lain.

Plagiarisme adalah tindakan penjiplakan atau pengambilan karangan, pendapat, dan sebagainya dari orang lain dan menjadikannya seolah karangan dan pendapat sendiri (KBBI. 1997: 775). Tindakan penjiplakan bukan suatu hal yang baru, adanya alat bantu seperti komputer dan *laptop*, memberikan kemudahan bagi seseorang untuk melakukan tindakan penjiplakan. Hal tersebut berdampak timbulnya perilaku yang terkesan malas dalam mengerjakan suatu tugas tertentu.

Kemajuan di bidang TIK juga berdampak positif dalam membantu menyelesaikan suatu permasalahan yang berbasis komputer. Untuk itu diperlukan suatu aplikasi yang dapat membantu mendeteksi plagiarisme pada dokumen teks dengan berbasis komputer.

Untuk menentukan kemiripan antara dua objek didalam *text mining* terdapat beberapa metode, diantaranya Knut Morris Pratt (KMP) (Utomo, Darmawan., dkk.). KMP dapat diterapkan dalam menentukan nilai kemiripan pada dua berkas dokumen teks. KMP memanfaatkan karakter-karakter pola yang sudah diketahui ada didalam teks sampai terjadinya ketidakcocokan untuk melakukan pergeseran.

Metode Knuth Morris Pratt tidak dapat menentukan kesamaan makna dari setiap kata. Setiap kata yang memiliki komponen huruf yang berbeda dianggap kata yang berbeda. Algoritma Nazief & Adriani dapat mengenal kesamaan makna dari setiap kata, dimana kata yang memiliki sufiks, prefiks dan konfiks diubah menjadi bentuk dasar. Sehingga kata aktif & pasif dan kata bersinonim dapat dikenali karena kesamaan maknanya. Selain itu, algoritma Nazief & Adriani dapat mengurangi jumlah kata yang nantinya akan dihitung menggunakan metode KMP.

Berdasarkan uraian diatas maka penulis tertarik untuk membangun suatu “**Aplikasi Pendeteksi Plagiarisme Pada Dokumen Teks**”.

BAB II

TINJAUAN PUSTAKA

3.1 *Text mining*

Text mining adalah turunan data mining yaitu suatu proses menganalisis teks untuk mengekstrak informasi yang berguna untuk tujuan tertentu (Ian H. Witten: 2001.1). Penambangan teks memiliki tujuan dan menggunakan proses yang sama dengan penambangan data, namun memiliki masukan yang berbeda. Masukan untuk penambangan teks adalah data yang tidak terstruktur, seperti dokumen Word, PDF, kutipan teks, dll, sedangkan masukan untuk penambangan data adalah data yang terstruktur.

Langkah-langkah pada text mining adalah sebagai berikut :

1) *Case Folding* dan *Tokenizing*

Case folding adalah mengubah semua huruf dalam dokumen menjadi huruf kecil. Hanya huruf a sampai dengan z yang diterima. Karakter selain huruf dihilangkan dan dianggap delimiter. Tahap *tokenizing/parsing* adalah tahap pemotongan string input berdasarkan tiap kata yang menyusunnya.

2) *Filtering*

Filtering adalah tahap mengambil kata-kata penting dari hasil token. Terdapat beberapa algoritma dalam filtering yaitu stoplist dan wordlist. Algoritma *stoplist* merupakan algoritma yang digunakan untuk mengeliminasi kata-kata yang tidak deskriptif. Algoritma *wordlist* adalah algoritma yang digunakan untuk menyimpan kata-kata memiliki nilai deskriptif.

3) *Stemming*

Stemming adalah proses untuk menggabungkan atau memecahkan setiap varian-varian suatu kata menjadi kata dasar. Proses *stemming* pada Bahasa Indonesia berbeda dengan *stemming* pada teks berbahasa Indonesia. Proses *stemming* pada Bahasa Inggris adalah proses

untuk mengeliminasi sufiks pada kata sementara proses *stemming* pada Bahasa Indonesia adalah proses untuk mengeliminasi sufiks, prefiks dan konfiks. Terdapat beberapa algoritma dalam *stemming*, antara lain algoritma Porter dan algoritma Nazief & Adriani.

4) *Analyzing*

Tahap *Analyzing* merupakan tahap penentuan seberapa jauh kemiripan antar dokumen teks. Terdapat beberapa metode untuk menentukan kemiripan antar dokumen teks antara lain metode *Eucliden Distance*, metode Rabin Karp, metode *Cosine Similarity*, metode *Jaccard Coefficient*, metode *Person Correlation Coefficient* dan metode *Average Kullback-Leibler Divergence*. Metode tersebut menggunakan persamaan matematika dalam menentukan nilai kemiripan antar *file* dokumen teks.

3.2 Algoritma Nazief & Adriani

Algoritma Nazief dan Mirna Adriani ini memiliki tahap-tahap sebagai berikut:(Agusta, 2009)

- 1) Cari kata yang akan distem dalam kamus. Jika ditemukan maka diasumsikan bahwa kata tersebut adalah *root word*. Maka algoritma berhenti
- 2) *Inflection Suffixes* (“-lah”, “-kah”, “-ku”, “-mu”, atau “-nya”) dibuang. Jika berupa *particles* (“-lah”, “-kah”, “-tah” atau “-pun”) maka langkah ini diulangi lagi untuk menghapus *Possesive Pronouns* (“-ku”, “-mu”, atau “-nya”), jika ada.
- 3) Hapus *Derivation Suffixes* (“-i”, “-an” atau “-kan”). Jika kata ditemukan di kamus, maka algoritma berhenti. Jika tidak maka ke langkah 3a
 - a) Jika “-an” telah dihapus dan huruf terakhir dari kata tersebut adalah “-k”, maka “-k” juga ikut dihapus. Jika kata tersebut ditemukan dalam kamus maka algoritma berhenti. Jika tidak ditemukan maka lakukan langkah 3b.
 - b) Akhiran yang dihapus (“-i”, “-an” atau “-kan”) dikembalikan, lanjut ke langkah 4.

- 4) Hapus *Derivation Prefix*. Jika pada langkah 3 ada sufiks yang dihapus maka pergi ke langkah 4a, jika tidak pergi ke langkah 4b.
- Periksa tabel kombinasi awalan-akhiran yang tidak diijinkan. Jika ditemukan maka algoritma berhenti, jika tidak pergi ke langkah 4b.
 - For $i = 1$ to 3, tentukan tipe awalan kemudian hapus awalan. Jika root word belum juga ditemukan lakukan langkah 5, jika sudah maka algoritma berhenti. Catatan: jika awalan kedua sama dengan awalan pertama algoritma berhenti.
- 5) Melakukan *Recoding*.
- 6) Jika semua langkah telah selesai tetapi tidak juga berhasil maka kata awal diasumsikan sebagai *root word*. Proses selesai.

Tipe awalan ditentukan melalui langkah-langkah berikut:

- Jika awalannya adalah: “di-”, “ke-”, atau “se-” maka tipe awalannya secara berturut-turut adalah “di-”, “ke-”, atau “se-”.
- Jika awalannya adalah “te-”, “me-”, “be-”, atau “pe-” maka dibutuhkan sebuah proses tambahan untuk menentukan tipe awalannya.
- Jika dua karakter pertama bukan “di-”, “ke-”, “se-”, “te-”, “be-”, “me-”, atau “pe-” maka berhenti.

Tabel 7.1. Kombinasi awalan akhiran yang tidak diijinkan (Agusta, 2009)

Awalan	Akhiran yang tidak diijinkan
be-	-i
di-	-an-
ke-	-i, -kan
me-	-an
se-	i-, -kan

Tabel 7.2. Jenis Awalan Berdasarkan Tipe Awalannya
(Agusta, 2009)

Tipe Awalan	Akhiran yang harus dihapus
di-	di-
Ke-	Ke-
Se-	Se-
Te-	Te-
Ter-	Ter-
Ter-luluh	Ter-

Untuk mengatasi keterbatasan pada algoritma di atas, maka ditambahkan aturan aturan dibawah ini:

- 1) Aturan untuk reduplikasi.
 - a) Jika kedua kata yang dihubungkan oleh kata penghubung adalah kata yang sama maka *root word* adalah bentuk tunggalnya, contoh : “buku-buku” *root word*-nya adalah “buku”.
 - b) Kata lain, misalnya “bolak-balik”, “berbalas-balasan, dan ”seolah-olah”. Untuk mendapatkan *root word*-nya, kedua kata diartikan secara terpisah. Jika keduanya memiliki *root word* yang sama maka diubah menjadi bentuk tunggal, contoh: kata “berbalas-balasan”, “berbalas” dan “balasan” memiliki *root word* yang sama yaitu “balas”, maka *root word* “berbalas-balasan” adalah “balas”. Sebaliknya, pada kata “bolak-balik”, “bolak” dan “balik” memiliki *root word* yang berbeda, maka *root word*-nya adalah “bolak-balik”
- 2) Tambahan bentuk awalan dan akhiran serta aturannya.

- a) Untuk tipe awalan “mem-“, kata yang diawali dengan awalan “memp-” memiliki tipe awalan “mem-”.
- b) Tipe awalan “meng-“, kata yang diawali dengan awalan “mengk-” memiliki tipe awalan “meng-”.

3.3 *Knut Morris Pratt*

Knut Morris Pratt (KMP) melakukan perbandingan karakter di teks dan karakter dipola dari kiri ke kanan. Ide dari algoritma ini adalah bagaimana dapat memanfaatkan karakter-karakter pola yang sudah diketahui ada didalam teks sampai terjadinya ketidakcocokan untuk melakukan pergeseran (Utomo, Darmawan., Dkk).

Misalkan string teks T mempunyai panjang n , indeksnya dinyatakan dengan i , serta string pola P , mempunyai panjang m , indeksnya dinyatakan dengan j . Jika terjadi ketidakcocokan terjadi di $P[j]=a$ dan $T[i+j]=b$, maka telah diketahui terdapat karakter-karakter pola yang terdapat pada teks yaitu $P[0..j-1]=T[i..i+j-1]=u$. Karakter-karakter ini dapat dimanfaatkan sehingga dapat memungkinkan melakukan pergeseran lebih jauh.

Jadi jika ketidakcocokan terjadi di $P[j]$, maka untuk menentukan besarnya pergeseran adalah dengan mencari prefix terpanjang dari pola $[0..j-1]$ yang merupakan suffix dari $P[1..j-1]$. Nilai prefix terpanjang ini akan disimpan dalam suatu table, yang disebut table next. Tabel next ini juga dinyatakan dalam array, yaitu $next[j]$.

BAB III

TUJUAN DAN MANFAAT PENELITIAN

4.1 Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian ini adalah Membangun aplikasi yang dapat mendeteksi plagiarisme pada dokumen teks.

4.2 Manfaat Penelitian

Manfaat yang ingin dicapai dalam penelitian ini adalah memberikan kemudahan dalam mendeteksi plagiarisme pada dokumen teks.

BAB IV

METODE PENELITIAN

3.1 Jenis Penelitian

Jenis penelitian yang dilakukan adalah penelitian terapan karena tujuan dari penelitian ini ialah menghasilkan sebuah aplikasi yang dapat mendeteksi adanya kemiripan antara dua dokumen teks dengan menggunakan metode Knut Morris Pratt.

Tempat dan Waktu Penelitian

Penelitian ini dilaksanakan pada bulan Juli 2013 – Desember 2013 di Universitas Bengkulu Kota Bengkulu. Peneliti berharap dengan diadakannya penelitian ini, maka akan didapatkan suatu sistem yang dapat mendeteksi kemiripan antara dua dokumen teks demi mengurangi atau mencegah plagiarisme.

Teknik Pengumpulan Data

1. Studi Pustaka

Studi Pustaka dilakukan dengan cara mempelajari teori-teori pada literatur dan buku-buku yang berhubungan dengan aplikasi yang akan dibangun dalam tugas akhir ini.

2. Studi Lapangan

Studi lapangan bertujuan mengumpulkan dokumen teks yang akan digunakan sebagai sampel untuk melakukan uji coba aplikasi.

Metode Pengembangan Sistem

Metode pengembangan sistem yang digunakan pada penelitian ini adalah metode waterfall. Metode waterfall mengusulkan pengerjaan suatu sistem dilakukan secara berurutan

Created with

atau secara linier. Jadi jika langkah satu belum dikerjakan maka tidak akan bisa melakukan pengerjaan langkah 2, dan seterusnya.

Fase – fase yang terdapat dalam model waterfall terdiri dari aktivitas sebagai berikut:

1. Analisis kebutuhan

A. Perangkat keras (*hardware*):

1) Satu unit laptop Acer Aspire One dengan spesifikasi Intel Dual Core RAM 2 GB dan kapasitas harddisk 320 GB.

2) *Mouse* atau *Touchpad*.

B. Perangkat lunak (*software*):

1) Sistem operasi Windows 7

2) Jdk1.7 dan *IDENetbeans* 7.3 sebagai perangkat lunak bantu *editor* listing program atau *source code*.

3) MySQL 5.0.2.6.

C. Data masukan

Sebagai data masukan diperlukan dua dokumen yang berfungsi sebagai dokumen master dan dokumen replika

D. Keluaran Aplikasi

Keluaran aplikasi berupa persentase kemiripan antara dokumen sumber (master) dan dokumen pembandingan(replika) dan letak kemiripan kata yang dibandingkan.

2. Desain perangkat lunak

Perancangan sistem dalam tahap ini meliputi pemodelan perangkat lunak yang akan dibangun dengan menggunakan UML (*Unified Modelling Language*) dan perancangan *database*.

3. Implementasi

Pada tahap ini, perancangan perangkat lunak direalisasikan sebagai serangkaian program yang terdiri dari *listing* program. Analisis, pendefinisian persyaratan dan perancangan perangkat lunak yang ada pada tahap sebelumnya, direalisasikan ke dalam bahasa pemrograman Java dan *database* MySQL pada sistem operasi Windows.

4. Pengujian sistem

Tahap pengujian sistem dalam penelitian ini dilakukan dengan menginputkan data sampel berupa contoh dokumen teks yang diperoleh dari pengumpulan data lalu melihat tingkat akurasi berupa persentase kemiripan antara dua dokumen teks. Batas minimum persentase kemiripan dokumen untuk menyatakan seseorang plagiat atau tidak dalam penelitian ini sebesar 10

BAB V

HASIL DAN PEMBAHASAN

5.1 Pengujian Perhitungan Kemiripan

Pengujian perhitungan persentase kemiripan antara dua dokumen teks yang diuji menggunakan persamaan (5.1) :

$$Kemiripan(\%) = \frac{\text{Jumlah kata yang sama}}{\text{Jumlah kata pada dokumen 1} + \text{Jumlah kata pada dokumen 2}} \times 100\%$$

(5.1)

Persamaan 5.1 diterapkan pada aplikasi untuk menguji persentase kemiripan terhadap beberapa sampel dokumen. Keluaran dari pengujian ini berupa nilai kemiripan dan waktu proses. Nilai kemiripan dokumen teks yang dibandingkan memiliki rentang nilai 0% - 50%. Jika output nilai kemiripan antara dua dokumen teks memiliki persentase 50%, maka 100% terjadi plagiarisme terhadap salah satu dokumen teks (dokumen pembanding). Tabel 5.1 merupakan sampel yang digunakan dalam pengujian aplikasi:

Tabel 5.1 Sampel pengujian kemiripan dokumen teks

No	Kode	Dokumen
1	D1	Perkembangan teknologi informasi berkembang dengan pesat dan memberikan dampak positif. Salah satu dampak positif perkembangan teknologi informasi berdampak pada mudahnya pertukaran informasi. Kemudahan yang ditawarkan oleh perkembangan teknologi sering sekali disalahgunakan oleh seseorang atau beberapa orang dalam menyelesaikan pekerjaan. Penyalahgunaan teknologi informasi ini merupakan salah satu dampak negatif dari perkembangan teknolog informasi. Tindakan penyalahgunaan sering terjadi dikalangan akademis terutama mahasiswa. Salah satu tindakan penyalahgunaan yang terjadi adalah melakukan penjiplakan terhadap seluruh atau beberapa teks dari satu atau beberapa berkas dokumen teks sumber ke berkas dokumen teks lain. Praktis penyalahgunaan ini sering terjadi pada proses pembuatan tugas mata kuliah, tugas praktikum dan tugas akhir.
2	D2	Perkembangan teknologi informasi berkembang dengan pesat dan memberikan dampak positif. Salah satu dampak positif perkembangan teknologi informasi berdampak pada mudahnya pertukaran informasi. Kemudahan yang ditawarkan oleh perkembangan teknologi sering sekali disalahgunakan oleh

Created with

		seseorang atau beberapa orang dalam menyelesaikan pekerjaan. Salah satu tindakan penyalahgunaan yang terjadi adalah melakukan penjiplakan terhadap seluruh atau beberapa teks dari satu atau beberapa berkas dokumen teks sumber ke berkas dokumen teks lain. Praktis penyalahgunaan ini sering terjadi pada proses pembuatan tugas mata kuliah, tugas praktikum dan tugas akhir. Penyalahgunaan teknologi informasi ini merupakan salah satu dampak negatif dari perkembangan teknolog informasi. Tindakan penyalahgunaan sering terjadi dikalangan akademis terutama mahasiswa.
3	D3	Perkembangan teknologi informasi berkembang dengan pesat dan memberikan dampak positif. Salah satu dampak positif perkembangan teknologi informasi berdampak pada mudahnya pertukaran informasi. Salah satu tindakan penyalahgunaan yang terjadi adalah melakukan penjiplakan terhadap seluruh atau beberapa teks dari satu atau beberapa berkas dokumen teks sumber ke berkas dokumen teks lain. Praktis penyalahgunaan ini sering terjadi pada proses pembuatan tugas mata kuliah, tugas praktikum dan tugas akhir. Penyalahgunaan teknologi informasi ini merupakan salah satu dampak negatif dari perkembangan teknolog informasi. Tindakan penyalahgunaan sering terjadi dikalangan akademis terutama mahasiswa.
4	D4	Perkembangan teknologi memberikan dampak positif dan negatif. Dampak positif berupa mudahnya pertukaran informasi. Kemudahan tersebut berdampak pada penyalahgunaan. Penyalagunaan tersebut merupakan dampak negatif yang diberikan oleh perkembangan teknologi. Salah satu praktek penyalahgunaan adalah melakukan penjiplakan terhadap seluruh atau sebagian dari satu atau beberapa dokumen teks.
5	D5	Perkembangan teknologi informasi berkembang dengan pesat dan memberikan dampak positif. Salah satu dampak positif perkembangan teknologi informasi adalah kemudahan dalam bertukar informasi. Kemudahan tersebut sering disalahgunakan oleh seseorang atau beberapa orang dalam penyelesaian pekerjaan. Penyalahgunaan tersebut merupakan salah satu dampak negatif dari perkembangan teknolog informasi. Penyalahgunaan ini sering terjadi di lingkungan akademis terutama oleh mahasiswa. Salah satu tindakan penyalahgunaan yang terjadi adalah melakukan penjiplakan terhadap seluruh atau beberapa teks dari satu atau beberapa berkas dokumen teks sumber ke berkas dokumen teks lain. Praktis penyalahgunaan ini sering terjadi pada proses pembuatan tugas mata kuliah, tugas praktikum dan tugas akhir. Plagiat adalah pengambilan karangan orang lain dan disiarkan sebagai karangan sendiri. Tindakan pengambilan karangan atau ide bukan suatu hal yang baru bagi mahasiswa. Adanya alat bantu seperti komputer dan <i>laptop</i> , memberikan kemudahan untuk melakukan tindakan penjiplakan. Hal tersebut berdampak pada perilaku yang terkesan malas dalam mengerjakan tugas mata kuliah, tugas praktikum dan tugas akhir. Hal tersebut juga berpengaruh pada tahap evaluasi hasil pembelajaran. Peserta mata kuliah yang tidak sedikit memberikan kesulitan bagi para pengajar untuk memberikan hasil yang objektif. Hal tersebut menimbulkan kecurigaan terhadap tugas-tugas yang nantinya akan

		diakumulasikan ke tahap nilai akhir. Sehingga para pengajar akan direpotkan untuk menganalisis satu per satu tugas mahasiswa dikarenakan jumlah peserta mata kuliah tidak sedikit. Beberapa pengajar masih menggunakan cara manual untuk menentukan kemiripan suatu dokumen teks. Cara yang dimaksud adalah membandingkan dua atau lebih berkas dokumen dan menelusuri paragraf per paragraf dan menentukan kata, kalimat atau paragraf yang memiliki kemiripan. Cara tersebut merupakan cara yang kurang efektif dan efisien.
--	--	--

Keterangan

D1 : Dokumen teks asli

D2 : Dokumen D1 yang diubah letak kalimatnya.

D3 : Dokumen D1 yang dipenggal beberapa kalimatnya.

D4 : Dokumen yang memiliki satu topik dengan Dokumen D1

D5 : Dokumen D1 namun jumlah kata yang lebih banyak

Dokumen D1 merupakan dokumen teks asli sementara untuk Dokumen D2, D2, D3 dan D4 merupakan modifikasi kata dari dokumen D1. Dokumen D2 adalah Dokumen D1 yang diubah letak kalimatnya. Dokumen D3 adalah dokumen yang dipenggal beberapa kalimatnya. Dokumen D4 adalah dokumen yang memiliki satu topik terhadap dokumen D1. Dokumen D5 merupakan dokumen D1 yang ditambah beberapa kalimat. Tabel 5.2 hasil percobaan perhitungan kemiripan menggunakan aplikasi:

Tabel 5.2 Hasil perhitungan kemiripan

	Dokumen Sumber	Dokumen Pemanding	Hasil Perhitungan (%)	Waktu Proses (detik)
1	D1	D2	50	1,56
2	D1	D3	49,08	1,527
3	D1	D4	40,48	1,498
4	D1	D5	27,27	1,668
Total Waktu Proses				6,253

Pada pengujian no.1 pada tabel 5.1, hasil perhitungan yang diperoleh dalam membandingkan dokumen D1 dan D2 adalah 50%. D2 merupakan dokumen D1 yang letak beberapa kalimatnya diubah namun tidak mengubah komposisi kata. Berdasarkan persamaan

Created with

5.1 nilai pembagi adalah jumlah kata pada kedua dokumen, sehingga nilai yang diperoleh bukan 100% melainkan 50%. Persamaan 5.1 menggunakan jumlah kata yang sama pada kedua dokumen sehingga tidak berpengaruh terhadap letak katanya. dapat disimpulkan bahwa aplikasi ini mampu memberikan hasil kemiripan yang sama walau dokumen yang dibandingkan mengalami perubahan letak kata atau kalimat dan tidak mengalami perubahan komposisi kata.

Pada pengujian no.2 pada tabel 5.2, hasil perhitungan yang diperoleh dalam membandingkan dokumen D1 dan D3 adalah 49.02%. D3 merupakan dokumen D1 yang beberapa kalimatnya dipenggal, sehingga mengubah komposisi kata. Perbedaan hasil antara pengujian no.1 dan no.2 karena terdapat perbedaan komposisi kata antara Dokumen D1 dan D3. Kesimpulan yang diperoleh bahwa perbedaan komposisi kata antar dua dokumen mempengaruhi hasil perhitungan kemiripan pada aplikasi.

Pada pengujian no.3 pada tabel 5.2, hasil perhitungan yang diperoleh dalam membandingkan dokumen D1 dan D4 adalah 40,48%. Dokumen D1 dan D4 memiliki topik yang sama namun memiliki beberapa perbedaan kata. komposisi kata pada keduanya tentu berbeda namun kedua dokumen ini memiliki kesamaan makna. Kesimpulan yang diperoleh bahwa walaupun kedua dokumen memiliki makna yang sama namun tidak memiliki komposisi kata yang sama, hasil perhitungan yang diperoleh belum tentu bernilai 100%.

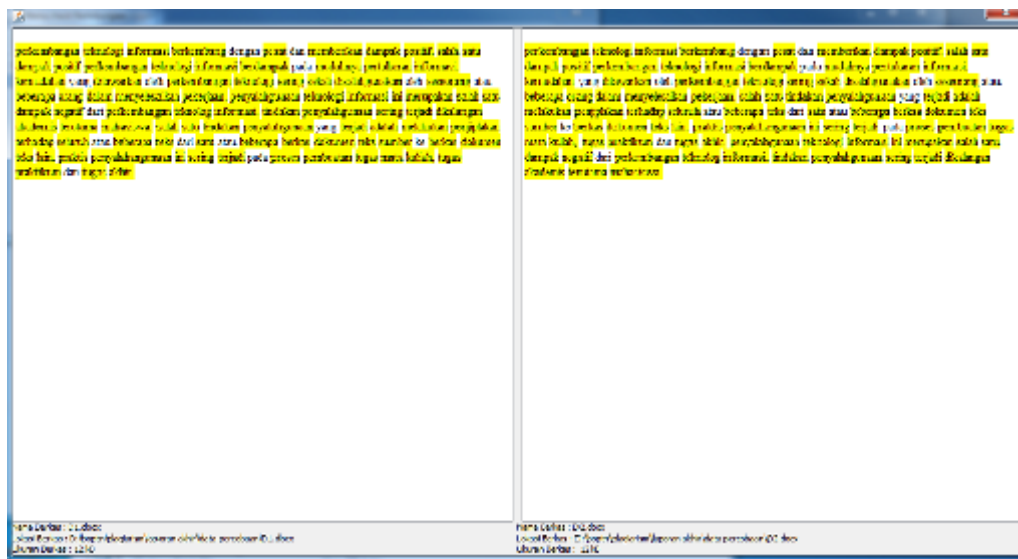
Pada pengujian no.4 pada tabel 5.2, hasil yang diperoleh dalam membandingkan dokumen D1 dan D5 adalah 27,27%. Dokumen D1 dan D5 memiliki beberapa baris kata yang sama namun D5 memiliki jumlah yang banyak dari D1. Letak kesamaan kata-kata antara D1 dan D5 terdapat pada awal paragraf namun berbeda pada kata-kata selanjutnya pada D5.

Berdasarkan percobaan no.1 pada tabel 5.1 dokumen D1 dan D2 memiliki kemiripan 100% yang membedakan adalah tata letak dari kata yang digunakan. Dengan menggunakan

persamaan 5.1 hasil kemiripan yang diperoleh adalah 50% bukan 100%, hal ini disebabkan karena perhitungan persentase kemiripan diperoleh dengan menjumlahkan kata yang sama antara dokumen D1 dan D2 dibagi dengan jumlah total kata yang terdapat pada D1 dan D2.

Dengan menggunakan batas minimum persentase kemiripan 10% dan berdasarkan pengujian no. 1, no 2, no. 3 dan no.4 maka terbukti adanya plagiarisme terhadap dokumen sumber (D1)

Gambar 5.1 merupakan user interface yang memperlihatkan letak kata yang memiliki kesamaan dengan berdasarkan pengujian no.1 yaitu membandingkan D1 dengan D2. Kata-kata yang sama ditandai dengan warna kuning.



Gambar 5.1 User Interface yang memperlihatkan letak kata yang memiliki kesamaan

5.2 Pengamatan Waktu Pengujian

Hasil pengujian pada tabel 5.2 didasarkan pada hasil keluaran aplikasi. Selama proses membandingkan berkas dokumen teks tidak terdapat aplikasi yang berjalan selain aplikasi ini. Berdasarkan pengujian pada tabel 5.2, total waktu yang dibutuhkan untuk membandingkan empat pasang dokumen teks selama 6,253 detik. Waktu yang dibutuhkan untuk membandingkan setiap pasang dokumen teks masing-masing kurang dari 2 detik.

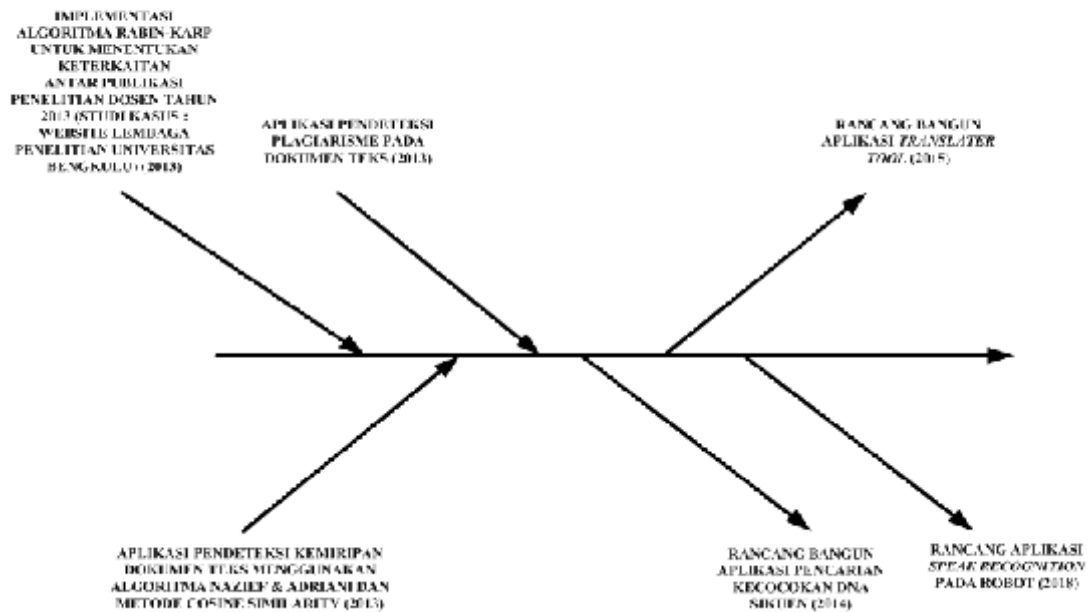
BAB VI

RENCANA TAHAPAN BERIKUTNYA

Created with



DIAGRAM FISHBONE PENELITIAN



Penelitian ini merupakan penelitian yang terkait dengan bidang ilmu *text mining* dan berdasarkan pengujian yang telah dilakukan, aplikasi yang dibangun cukup berhasil untuk mendeteksi plagiarisme. Selanjutnya Penelitian akan dikembangkan untuk bidang bioinformatika dimana *text mining* diterapkan untuk mencari kecocokan DNA dengan menggunakan pola dari DNA sekuens.

BAB VII

KESIMPULAN DAN SARAN

Created with

7.1 KESIMPULAN

Dari penelitian ini diperoleh kesimpulan :

1. Aplikasi yang dibangun telah berhasil mendeteksi plagiarisme.
2. Dalam melakukan deteksi kemiripan antara 4 dokumen teks diperoleh total waktu 6,253 detik.
3. Melalui aplikasi yang dibangun juga dapat dilihat letak kemiripan kata yang sama

7.2 SARAN

Aplikasi yang dibangun hanya bisa mendeteksi kata-kata yang sama pada dokumen yang berbeda., tetapi belum mampu untuk mendeteksi makna yang sama dari kata yang berbeda. Kedepan diharapkan aplikasi yang dibangun juga mampu mendeteksi kata-kata yang memiliki kesamaan makna.

DAFTAR PUSTAKA

Created with

Agusta, Ledy. Perbandingan Algoritma Stemming Porter Dengan Algoritma Nazief & Adriani Untuk Stemming Dokumen Teks Bahasa Indonesia ?. [Online] Tersedia :<http://yudiagusta.files.wordpress.com/2009/11/196-201-knsi09-036-perbandingan-algoritma-stemming-porter-dengan-algoritma-nazief-adriani-untuk-stemming-dokumen-teks-bahasa-indonesia.pdf> [17 Februari 2013].

Utomo, Darmawan., Harja, Eric W., Handoko., Perbandingan Algoritma String Searching Brute Force, Knuth Morris Pratt, Boyer Moore, Dan Karp Rabin Pada Teks Alkitab Bahasa Indonesia.

KBBI. 1997: 775.

Witten, Ian H. 2001. Adaptive Text Mining: Inferring Structure from Sequences ?. [Online] Tersedia :<http://www.cs.waikato.ac.nz/~ihw/papers/01IHW-Adaptivetextmining.pdf>. [16 Februari 2013]

Created with

