

**IMPLEMENTASI ALGORITMA SMITH WATERMAN  
UNTUK PENGECEKAN KESALAHAN EJAAN  
KEYWORD QUERY**

**Skripsi**



Oleh:

**FRISKA MARTHIN**

**22094764**

PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNOLOGI INFORMASI  
UNIVERSITAS KRISTEN DUTA WACANA  
YOGYAKARTA

2013

**IMPLEMENTASI ALGORITMA SMITH WATERMAN  
UNTUK PENGECEKAN KESALAHAN EJAAN  
KEYWORD QUERY**

Skripsi



©  
Diajukan kepada Program Studi Teknik Informatika Fakultas Teknologi Informasi  
Universitas Kristen Duta Wacana  
Sebagai Salah Satu Syarat dalam Memperoleh Gelar  
Sarjana Komputer

**Disusun oleh:**  
**FRISKA MARTHIN**  
**22094764**

PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNOLOGI INFORMASI  
UNIVERSITAS KRISTEN DUTA WACANA  
YOGYAKARTA

2013

## PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

### **IMPLEMENTASI ALGORITMA SMITH WATERMAN UNTUK PENGECEKAN KESALAHAN EJAAN KEYWORD QUERY**

yang saya kerjakan untuk melengkapi sebagian persyaratan menjadi Sarjana Komputer pada pendidikan Sarjana Program Studi Teknik Informatika Fakultas Teknologi Informasi Universitas Kristen Duta Wacana, bukan merupakan tiruan atau duplikasi dari skripsi kesarjanaan di lingkungan Universitas Kristen Duta Wacana maupun di Perguruan Tinggi atau instansi manapun, kecuali bagian yang sumber informasinya dicantumkan sebagaimana mestinya.

Jika dikemudian hari didapati bahwa hasil skripsi ini adalah hasil plagiasi atau tiruan dari skripsi lain, saya bersedia dikenai sanksi yakni pencabutan gelar kesarjanaan saya.

Yogyakarta, 24 Mei 2013



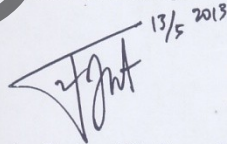
FRISKA MARTHIN  
22094764

## HALAMAN PERSETUJUAN

Judul Skripsi : IMPLEMENTASI ALGORITMA SMITH WATERMAN  
UNTUK PENGECEKAN KESALAHAN EJAAN  
KEYWORD QUERY  
Nama Mahasiswa : FRISKA MARTHIN  
N I M : 22094764  
Matakuliah : Skripsi (Tugas Akhir)  
Kode : TIW276  
Semester : Gasal  
Tahun Akademik : 2012/2013

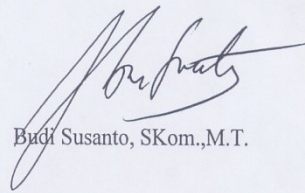
Telah diperiksa dan disetujui di  
Yogyakarta,  
Pada tanggal 13 Mei 2013

Dosen Pembimbing I



Antonius Rachmat C., Skom.,M.Cs

Dosen Pembimbing II



Budi Susanto, SKom.,M.T.

**HALAMAN PENGESAHAN**

**IMPLEMENTASI ALGORITMA SMITH WATERMAN UNTUK  
PENGECEKAN KESALAHAN EJAAN KEYWORD QUERY**

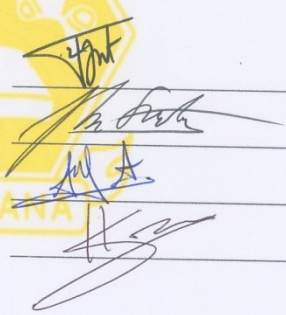
Oleh: FRISKA MARTHIN / 22094764

Dipertahankan di depan Dewan Penguji Skripsi  
Program Studi Teknik Informatika Fakultas Teknologi Informasi  
Universitas Kristen Duta Wacana - Yogyakarta  
Dan dinyatakan diterima untuk memenuhi salah satu syarat memperoleh gelar  
Sarjana Komputer  
pada tanggal 20 Mei 2013

Yogyakarta, 24 Mei 2013  
Mengesahkan,

Dewan Penguji:

1. Antonius Rachmat C., SKom., M.Cs
2. Budi Susanto, SKom., M.T.
3. Aditya Wikan Mahastama, S.Kom
4. Junius Karel, M.T.




Dekan



(Drs. Wimmie Handiwidjojo, MIT.)

Ketua Program Studi



(Nugroho Agus Haryono, M.Si)

## UCAPAN TERIMA KASIH

Puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa, karena rahmat dan anugerah serta penyertaan-Nya yang sempurna, sehingga penulis dapat menyelesaikan Tugas Akhir dengan judul Implementasi Algoritma *Smith Waterman* untuk Pengecekan Kesalahan Ejaan Keyword Query.

Penulisan laporan ini merupakan kelengkapan dan pemenuhan dari salah satu syarat dalam memperoleh gelar Sarjana Komputer. Selain itu bertujuan dalam melatih mahasiswa untuk dapat menghasilkan suatu karya ilmiah yang dapat dipertanggungjawabkan dan bermanfaat bagi penggunaannya.

Dalam proses penyelesaian dan pembuatan program serta laporan Tugas Akhir ini, penulis telah banyak menerima bimbingan, masukan, saran dan dukungan dari berbagai pihak, baik secara langsung maupun tidak langsung. Untuk itu, dengan segala kerendahan hati, penulis menyampaikan ucapan terima kasih kepada :

1. Bapak Antonius Rachmat C., S.Kom., M.Cs selaku dosen pembimbing I yang telah banyak memberikan bimbingan, petunjuk, saran, masukan dan dukungan dengan sabar dan hangat kepada penulis dari awal pembuatan Tugas Akhir sampai akhir.
2. Bapak Budi Susanto, S.Kom. M.T. selaku dosen pembimbing II dan juga selaku koordinator TA yang telah banyak memberikan bimbingan, petunjuk, saran, masukan dan dukungan dengan sabar dan hangat kepada penulis dari awal pembuatan Tugas Akhir sampai akhir.
3. Dosen-dosen Universitas Kristen Duta Wacana yang juga telah membantu memberikan bimbingan, petunjuk, saran, masukan dan dukungan kepada penulis.
4. Papa, Mama dan Keluarga tercinta yang tak henti-hentinya memberikan doa, dukungan, semangat, motivasi dan dorongan baik secara materil maupun moril sehingga Tugas Akhir ini dapat terselesaikan dengan baik dan tepat waktu.

5. Para sahabat tercinta : Putri, Fanny, Rieka, Ria, Ela', Rosie, Stepdes, Anton, Gothak, Eveline, Ratih, Ari, Lina, Siska dan para sahabat lainnya yang selalu setia dan tulus memberikan doa, dukungan, semangat, motivasi dan dorongan sehingga Tugas Akhir ini dapat terselesaikan dengan baik dan tepat waktu.
6. Para teman seperjuangan yang sama-sama mengambil Tugas Akhir terkhusus angkatan 2009, yang berjuang bersama untuk menyelesaikan Tugas Akhir ini dengan saling memotivasi dan mengingatkan serta saling memberikan dukungan.
7. Para saudara-saudari tercinta "JM" (Hezron, Steven, Rista, Keken, Orlando, Abhi, Devis, Kristo, Ety, Dian, Morgan, Ryon, Alvin, Ika) yang selalu setia dan tulus memberikan doa, dukungan, semangat, motivasi dan dorongan selama Tugas Akhir ini berlangsung.
8. Para sahabat "Data Skrip" (Putri, Nina, Gaby, Nike, Agnes) yang senantiasa memberikan doa, dukungan dan semangat selama pengerjaan Tugas Akhir ini berlangsung.
9. Pihak lain yang tidak dapat penulis sebutkan satu per satu, sehingga Tugas Akhir ini dapat terselesaikan dengan baik dan tepat waktu.

Penulis menyadari bahwa program dan laporan dalam Tugas Akhir ini masih jauh dari pada kata sempurna. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang dapat membangun dari para pembaca sekalian, sehingga dalam waktu kedepan, penulis dapat memberikan karya yang lebih baik lagi.

Akhir kata, penulis ingin meminta maaf yang sebesar-besarnya apabila dalam proses pembuatan program dan penyusunan laporan Tugas Akhir ini terselip kata-kata yang salah maupun kesalahan yang pernah penulis lakukan. Semoga Tugas Akhir ini dapat berguna dan bermanfaat bagi kita semua.

Yogyakarta, 14 Mei 2013

Penulis

## INTISARI

### Implementasi Algoritma *Smith Waterman* untuk Pengecekan Kesalahan Ejaan Keyword Query

Dalam sistem pencarian dokumen, salah satu yang menjadi kesalahan pencarian adalah inputan user. Kesalahan dapat disebabkan oleh kesalahan pengetikan karena ketidaktelitian user saat menginputkan kata kunci pencarian ataupun karena kesalahan literasi yaitu kesalahan karena kata kunci yang diinputkan user saat melakukan pencarian dokumen tidak tepat.

Melihat latar belakang di atas, penulis membangun sebuah sistem pencarian yang mampu melakukan pengecekan terhadap kesalahan penulisan kata kunci yang diinputkan user saat akan melakukan pencarian dokumen dengan menggunakan algoritma *Smith Waterman*.

Sistem pengecekan kesalahan penulisan kata kunci menggunakan algoritma *Smith Waterman* pada sistem pencarian yang telah dibangun ini menghasilkan tingkat kebenaran kata yang cukup tinggi terhadap hasil uji coba beberapa kata kunci, yakni mencapai 86.648%.

Kata Kunci: pengecekan, kesalahan penulisan, *Smith Waterman*



## DAFTAR ISI

HALAMAN SAMPUL .....	
PERNYATAAN KEASLIAN SKRIPSI.....	<b>Error! Bookmark not defined.</b>
HALAMAN PERSETUJUAN.....	<b>Error! Bookmark not defined.</b>
HALAMAN PENGESAHAN.....	<b>Error! Bookmark not defined.</b>
UCAPAN TERIMA KASIH.....	v
INTISARI.....	viii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xii
DAFTAR GAMBAR.....	xiv
BAB 1 PENDAHULUAN .....	1
1.1 Latar Belakang Masalah.....	1
1.2 Perumusan Masalah.....	2
1.3 Batasan Masalah.....	2
1.4 Tujuan Penelitian.....	3
1.5 Metode Penelitian.....	3
1.6 Sistematika Penulisan.....	4
BAB 2 TINJAUAN PUSTAKA DAN LANDASAN TEORI.....	5
2.1 Tinjauan Pustaka .....	5
2.2 Landasan Teori.....	6
2.2.1 Frase.....	6
2.2.2 Text Mining.....	6
2.2.3 Information Retrieval.....	8
2.2.4 TF-IDF .....	9
2.2.5 Vektor Space Model.....	10
2.2.6 Algoritma Smith Waterman .....	12
2.2.7 Evaluasi Sistem.....	15
BAB 3 PERANCANGAN SISTEM.....	19
3.1 Spesifikasi Kebutuhan Non Fungsional dan Fungsional.....	19
3.1.1 Spesifikasi Kebutuhan Non Fungsional .....	19

3.1.2	Spesifikasi Kebutuhan Fungsional Kebutuhan Perangkat Keras dan Perangkat Lunak .....	21
3.2	Perancangan Proses Sistem .....	23
3.2.1	Flowchart .....	23
3.3	Kamus Data .....	34
3.3.1	Tabel Dokumen .....	34
3.3.2	Tabel Token .....	34
3.3.3	Tabel Bobot .....	35
3.3.4	Tabel BiGram .....	35
3.3.5	Tabel Stopword .....	36
3.3.6	Tabel User .....	36
3.4	Diagram Skema .....	37
3.5	Rancangan Antarmuka Sistem .....	38
3.5.1	Rancangan Antarmuka Halaman Awal Sistem .....	38
3.5.2	Rancangan Antarmuka Halaman Hasil Pencarian Dokumen Abstrak 39	
3.5.3	Rancangan Antarmuka Halaman Abstrak .....	40
3.5.4	Rancangan Antarmuka Halaman Login Admin .....	40
3.5.5	Rancangan Antarmuka Halaman Awal Setelah Login .....	41
3.5.6	Rancangan Antarmuka Halaman Update Dokumen .....	42
3.5.7	Rancangan Antarmuka Halaman Tambah Stopword .....	43
3.6	Rancangan Pengujian Sistem .....	43
3.6.1.	Pengujian Hasil Rekomendasi Perbaikan Kata .....	43
3.6.2.	Pengujian Hasil Pencarian Dokumen .....	45
4.1	Implementasi Sistem .....	47
4.1.1	Implementasi Proses Tokenisasi .....	47
4.1.2	Implementasi Proses Perhitungan Nilai TF-IDF .....	48
4.1.3	Implementasi Proses Perhitungan Normalisasi Nilai TF-IDF .....	49
4.1.4	Implementasi Proses Bigram Kata .....	50
4.1.5	Implementasi Proses Perhitungan Nilai Chi-Square .....	51
4.1.6	Implementasi Proses Perhitungan Nilai Similarity Smith Waterman 53	

4.1.7	Implementasi Proses Pencarian Vector Space Model .....	56
4.2	Implementasi Antarmuka .....	59
4.2.1	Halaman Utama - User .....	59
4.2.2	Halaman Hasil Pencarian-User .....	61
4.2.3	Halaman Selengkapnya .....	61
4.2.4	Halaman Login Admin .....	62
4.2.5	Halaman Utama Sistem-Admin .....	62
4.2.6	Halaman Hasil Perhitungan Similarity Smith Waterman-Admin ...	63
4.2.7	Halaman Hasil Perhitungan Chi-Square-Admin .....	64
4.2.8	Halaman Hasil Pencarian-Admin .....	64
4.2.9	Halaman Detail Hasil Pencarian-Admin .....	65
4.2.10	Halaman Lihat Dokumen-Admin .....	66
4.2.11	Halaman Tambah Dokumen-Admin .....	67
4.2.12	Halaman Tambah Stopword-Admin .....	68
4.3	Analisa Sistem .....	68
4.3.1	Analisa Terhadap Hasil Rekomendasi Perbaikan Kata .....	68
4.3.2	Analisa Terhadap Hasil Pencarian Dokumen .....	88
BAB 5 KESIMPULAN DAN SARAN .....		103
5.1.	Kesimpulan .....	103
5.2.	Saran .....	103
DAFTAR PUSTAKA .....		103
LAMPIRAN .....		103

## DAFTAR TABEL

Tabel 2. 1 Contoh hasil perhitungan TF-IDF .....	10
Tabel 2. 2 Contoh hasil perhitungan Vector Space Model .....	12
Tabel 2. 2 Contoh hasil perhitungan Vector Space Model .....	12
Tabel 3. 1 Tabel Dokumen .....	34
Tabel 3. 2 Tabel Token .....	35
Tabel 3. 3 Tabel Bobot .....	35
Tabel 3. 4 Tabel Bigram .....	36
Tabel 3. 5 Tabel Stopword .....	36
Tabel 3. 6 Tabel User .....	36
Tabel 3. 7 Kata-kata untuk Pengujian Hasil Rekomendasi Perbaikan Kata .....	44
Tabel 3. 8 Tabel Query dan Dokumen yang Relevan .....	46
Tabel 4. 1 Pseudo-code Proses Tokenisasi .....	47
Tabel 4. 2 Pseudo-code Proses Perhitungan Nilai TF-IDF .....	48
Tabel 4. 3 Pseudo-code Proses Perhitungan Normalisasi Nilai TF-IDF .....	50
Tabel 4. 4 Pseudo-code Proses Bigram Kata .....	50
Tabel 4. 5 Pseudo-code Proses Perhitungan Nilai Chi-Square .....	52
Tabel 4. 6 Pseudo-code Proses Perhitungan Nilai Similarity Smith Waterman .....	54
Tabel 4. 7 Pseudo-code Proses Pencarian Vector Space Model .....	57
Tabel 4. 8 Pengujian Kata “case” .....	69
Tabel 4. 9 Pengujian Kata “file” .....	70
Tabel 4. 10 Pengujian Kata “akses” .....	71
Tabel 4. 11 Pengujian Kata “biaya” .....	72
Tabel 4. 12 Pengujian Kata “cepat” .....	73
Tabel 4. 13 Pengujian Kata “analisa” .....	74
Tabel 4. 14 Pengujian Kata “ketelitian” .....	76
Tabel 4. 15 Pengujian Kata “perancangan” .....	79
Tabel 4. 16 Pengujian Kata “memperhitungkan” .....	81
Tabel 4. 17 Pengujian Kata “mempertimbangkan” .....	83
Tabel 4. 18 Persentasi Tingkat Kebenaran Kata yang Diuji .....	87
Tabel 4. 19 Hasil Pengujian Query Pertama .....	88

Tabel 4. 20 Hasil Pengujian Query Kedua.....	90
Tabel 4. 21 Hasil Pengujian Query Ketiga.....	91
Tabel 4. 22 Hasil Pengujian Query Keempat.....	92
Tabel 4. 23 Hasil Pengujian Query Kelima.....	93
Tabel 4. 24 Hasil Pengujian Query Keenam.....	94
Tabel 4. 25 Hasil Pengujian Query Ketujuh.....	96
Tabel 4. 26 Hasil Pengujian Query Kedelapan.....	97
Tabel 4. 27 Hasil Pengujian Query Kesembilan.....	98
Tabel 4. 28 Hasil Pengujian Query Kesepuluh.....	100
Tabel 4. 29 Rata-rata Hasil Pengujian.....	101

©UKDW

## DAFTAR GAMBAR

Gambar 2. 1 Nilai kesamaan yang mungkin dari 2 sekuens : Perbandingan sekuens (b) merupakan kemungkinan yang terbaik .....	13
Gambar 2. 2 Proses membangun tabel matriks menggunakan algoritma smith waterman .....	15
Gambar 2. 3 Gambar tabel kontingensi proses precision dan recall .....	16
Gambar 2. 4 Gambar kurva precision / recall untuk ranked retrieved .....	17
Gambar 2. 5 Gambar kurva averaged 11-point precision / recall .....	18
Gambar 3. 1 Diagram Use Case .....	21
Gambar 3. 2 Flowchart Dokumen Training .....	24
Gambar 3. 3 Subflowchart Perhitungan Nilai TF-IDF .....	25
Gambar 3. 4 Subflowchart Perhitungan Normalisasi Nilai TF-IDF .....	26
Gambar 3. 5 Flowchart Bi-Gram .....	28
Gambar 3. 6 Subflowchart Perhitungan Nilai Chi-Square .....	29
Gambar 3. 7 Flowchart Input Kata .....	30
Gambar 3. 8 Subflowchart Perhitungan Smith Waterman .....	31
Gambar 3. 9 Subflowchart Backtraking Smith Waterman .....	32
Gambar 3. 10 Flowchart Sistem Pencarian Vector Space Model .....	33
Gambar 3. 11 Diagram Skema .....	37
Gambar 3. 12 Rancangan Antarmuka Halaman Awal Sistem .....	38
Gambar 3. 13 Rancangan Antarmuka Halaman Hasil Pencarian Dokumen Abstrak .....	39
Gambar 3. 14 Rancangan Antarmuka Halaman Abstrak .....	40
Gambar 3. 15 Rancangan Antarmuka Halaman Login Admin .....	40
Gambar 3. 16 Rancangan Antarmuka Halaman Awal Setelah Login .....	41
Gambar 3. 17 Rancangan Antarmuka Halaman Update Dokumen .....	42
Gambar 3. 18 Rancangan Antarmuka Halaman Tambah Stopword .....	43
Gambar 4. 1 Implementasi Halaman Utama - User .....	59
Gambar 4. 2 Implementasi Halaman Utama – User (Rekomendasi Perbaikan Kata) .....	60
Gambar 4. 3 Implementasi Halaman Utama – User (Rekomendasi Pasangan Kata) .....	60
Gambar 4. 4 Implementasi Halaman Hasil Pencarian Dokumen – User .....	61

Gambar 4. 5 Implementasi Halaman Selengkapnya dari Tiap Dokumen – User	61
Gambar 4. 6 Implementasi Halaman Login - Admin	62
Gambar 4. 7 Implementasi Halaman Utama – Admin	62
Gambar 4. 8 Implementasi Halaman Hasil Perhitungan Similarity Smith Waterman – Admin	63
Gambar 4. 9 Implementasi Halaman Hasil Perhitungan Frekuensi Chi-Square – Admin	64
Gambar 4. 10 Implementasi Halaman Hasil Pencarian Dokumen – Admin	64
Gambar 4. 11 Implementasi Halaman Detail Hasil Pencarian Dokumen – Admin	65
Gambar 4. 12 Implementasi Halaman Lihat Dokumen – Admin	66
Gambar 4. 13 Implementasi Halaman Ubah Data Dokumen – Admin	66
Gambar 4. 14 Implementasi Halaman Tambah Dokumen – Admin	67
Gambar 4. 15 Implementasi Halaman Tambah Stopword – Admin	68
Gambar 4. 16 Grafik Interpolasi	102

© UKYD

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

Dalam sistem pencarian dokumen, salah satu yang menjadi kesalahan pencarian adalah inputan user. Kesalahan dapat disebabkan karena kesalahan pengetikan ataupun karena kesalahan literasi. Kesalahan pengetikan dapat disebabkan karena ketidakteelitian user saat menginputkan kata kunci pencarian. Sedangkan kesalahan literasi dapat disebabkan karena kata kunci yang diinputkan user saat melakukan pencarian dokumen tidak tepat. Rekomendasi pasangan kata yang biasa disebut juga frase dapat membantu user memilih kata kunci yang dapat digunakan saat melakukan pencarian dokumen. Frase merupakan kelompok kata atau gabungan dua kata atau lebih yang membentuk satu kesatuan.

Ada beberapa algoritma yang dapat digunakan untuk mendeteksi kemiripan, beberapa diantaranya yaitu algoritma *Levenshtein* dan algoritma *Smith Waterman*. Algoritma *Smith Waterman* memiliki kemiripan dengan algoritma *Levenshtein*. Algoritma *Levenshtein* mengukur jarak kemiripan antara dua string, sedangkan algoritma *Smith Waterman* mencari pencocokan terbaik (*best matching*) dengan memeriksa setiap kemungkinan kemiripan antara dua string yang dibandingkan, memberikan nilai terhadap karakter yang berbeda (*mismatch*), kesenjangan antara karakter (*gap*) dan karakter yang sama (*match*). Algoritma *Smith-Waterman* biasa disebut sebagai *local sequence alignment* yaitu metode untuk menyelaraskan *segmen* dari dua *sekuens* dengan tingkat kemiripan yang maksimal. Sedangkan untuk memberikan rekomendasi pasangan kata yang umum digunakan dalam dokumen digunakan perhitungan frekuensi *chi square*. Perhitungan frekuensi *chi square* dapat menguji keterikatan dua komponen yang berdampingan.

Berdasarkan permasalahan yang telah dijelaskan pada paragraph sebelumnya, maka perlu untuk membuat suatu sistem pencarian yang mampu melakukan pengecekan terhadap kesalahan penulisan kata kunci yang diinputkan user saat akan melakukan pencarian dokumen. Selain mengecek kesalahan



penulisan, sistem juga dapat menampilkan rekomendasi perbaikan kata terhadap kata kunci yang diinputkan. Tidak hanya menampilkan rekomendasi perbaikan kata, sistem juga akan menampilkan rekomendasi pasangan kata yang umum digunakan. Pasangan kata yang direkomendasikan didasarkan pada kata kunci yang diinputkan.

## 1.2 Perumusan Masalah

Berdasarkan latar belakang masalah diatas, akan diuji bagaimana tingkat kebenaran algoritma *Smith Waterman* dalam merekomendasikan kata yang memiliki kemiripan dengan kata yang diinputkan user, serta bagaimana sistem menghasilkan dokumen hasil pencarian yang relevan dengan menggunakan algoritma *Vector Space Model*.

## 1.3 Batasan Masalah

Adapun batasan sistem yang akan dibuat dalam penelitian ini adalah sebagai berikut :

- Dokumen yang digunakan merupakan dokumen abstraksi tugas akhir mahasiswa Teknik Informatika Universitas Kristen Duta Wacana yang dikumpulkan secara manual dari situs <http://sinta.ukdw.ac.id/sinta/search.jsp?query=prodi:22>.
- Kata yang akan dicek hanya kata yang memiliki minimal panjang karakter tiga (3), untuk kata-kata yang hanya terdiri dari satu (1) atau dua (2) karakter tidak akan dibandingkan dalam kamus kata.
- Kesalahan tata bahasa (*Grammar*) akan diabaikan karena sistem hanya akan memeriksa kesalahan penulisan.

#### 1.4 Tujuan Penelitian

Penelitian ini bertujuan untuk menguji seberapa tingkat kebenaran algoritma *Smith Waterman* dalam merekomendasikan kata-kata yang memiliki kemiripan dengan kata yang diinputkan user. Selain itu penelitian ini juga bertujuan untuk mengurangi terjadinya kesalahan pencarian saat user melakukan pencarian dokumen. Seperti yang dijelaskan sebelumnya bahwa salah satu yang menjadi kesalahan pencarian adalah inputan user yang dapat disebabkan oleh salah pengetikan ataupun kesalahan literasi.

#### 1.5 Metode Penelitian

Metode yang digunakan dalam penelitian ini :

1. Studi Pustaka

Studi pustaka dilakukan dengan mempelajari teori-teori melalui buku, artikel, jurnal dan bahan lain yang berkaitan dengan metode pengecekan ejaan dan pencarian dokumen.

2. Perancangan Sistem

Tahap ini berisi perancangan basis data dan perancangan antarmuka untuk sistem yang akan dibangun.

3. Pembangunan Sistem

Tahap ini merupakan tahap pembuatan program sesuai dengan rancangan yang telah dibuat.

4. Implementasi dan Testing

Pada tahap ini dilakukan pengujian terhadap program, dengan menampilkan kata-kata sebagai hasil pengecekan kata serta menampilkan dokumen-dokumen berdasarkan query yang diinputkan.

5. Analisis Hasil Percobaan dan Evaluasi

Setelah dilakukan pengujian terhadap program, tahap selanjutnya adalah menganalisis keefektifan program dalam melakukan menampilkan kata-kata yang disarankan oleh sistem.

## **1.6 Sistematika Penulisan**

Sistematika penulisan laporan tugas akhir ini dibagi dalam 5 Bab yang didalamnya terbagi lagi menjadi beberapa bagian yang merupakan satu kesatuan dan saling terkait.

### **Bab 1 Pendahuluan**

Berisi latar belakang masalah, perumusan masalah, batasan masalah, tujuan penelitian, metode penelitian, dan sistematika penulisan dari tugas akhir ini.

### **Bab 2 Landasan Teori**

Berisi teori-teori yang mendasari penelitian dalam tugas akhir ini.

### **Bab 3 Perancangan Sistem**

Berisi perancangan sistem yang akan dibangun meliputi alur kerja sistem, antar muka sistem, dan perancangan basis data pada sistem.

### **Bab 4 Implementasi Sistem**

Berisi metode penelitian yang digunakan beserta langkah-langkah serta hasil penelitian. Dalam bab ini diketahui lebih lanjut mengenai kelebihan dan kekurangan sistem.

### **Bab 5 Kesimpulan dan Saran**

Berisi hal-hal penting yang dapat diambil dari penelitian dan saran pengembangan guna penelitian berikutnya.

## BAB 5

### KESIMPULAN DAN SARAN

#### 5.1. Kesimpulan

Berikut ini adalah beberapa kesimpulan yang dapat diambil berdasarkan analisis terhadap hasil pengujian terhadap sistem :

- Dari hasil pengujian pengecekan kesalahan penulisan kata menggunakan algoritma *smith waterman* oleh sistem diperoleh nilai tingkat kebenaran mencapai 86.648%.
- Sedangkan untuk pengujian terhadap sistem pencarian dengan menggunakan *vector space model* didapat nilai MAP (Mean Average Precision) 0.828, yang berarti sistem dapat mengembalikan dokumen yang relevan dengan baik sesuai dengan query yang diinputkan user.

#### 5.2. Saran

Saran yang diajukan penulis untuk pengembangan dan perbaikan sistem adalah :

- Perlu adanya perbaikan algoritma untuk mempercepat proses pencocokan karakter, sehingga user tidak terlalu lama menunggu saat proses pencarian rekomendasi kata berlangsung.
- Sistem dapat melakukan pencarian terhadap dokumen selain abstrak yaitu bab 1, bab 2, bab 3, bab 4, dan bab 5.
- Sistem rekomendasi kata dan pencarian dapat diintegrasikan ke blog ataupun website lain.

## DAFTAR PUSTAKA

- David A.grossman., & Ophir Frieder (2007). *Information Retrieval Algorithms and Heuristics Second Edtion*.Chicago, IL, U.S.A : Illinois Institute of Technology.
- Feldman, R., & Sanger, J. (2007). *The Text Mining Handbook*. New York: Cambridge University Press.
- Irving, Robert W. (2004). Plagiarism and Collusion Detection using the Smith-Waterman Algorithm. Diakses tanggal 11 Januari 2013 dari <http://www.dcs.gla.ac.uk/publications/PAPERS/7444/TR-2004-164.pdf>
- Ryan, Mudafiq Pratama., Eko, Budi Cahyono., Gita, Indah Marthasari. *Aplikasi Pendeteksi Duplikasi Dokumen Teks Bahasa Indonesia Menggunakan Algoritma Winnowing dengan Metode K-Gram dan Synonym Recognition*. Diakses tanggal 05 Desember 2012 dari <http://mudafiqriyan.com/2012/03/aplikasi-pendeteksi-duplikasi-dokumen-teks-bahasa-indonesia-menggunakan-algoritma-winnowing-dengan-metode-k-gram-dan-synonym-recognition/>
- Saeed, Ali Khajeh., Poole, Stephen., Perot, Blair J. (2010). Acceleration of The Smith-Waterman Algorithm Using Single and Multiple Graphics Processors, *Journal of Computational Physics*. Vol.229, Elsevier Inc, 4247-4258.
- Shehab, Sara A., Keshk, Arabi., Mahgoub, Hany. (2012). Fast Dynamic Algorithm for Sequence Alignment based on Bioinformatics, *International Journal of Computer Applications*. Vol.3, No.7, 54-61.
- Su, Zhan., Ahn, Byung-Ryul., Eom, Ki-Yol., Kang, Min-Koo., Kim, Jin-Pyung., Kim, Moon-Kyun. (2008). Plagiarism Detection Using the Levenshtein Distance and Smith-Waterman Algorithm. Diakses tanggal 04 Maret 2013 dari <http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=4603758&url=http%>

3A%2F%2Fieeexplore.ieee.org%2Fstamp%2Fstamp.jsp%3Ftp%3D%26arn  
umber%3D4603758

Thalib Farid., Ratih Kusumawati. (2010). Pembuatan Program Aplikasi untuk  
Pendeteksian Kemiripan Dokumen Teks dengan Algoritma Smith-  
Waterman. Diakses tanggal 05 Desember 2012 dari  
[http://repository.gunadarma.ac.id/bitstream/123456789/2383/1/03-01-004-  
Pembuatan%5BFarid%5D.pdf](http://repository.gunadarma.ac.id/bitstream/123456789/2383/1/03-01-004-Pembuatan%5BFarid%5D.pdf)

©UKDW