

**Pengaruh *Synonym Recognition* Dalam Sistem Deteksi Kemiripan Teks
Menggunakan Algoritma *Needleman-Wunsch***

*Diajukan Sebagai Syarat Untuk Menyelesaikan
Pendidikan Program Strata-1 Pada
Jurusan Teknik Informatika*



Oleh :

Anggi Juwita
NIM : 09021181520035

**JURUSAN TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2020**

LEMBAR PENGESAHAN TUGAS AKHIR

PENGARUH *SYNONYM RECOGNITION* DALAM SISTEM DETEKSI KEMIRIPAN TEKS MENGGUNAKAN ALGORITMA *NEEDLEMAN-WUNSCH*

Oleh :

ANGGI JUWITA
NIM : 09021181520035

Indaralaya, Juli 2020

Pembimbing I,

Novi Yusliani, M.T
NIP. 198211082012122001

Pembimbing II,

Osvari Arselan, M.T
NIP. 198805282018031001

Mengetahui,
Ketua Jurusan Teknik Informatika



Rifkie Primartha, M.T
NIP. 197706012009121004

TANDA LULUS UJIAN SIDANG TUGAS AKHIR

Pada hari tanggal 06 Juli 2020 telah dilaksanakan ujian sidang tugas akhir oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Anggi Juwita
NIM : 09021181520035
Judul : Pengaruh *Synonym Recognition* Dalam Sistem Deteksi Kemiripan Teks Menggunakan Algoritma *Needleman-Wunsch*

1. Pembimbing I

Novi Yusliani, M.T
NIP. 198211082012122001



2. Pembimbing II

Osvari Arsalan, M.T
NIP. 198806282018031001



3. Penguji I

M. Fachrurrozi, M.T
NIP. 198005222008121002



4. Penguji II

Kanda Januar Miraswan, M.T
NIK. 1671080901900006



Mengetahui,

Ketua Jurusan Teknik Informatika



Rifkie Primārtha, M.T
NIP. 197706012009121004

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama : Anggi Juwita
NIM : 092021181520035
Program Studi : Teknik Informatika
Judul Skripsi : Pengaruh *Synonym Recognition* Dalam Sistem
Deteksi Kemiripan Teks Menggunakan
Algoritma *Needleman-Wunsch*

Hasil Pengecekan *Software iThenticate/Turnitin* : 7%

Menyatakan bahwa Laporan Proyek saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan proyek ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian, pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.



Indralaya, Juli 2020



(Anggi Juwita)
NIM. 09021181520035

MOTTO DAN PERSEMBAHAN

Manusia bisa menyusun rencana dalam bayangannya, namun hanya Allah Subhana Wa Ta'ala yang mampu memberikan rencana pasti dan jalan terbaiknya.

“Pertolongan Allah tak pernah telat hanya saja datang di waktu yang tepat”

“Berdo'alah kepada Ku, pastilah Aku kabulkan untukmu”
(Q.S. Al-Mukmin: 60)

Kupersembahkan karya tulis ini kepada:

- ♥ Allah Subhana wa Ta'ala
- ♥ Kedua OrangTua ku tercinta
- ♥ Kakak dan Adikku tersayang
- ♥ Keluarga Besarku
- ♥ Sahabat dan teman-teman seperjuanganku
- ♥ Dosen Pembimbingku
- ♥ Dosen dan Staff Jurusan Teknik Informatika
- ♥ Almamaterku Universitas Sriwijaya

**THE EFFECT OF SYNONYM RECOGNITION IN THE TEXT
SIMILARITY DETECTION SYSTEM USING THE *NEEDLEMAN-
WUNSCH* ALGORITHM**

**By:
Anggi Juwita
09021181520035**

ABSTRACT

Detection of similarities in the Final Project becomes very important because it can be used to avoid plagiarism behavior. Similarity detection can be done automatically using string matching techniques with the *Needleman-Wunsch* algorithm to determine the percentage of similarity in a document. With the introduction of the concept of synonyms using Synonym Recognition is intended to recognize words that contain synonyms in the text that allegedly plagiarism of the source text. The test is carried out by comparing the similarity percentage of the similarity value of the two documents and the length of the process carried out using the *Needleman-Wunsch* algorithm with Synonym Recognition and without Synonym Recognition. The percentages obtained from the two types of tests produced the highest similarity values of 98,718% and 95,633%. The results of the lowest processing time with Synonym Recognition reached an average value of 15,648 seconds, which is longer than without using synonym recognition with an average processing time of 7.83 seconds. The more terms contained in a synonym dictionary, the higher the percentage of similarities produced. The more words in the document text, the longer the processing time needed.

Keywords: Detection of Text Similarity, Needleman-Wunsch, Percentage of Similarity, Processing Time, Synonym Recognition

**PENGARUH *SYNONYM RECOGNITION* DALAM SISTEM DETEKSI
KEMIRIPAN TEKS MENGGUNAKAN ALGORITMA
*NEEDLEMAN-WUNSCH***

**Oleh:
Anggi Juwita
09021181520035**

ABSTRAK

Pendeteksian kemiripan pada Tugas Akhir menjadi sangat penting karena dapat digunakan untuk menghindari perilaku plagiarisme. Pendeteksian kemiripan dapat dilakukan secara otomatis menggunakan teknik *string matching* dengan algoritma *Needleman-Wunsch* untuk mengetahui persentase kemiripan suatu dokumen. Dengan pengenalan konsep sinonim menggunakan *Synonym Recognition* dimaksudkan untuk mengenali kata-kata yang mengandung sinonim pada teks yang diduga plagiat terhadap teks sumber. Pengujian dilakukan dengan membandingkan persentase kemiripan dari nilai kemiripan kedua dokumen dan lama proses yang dilakukan menggunakan algoritma *Needleman-Wunsch* dengan *Synonym Recognition* dan tanpa *Synonym Recognition*. Persentase yang diperoleh dari kedua jenis pengujian tersebut menghasilkan nilai kemiripan tertinggi sebesar 95.355% dan 74.145%. Hasil waktu proses terendah dengan *Synonym Recognition* mencapai rata-rata nilai sebesar 15.648 detik yang lebih lama dibandingkan tanpa menggunakan *synonym recognition* dengan rata-rata lama proses 7.83 detik. Semakin banyak *term* yang terdapat dalam kamus sinonim, semakin tinggi persentase kemiripan yang dihasilkan. Semakin banyak kata dalam teks dokumen, semakin lama waktu proses yang dibutuhkan.

Kata Kunci: Deteksi Kemiripan Teks, *Needleman-Wunsch*, Persentase Kemiripan, Waktu Proses dan *Synonym Recognition*

KATA PENGANTAR

Puji syukur kepada Allah Subhanawa Ta'ala atas berkat dan rahmat-Nya yang telah diberikan kepada Penulis sehingga dapat menyelesaikan Tugas Akhir ini dengan baik. Tugas akhir yang berjudul **“Pengaruh *Synonym Recognition* Dalam Sistem Deteksi Kemiripan Teks Menggunakan Algoritma *Needleman-Wunsch*”** disusun untuk memenuhi salah satu syarat guna menyelesaikan pendidikan program Strata-1 pada Fakultas Ilmu Komputer Program Studi Teknik Informatika di Universitas Sriwijaya.

Dalam menyelesaikan Tugas Akhir ini banyak pihak yang telah memberikan bantuan dan dukungan baik secara langsung maupun secara tidak langsung. Penulis ingin menyampaikan rasa terima kasih kepada:

1. Kedua OrangTua ku, Bapak Agus Susanto dan Ibu Masula yang selalu bersabar, setia mendo'akan, memberikan motivasi, kasih sayang yang sangat tulus serta memberikan dukungan baik moril maupun materil.
2. Kakakku, Angga Sanjaya, S.Sos dan Adikku, Putri Hamida yang selalu mendukung dan menjadi penyemangat untuk segera menyelesaikan Tugas Akhir ini.
3. Keluarga besarku yang turut memberikan nasihat, motivasi dan kritikkan selama menjalankan masa perkuliahan hingga sampai kini.
4. Bapak Jaidan Jauhari, S.Pd, M.T. selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
5. Bapak Rifkie Primartha, S.T., M.T. selaku Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

6. Ibu Anggina Primanita, M.T dan Bapak Ali Ibrahim, M.T selaku dosen pembimbing akademik yang senantiasa memberikan masukan dan inspirasi baik dalam bidang akademik maupun non akademik
7. Ibu Novi Yusliani, M.T selaku dosen pembimbing I dan Bapak Osvari Arsalan, M.T selaku pembimbing II, yang telah banyak memberikan bimbingan, masukan, dan bantuan dalam proses menyelesaikan Tugas Akhir ini.
8. Bapak M. Fachrurrozi, M.T dan Bapak Kanda Januar Miraswan, M.T selaku dosen penguji yang telah memberikan koreksi dan masukan pada Tugas Akhir penulis.
9. Seluruh Bapak dan Ibu Dosen Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya yang telah memberikan banyak ilmu pengetahuannya selama kegiatan akademik berlangsung.
10. Mbak Winda, Kak Ricy serta seluruh Staf Tata Usaha yang telah membantu dalam kelancaran proses administrasi dan akademik selama penulis menjalankan masa perkuliahan.
11. Okta, Elin, Syazili dan partner di Jurusan Teknik Informatika yang selalu siap mendengarkan keluh kesahku dan membuatku tau arti untuk tetap tersenyum meskipun dalam kondisi apapun. Terimakasih buat canda tawamu yang ceria, my sweet friends.
12. Lili Arista Repi, S. Si., Koni, Yolak, Depi, teman sekaligus saudara perempuan dari masa kecil yang selalu memberikan semangat dan do'a tanpa mengenal kesibukan masing-masing. Terima kasih atas perhatian, motivasi

dan do'anya semoga semua cita-cita kita diijabah oleh Allah Subhana Wa Ta'ala, Aamiin Ya Rabbal Alaamiin.

13. Teman satu bimbingan Tugas Akhir yang senantiasa saling berbagi informasi dan masukan selama proses bimbingan yang dilaksanakan oleh penulis.

Semua orang yang telah terlibat dan bekerjasama dalam penyusunan tugas akhir ini yang tidak dapat disebutkan namanya satu persatu.

Semoga Allah Subhana Wa Ta'ala selalu melimpahkan berkah, rahmat dan hidayah-Nya kepada kita semua. Saya menyadari masih banyak sekali kesalahan pada laporan Tugas Akhir ini, semua ini disebabkan karena keterbatasan pengetahuan dan kemampuan saya, maka saya dengan senang hati menerima saran dan kritik yang bersifat membangun dari pembaca, dan berharap laporan ini bermanfaat bagi kita semua. Aamiin Ya Rabbal Alaamiin.

Indralaya, Juli 2020

Anggi Juwita

Daftar Isi

	Halaman
HALAMAN JUDUL	i
LEMBAR PENGESAHAN SKRIPSI	ii
TANDA LULUS UJIAN SIDANG SKRIPSI.....	iii
HALAMAN PERNYATAAN.....	iv
MOTTO DAN PERSEMBAHAN.....	v
ABSTRACT	vi
ABSTRAK	vii
KATA PENGANTAR	viii
DAFTAR ISI	xi
DAFTAR TABEL	xv
DAFTAR GAMBAR	xviii
DAFTAR LAMPIRAN.....	xx

BAB I PENDAHULUAN

<u>1.1 Pendahuluan</u>	I-1
<u>1.2 Latar Belakang</u>	I-1
<u>1.3 Rumusan Masalah</u>	I-3
<u>1.4 Tujuan Penelitian</u>	I-4
<u>1.5 Manfaat Penelitian</u>	I-4
<u>1.6 Batasan Masalah</u>	I-5
<u>1.7 Sistematika Penulisan</u>	I-5
<u>1.8 Kesimpulan</u>	I-7

BAB II KAJIAN LITERATUR

<u>2.1 Pendahuluan</u>	II-1
<u>2.2 Landasan Teori</u>	II-1
<u>2.2.1 Plagiarisme</u>	II-1

<u>2.2.2 Preprocessing</u>	II-2
<u>2.2.3 Synonym Recognition</u>	II-5
<u>2.2.4 Algoritma Needleman-Wunsch</u>	II-6
<u>2.2.5 Penelitian Lain Yang Relevan</u>	II-11
<u>2.3 Kesimpulan</u>	II-13

BAB III METODOLOGI PENELITIAN

<u>3.1 Pendahuluan</u>	III-1
<u>3.2 Pengumpulan Data</u>	III-1
<u>3.2.1 Jenis dan Sumber Data</u>	III-1
<u>3.2.2 Metode Pengumpulan Data</u>	III-1
<u>3.3 Tahapan Penelitian</u>	III-2
<u>3.3.1 Menetapkan Kerangka Kerja/Framework</u>	III-2
<u>a. Prapengolahan</u>	III-2
<u>b. Menghitung Hasil Kemiripan dengan Needleman- Wunsch</u>	III-5
<u>3.3.2 Menetapkan Kriteria Pengujian</u>	III-8
<u>3.3.3 Menetapkan Format Data Pengujian</u>	III-7
<u>3.3.4 Menetapkan Alat yang digunakan dalam Pelaksanaan Penelitian</u>	III-9
<u>3.3.5 Melakukan Pengujian Penelitian</u>	III-10
<u>3.3.6 Melakukan Analisis Hasil Pengujian dan Membuat Kesimpulan</u>	III-11
<u>3.4 Metode Pengembangan Perangkat Lunak</u>	III-12
<u>3.4.1 Fase Insepsi</u>	III-13
<u>3.4.2 Fase Elaborasi</u>	III-13
<u>3.4.3 Fase Konstruksi</u>	III-14
<u>3.4.4 Fase Transisi</u>	III-14
<u>3.5 Manajemen Proyek Penelitian</u>	III-15
<u>3.6 Kesimpulan</u>	III-15

BAB IV PENGEMBANGAN PERANGKAT LUNAK

<u>4.1</u>	<u>Pendahuluan</u>	IV-1
<u>4.2</u>	<u>Fase Insepsi</u>	IV-1
4.2.1	<u>Pemodelan Bisnis</u>	IV-1
4.2.2	<u>Kebutuhan Sistem</u>	IV-3
4.2.2.1	<u>Fitur Cek Kemiripan</u>	IV-4
4.2.3	<u>Analisis dan Desain</u>	IV-4
4.2.3.1	<u>Analisis Kebutuhan Perangkat Lunak</u>	IV-4
4.2.3.2	<u>Analisis Data</u>	IV-6
4.2.3.3	<u>Analisis <i>Praprocessing</i></u>	IV-8
4.2.3.4	<u>Analisis Pendeteksian Sinonim menggunakan <i>Synonym Recognition</i></u>	IV-10
4.2.3.5	<u>Analisis Algoritma <i>Needleman-Wunsch</i></u>	IV-12
4.2.3.6	<u>Desain Perangkat Lunak</u>	IV-46
<u>4.3</u>	<u>Fase Elaborasi</u>	IV-53
4.3.1	<u>Pemodelan Bisnis</u>	IV-54
4.3.2	<u>Perancangan Data</u>	IV-54
4.3.3	<u>Perancangan Antarmuka Sistem</u>	IV-54
4.3.4	<u>Kebutuhan Sistem</u>	IV-56
4.3.5	<u>Diagram Aktivitas (<i>Activity Diagram</i>)</u>	IV-56
4.3.6	<u>Diagram Alur (<i>Sequence Diagram</i>)</u>	IV-60
<u>4.4</u>	<u>Fase Konstruksi</u>	IV-63
4.4.1	<u>Kebutuhan Sistem</u>	IV-63
4.4.2	<u>Diagram Kelas (<i>Class Diagram</i>)</u>	IV-63
4.4.2.1	<u>Implementasi</u>	IV-65
4.4.2.2	<u>Implementasi Kelas</u>	IV-65
4.4.2.3	<u>Implementasi Antarmuka</u>	IV-68
<u>4.5</u>	<u>Fase Transisi</u>	IV-69
4.5.1	<u>Pemodelan Bisnis</u>	IV-69
4.5.2	<u>Kebutuhan Sistem</u>	IV-69
4.5.3	<u>Rencana Pengujian</u>	IV-70

<u>4.5.3.1 Rencana Pengujian Use Case Memilih File</u>	
<u>Sumber</u>	IV-70
<u>4.5.3.2 Rencana Pengujian Use Case Memilih File</u>	
<u>Banding</u>	IV-71
<u>4.5.3.2 Rencana Pengujian Use Case Mengecek</u>	
<u>Kemiripan</u>	IV-71
<u>4.5.4 Implementasi</u>	IV-72
<u>4.5.4.1 Pengujian Use case Memilih File Sumber</u>	IV-73
<u>4.5.4.2 Pengujian Use case Memilih File Banding</u>	IV-74
<u>4.5.4.3 Pengujian Use case Mengecek Kemiripan</u>	IV-76
<u>4.6 Kesimpulan</u>	IV-77

BAB V HASIL DAN ANALISIS PENELITIAN

<u>5.1 Pendahuluan</u>	V-1
<u>5.2 Data Hasil Penelitian</u>	V-1
<u>5.2.1 Konfigurasi Percobaan</u>	V-1
<u>5.2.2 Data Hasil Konfigurasi I</u>	V-2
<u>5.2.3 Data Hasil Konfigurasi II</u>	V-6
<u>5.2.4 Data Hasil Konfigurasi III</u>	V-7
<u>5.2.5 Analisis Data Pengujian</u>	V-7
<u>5.3 Kesimpulan</u>	V-11

BAB VI KESIMPULAN DAN SARAN

<u>6.1 Pendahuluan</u>	VI-1
<u>6.2 Kesimpulan</u>	VI-1
<u>6.3 Saran</u>	VI-2

<u>DAFTAR PUSTAKA</u>	xxi
-----------------------------	-----

DAFTAR TABEL

	Halaman
Tabel II-1. Contoh Teks Dokumen Sebelum Diproses	II-3
Tabel II-2. Hasil Teks Dari Proses Pre-prosesing Pada Dokumen_Sumber_A50%	II-4
Tabel II-3. Hasil Teks Dari Proses Pre-prosesing Pada Dokumen_Banding_A50%	II-4
Tabel II-4. Contoh Hasil Proses <i>Synonym Recognition</i>	II-5
Tabel II-5. Inialisasi Matrik (i,0) dan (0,j) dengan <i>Needleman-Wunsch</i>	II-9
Tabel II-6. Nilai Matrik (i,j) dengan <i>Needleman-Wunsch</i>	II-10
Tabel II-7. Hasil susunan sekuen dari proses <i>traceback</i>	II-10
Tabel III-1. Rancangan Hasil Pendeteksian Kemiripan Teks	III-8
Tabel III-2. Rancangan Tabel Analisis Pengujian Rata-Rata Hasil Pengujian	III-10
Tabel III-3. Rancangan Tabel Analisis Pengujian Persentase Maksimum	III-11
Tabel III-4. Penjadwalan Penelitian dalam Bentuk <i>Work Breakdown Structure</i>	III-15
Tabel IV-1. Kebutuhan Fungsional	IV-3
Tabel IV-2. Kebutuhan non-Fungsional	IV-3
Tabel IV-3. Contoh Teks Masukkan.....	IV-8
Tabel IV-4. Hasil Proses <i>Casefolding</i>	IV-8
Tabel IV-5. Hasil Proses <i>Tokenizing</i>	IV-9
Tabel IV-6. Hasil Proses <i>Stopword Removal</i>	IV-9
Tabel IV-7. Hasil Proses <i>Stemming</i>	IV-10
Tabel IV-8. Hasil Teks Dengan Proses <i>Synonym Recognition</i>	IV-11
Tabel IV-9. Hasil Teks Tanpa Proses <i>Synonym Recognition</i>	IV-11
Tabel IV-10. Pengecekan Kata Dalam Isi Teks Dengan Kamus Sinonim.....	IV-11
Tabel IV-11. Perubahan Kata Pada Dokumen Sumber Dan Dokumen Banding	IV-12

Tabel IV-12. Hasil Perhitungan Inisialisasi Matrik menggunakan <i>Needleman-Wunsch</i> dengan <i>Synonym Recognition</i>	IV-14
Tabel IV-13. Perhitungan Inisialisasi Matrik menggunakan <i>Needleman-Wunsch</i> Dengan <i>Synonym Recognition</i>	IV-14
Tabel IV-14. Hasil Perhitungan Needleman-Wunsch Pada Matrik M (i, j) Dengan <i>Synonym Recognition</i>	IV-15
Tabel IV-15. Perhitungan Needleman-Wunsch Pada Matrik M (i, j) Dengan <i>Synonym Recognition</i>	IV-15
Tabel IV-16. Hasil Susunan Substring Dari Proses Traceback Dengan <i>Synonym Recognition</i>	IV-30
Tabel IV-17. Hasil Perhitungan Inisialisasi Matrik menggunakan Needleman-Wunsch Tanpa <i>Synonym Recognition</i>	IV-30
Tabel IV-18. Perhitungan Inisialisasi Matrik menggunakan Needleman-Wunsch Tanpa <i>Synonym Recognition</i>	IV-30
Tabel IV-19. Hasil Perhitungan Needleman-Wunsch Pada Matrik M (i, j) Tanpa <i>Synonym Recognition</i>	IV-31
Tabel IV-20. Perhitungan Needleman-Wunsch Pada Matrik M (i, j) Tanpa <i>Synonym Recognition</i>	IV-31
Tabel IV-21. Hasil Susunan Substring Dari Proses Traceback Tanpa <i>Synonym Recognition</i>	IV-46
Tabel IV-22. Definisi Aktor	IV-47
Tabel IV-23. Definisi <i>Use Case</i>	IV-48
Tabel IV-24. Skenario <i>Use Case</i> Memilih File Sumber	IV-48
Tabel IV-25. Skenario <i>Use Case</i> Memilih File Banding	IV-50
Tabel IV-26. Skenario <i>Use Case</i> Mengecek Kemiripan	IV-52
Tabel IV-27. Implementasi Kelas	IV-65
Tabel IV-28. Rencana Pengujian <i>Use Case</i> Memilih File Sumber	IV-70
Tabel IV-29. Rencana Pengujian <i>Use Case</i> Memilih File Banding	IV-71
Tabel IV-30. Rencana Pengujian <i>Use Case</i> Mengecek Kemiripan	IV-72
Tabel IV-31. Pengujian <i>Use Case</i> Memilih File Sumber	IV-73
Tabel IV-32. Pengujian <i>Use Case</i> Memilih File Banding	IV-75

Tabel IV-33. Skenario <i>Use Case</i> Mengecek Kemiripan.....	IV-76
Tabel V-1. Hasil Deteksi Kemiripan Teks Menggunakan <i>Needleman-Wunsch</i>	V-2
Tabel V-2. Rata-Rata Persentase Kemiripan Dan Waktu Pemrosesan Dengan <i>Synonym Recognition</i> Dan Tanpa <i>Synonym Recognition</i>	V-6
Tabel V-3. Persentase maksimum dengan dan tanpa <i>Synonym Recognition</i>	V-7

DAFTAR GAMBAR

	Halaman
Gambar III-1. Diagram Proses Pra-pengolahan	III-4
Gambar III-2. Flowchart Perangkat Lunak Tanpa Synonym Recognition	III-6
Gambar III-3. Flowchart Perangkat Lunak Dengan Synonym Recognition	III-7
Gambar III-4. Diagram Tahapan Pengujian	III-10
Gambar III-5. Penjadwalan untuk Tahap Menentukan Ruang Lingkup dan Unit Penelitian	III-23
Gambar III-6. Penjadwalan untuk Tahap Menentukan Dasar Teori yang Berkaitan dengan Penelitian	III-23
Gambar III-7. Penjadwalan untuk Tahap Menentukan Kriteria Pengujian	III-24
Gambar III-8. Penjadwalan untuk Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian Fase Insepsi	III-24
Gambar III-9. Penjadwalan untuk Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian Fase Elaborasi	III-25
Gambar III-10. Penjadwalan untuk Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian Fase Konstruksi	III-26
Gambar III-11. Penjadwalan untuk Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian Fase Transisi	III-27
Gambar III-12. Penjadwalan untuk Tahap Melakukan Pengujian Penelitian	III-28
Gambar III-13. Penjadwalan untuk Tahap Analisa Hasil Pengujian Penelitian dan Membuat Kesimpulan	III-28
Gambar IV-1. Diagram Use Case	IV-47
Gambar IV-2. Perancangan Antarmuka Sistem	IV-55
Gambar IV-3. Diagram Aktivitas Memilih File Sumber	IV-57
Gambar IV-4. Diagram Aktivitas Memilih File Banding.....	IV-58
Gambar IV-5. Diagram Aktivitas Mengecek Kemiripan.....	IV-59
Gambar IV-6. Diagram <i>Sequence</i> Memilih File Sumber	IV-60
Gambar IV-7. Diagram <i>Sequence</i> Memilih File Banding	IV-61
Gambar IV-8. Diagram <i>Sequence</i> Mengecek Kemiripan	IV-62

Gambar IV-9. Diagram Kelas Perangkat Lunak.....	IV-64
Gambar IV-10. Implementasi Antarmuka.....	IV-68
Gambar V-1. Perbandingan Rata-Rata Persentase Kemiripan Menggunakan Algoritma Needleman-Wunsch	V-8
Gambar V-2. Perbandingan Rata-Rata Waktu Proses Menggunakan Algoritma Needleman-Wunsch	V-8
Gambar V-3. Perbandingan Persentase Maksimum Menggunakan Algoritma Needleman-Wunsch.....	V-10

DAFTAR LAMPIRAN

1. Teks Dokumen Penelitian
2. *Source Code* Program



ILMU ALAT PENGABDIAN

BAB I

PENDAHULUAN

1.1 Pendahuluan

Bab pendahuluan akan membahas mengenai alasan dilaksanakannya penelitian yang mengimplementasikan algoritma *Needleman-Wunsch* dan *Synonym Recognition* untuk sistem pendeteksian kemiripan dokumen. Penelitian tersebut dijelaskan dalam latar belakang, rumusan masalah, tujuan, manfaat dan batasan masalah serta menuliskan kesimpulan yang diambil dari keseluruhan isi yang disampaikan dalam sub-bab 1 pendahuluan.

1.2 Latar Belakang

Penyalinan informasi atau plagiarisme terhadap sebuah dokumen sering dilakukan di berbagai kalangan, terutama di dunia pendidikan. Beberapa hal yang mendorong plagiator melakukan tindak plagiat diantaranya disebabkan karena timbulnya rasa malas dalam berpikir, *deadline* tugas yang segera dikumpulkan bahkan juga akibat pengaruh buruk dari kemajuan teknologi (Hasan, 2016).

Perkembangan internet merupakan salah satu dampak dari kemajuan teknologi, sehingga masyarakat yang mahir dalam menggunakannya dengan mudah menyaring dan mengakses informasi yang dibutuhkan dalam penulisan karya ilmiah (Geravand and Ahmadi, 2014). Kegiatan ini menyebabkan semakin meningkatnya tindakan plagiarisme. Oleh karena itu, untuk mencegah dan

mengurangi tindak plagiarisme maka dibutuhkan metode atau algoritma untuk membangun sebuah sistem pendeteksian plagiarisme tersebut.

Algoritma *Needleman-Wunsch* merupakan algoritma yang termasuk dalam pendekatan *exact match* (Singla and Garg, 2012) yang dapat membandingkan antara dua dokumen yang diuji dalam sistem pendeteksian plagiarisme (Leonardo and Hansun, 2017). Berbeda dengan *Smith-Waterman*, algoritma *Needleman-Wunsch* melakukan pensejajaran sekuen secara *global* (Patel, 2016). Algoritma ini berdasarkan perhitungan dalam bentuk skor matrik dan biasa digunakan untuk membandingkan urutan biologis dan teks dokumen (Rachmawati, Khodra and Supriana, 2016). Algoritma ini pun sangat akurat dalam menemukan penjajaran *substring* yang optimal dengan cepat (Ray *et al.*, 2016).

Meskipun membutuhkan waktu yang sedikit lama pada perbandingan dengan jumlah *substring* yang lebih panjang, namun dalam penelitian yang dilakukan oleh (Bipin Nair, Shibin and Shamcy, 2017) membuktikan bahwa algoritma ini dapat menghasilkan nilai akurasi sebesar 89%. Penelitian yang diuji oleh (Muflikhah and Santoso, 2017) terhadap pendeteksian virus HBV dan HCC menunjukkan tingkat kemiripan yang sangat baik yaitu 96.55% dengan nilai akurasi mencapai sebesar 99.85% tanpa dilakukannya modifikasi terhadap algoritma *Needleman-Wunsch*.

Synonym Recognition dapat dimanfaatkan untuk mendeteksi kata yang bersinonim (Wibowo *et al.*, 2015). Hasil dari penelitian (Wibowo *et al.*, 2015) menunjukkan bahwa *Synonym Recognition* dapat memberikan hasil persentase kemiripan yang lebih baik dan meningkatkan nilai akurasi terhadap pengujian dalam pengembangan kalimat atau paraphrase dokumen dibandingkan tanpa

menggunakan *Synonym Recognition*. Berdasarkan penelitian (Yusuf, Fauzi and Brata, 2018) menunjukkan bahwa *Synonym Recognition* juga sangat baik dalam sistem *information retrieval*. Penelitian (Yusuf, Fauzi and Brata, 2018) memberikan hasil kenaikan pada nilai *recall* dan *f-measure* dengan masing-masing sebesar 0.594 dan 0.404 pada threshold 15 dan mencapai rata-rata nilai *precision* 0.52 dengan threshold 5.

Dari penjelasan yang telah disampaikan diatas, maka akan dilakukan penelitian untuk membangun sebuah sistem yang mendeteksi kemiripan dokumen dengan menggunakan algoritma *Needleman-Wunsch* dan *Synonym Recognition*, sehingga dapat diketahui besarnya pengaruh yang dihasilkan oleh *Synonym Recognition* dalam penelitian ini.

1.3 Rumusan Masalah

Rumusan masalah dalam penelitian ini adalah

1. Bagaimana hasil persentase kemiripan dokumen dengan menerapkan proses *Synonym Recognition* dan tanpa *Synonym Recognition* dalam sistem deteksi kemiripan teks Bahasa Indonesia menggunakan algoritma *Needleman-Wunsch*?
2. Bagaimana perbandingan waktu proses yang didapatkan dari hasil pengujian dengan mengimplementasikan *Synonym Recognition* dan tanpa *Synonym Recognition* menggunakan algoritma *Needleman-Wunsch* dalam mendeteksi kemiripan teks Bahasa Indonesia?

1.4 Tujuan Penelitian

Adapun tujuan penelitian dari rumusan masalah, yaitu untuk :

1. Mengetahui hasil persentase kemiripan yang menerapkan *Synonym Recognition* dan tanpa *Synonym Recognition* menggunakan algoritma *Needleman-Wunsch* dalam sistem pendeteksian kemiripan teks Bahasa Indonesia.
2. Mengetahui waktu proses yang dihasilkan dari sistem deteksi kemiripan teks Bahasa Indonesia yang mengimplementasikan proses *Synonym Recognition* dan tanpa *Synonym Recognition* menggunakan algoritma *Needleman-Wunsch*.

1.5 Manfaat Penelitian

Manfaat dari penelitian yang dilakukan, yaitu :

1. Mengasah kemampuan dalam membangun sistem pendeteksian kemiripan dokumen.
2. Hasil penelitian ini dapat dijadikan sebagai sarana pengetahuan dalam mengembangkan sistem pendeteksian kemiripan dokumen berbahasa Indonesia menggunakan pengenalan sinonim dengan suatu algoritma.
3. Penelitian ini sebagai bahan referensi dalam menerapkan algoritma *Needleman-Wunsch* untuk pendeteksian kemiripan dokumen bahasa Indonesia.

1.6 Batasan Masalah

Batasan masalah untuk penelitian ini adalah :

1. Data yang digunakan adalah data berupa teks Bahasa Indonesia dan daftar kata-kata bersinonim.
2. Proses pengujian hanya dilakukan terhadap data berupa huruf.
3. Sistem tidak memperbaiki kesalahan ejaan, tetapi tetap menguji terhadap data yang memiliki kesalahan dalam penulisan kata.
4. Perangkat lunak hanya menerima dan memproses masukan dokumen berekstensi *.txt*.

1.7 Sistematika Penulisan

Untuk memahami penelitian dengan jelas, maka pemaparan materi dikelompokkan menjadi beberapa Bab dengan sistematika penulisan sebagai berikut:

BAB I. PENDAHULUAN

Pada Bab pendahuluan menjabarkan mengenai latar belakang, perumusan masalah, tujuan dan manfaat penelitian, batasan masalah atau ruang lingkup serta sistematika penulisan.

BAB II. KAJIAN LITERATUR

Dalam Bab II memaparkan mengenai pembahasan dan dasar-dasar teori yang terkait dengan penelitian serta bahan materi penyusunan penelitian tugas akhir.

BAB III. METODELOGI PENELITIAN

Pada Bab ini membahas mengenai tahapan dan proses yang akan dilaksanakan. Setiap tahapan dan proses tersebut dibuat dalam bentuk kerangka kerja sehingga terdefiniskan secara jelas. Dan diakhir Bab ini disusun penjadwalan penelitian dalam bentuk tabel untuk memanejemenkan pelaksanaan penelitian.

BAB IV PENGEMBANGAN PERANGKAT LUNAK

Pada bab ini membahas mengenai proses pengembangan perangkat lunak yang digunakan sebagai alat penelitian, penulis mengembangkan perangkat lunak dengan metode pemrograman berorientasi obyek berdasarkan panduan *Rational Unified Process*. Di dalam *Rational Unified Process* terdapat empat fase pengembangan perangkat lunak yaitu fase insepisi, elaborasi, konstruksi, dan transisi. Setiap fase memiliki pemodelan bisnis, kebutuhan, analisis dan desain, implementasi, dan pengujian.

BAB V HASIL DAN ANALISIS PENELITIAN

Pada bab ini akan menyajikan kata salah ejaan beserta kandidatnya dan evaluasi perhitungan nilai MRR menggunakan perangkat lunak.

BAB VI KESIMPULAN DAN SARAN

Pada bab ini akan membahas kesimpulan dan saran dari penelitian ini.

1.8 Kesimpulan

Berdasarkan penjelasan sebelumnya, maka penelitian ini membangun sebuah sistem yang melakukan pendeteksian kemiripan pada teks berbahasa Indonesia. Dengan algoritma *Needleman-Wunsch* yang digunakan dalam penelitian ini, maka diketahui perbandingan persentase kemiripan dan waktu proses yang dihasilkan terhadap proses pengujian pada penggunaan *Synonym Recognition* dan tanpa *Synonym Recognition* dalam pengenalan kata bersinonim.

DAFTAR PUSTAKA

- Bipin Nair, B. J., Shibin, K. and Shamcy, O. (2017) ‘An hybrid method for comparing brainstem glioma sequences using needleman Wunsch and waterman smith algorithms’, in *2017 2nd International Conference for Convergence in Technology (I2CT)*. IEEE, pp. 867–872. doi: 10.1109/I2CT.2017.8226253.
- Chua, S. L. and Foo, L. K. (2017) ‘Tree alignment based on needleman-wunsch algorithm for sensor selection in smart homes’, *Sensors (Switzerland)*, 17(8). doi: 10.3390/s17081902.
- Day, R. F. (2010) ‘Examining the validity of the Needleman-Wunsch algorithm in identifying decision strategy with eye-movement data’, *Decision Support Systems*. Elsevier B.V., 49(4), pp. 396–403. doi: 10.1016/j.dss.2010.05.001.
- EZZIKOURI, H., ERRITALI, M. and OUKESSOU, M. (2017) ‘Fuzzy-Semantic Similarity for Automatic Multilingual Plagiarism Detection’, *International Journal of Advanced Computer Science and Applications*, 8(9), pp. 86–90. doi: 10.14569/ijacsa.2017.080912.
- Fakirah, M. *et al.* (2016) ‘Accelerating Needleman-Wunsch global alignment algorithm with GPUs’, *Proceedings of IEEE/ACS International Conference on Computer Systems and Applications, AICCSA*, 2016-July. doi: 10.1109/AICCSA.2015.7507113.
- Gao, L. and Chen, H. (2018) ‘An automatic extraction method based on synonym dictionary for web reptile question and answer’, *Proceedings of the 13th IEEE Conference on Industrial Electronics and Applications, ICIEA 2018*. IEEE, pp. 375–378. doi: 10.1109/ICIEA.2018.8397745.
- Geravand, S. and Ahmadi, M. (2014) ‘An efficient and scalable plagiarism checking system using Bloom filters q’, *Computers and Electrical Engineering*. Elsevier Ltd, 40(6), pp. 1789–1800. doi: 10.1016/j.compeleceng.2014.06.003.
- Gomaa, Wael H; Fahmy, A. A. (2012) ‘Short Answer Grading Using String Similarity And Corpus-Based Similarity’, *International Journal of Advanced Computer Science and Applications*, 3(11). doi: 10.14569/IJACSA.2012.031119.
- Hasan, A. (2016) ‘Fenomena Plagiarisme Mahasiswa’, *Mei*, IV(1), pp. 20–29. Available at: <https://media.neliti.com/media/publications/70308-ID-fenomena-plagiarisme-mahasiswa.pdf>.

- Jayanti, L. *et al.* (2016) 'Analisa Pola Penyalahgunaan Facebook Sebagai Alat Kejahatan Trafficking Menggunakan Data Mining', *E-journal Teknik Informatika*, 8(1), pp. 30–35. Available at: <https://ejournal.unsrat.ac.id/index.php/informatika/article/viewFile/12231/1811>.
- Latief, M., Kandowangko, N. and Yusuf, R. (2018) 'Pengembangan Sistem Informasi Tanaman Obat Daerah Gorontalo Berbasis Web dan Mobile', *Jurnal Rekayasa Elektrika*, 13(3), p. 152. doi: 10.17529/jre.v13i3.8532.
- Leonardo, B. and Hansun, S. (2017) 'Text Documents Plagiarism Detection using Rabin-Karp and Jaro-Winkler Distance Algorithms', *Indonesian Journal of Electrical Engineering and Computer Science*, 5(2), p. 462. doi: 10.11591/ijeecs.v5.i2.pp462-471.
- Malendes, M. A. and Bunyamin, H. (2017) 'Perbandingan Needleman-Wunsch dan Lempel-Ziv dalam Teknik Global Sequence Alignment: Keunggulan Faktorisasi Sempurna', 3(April), pp. 57–68.
- Muflikhah, L. and Santoso, E. (2017) 'Pairwise Sequence Alignment between HBV and HCC Using Modified Needleman Wunsch Algorithm', *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, 15(4), p. 1785. doi: 10.12928/telkomnika.v15i4.5813.
- Mutiara, A. B. and Agustina, S. (2008) 'Anti Plagiarism Application with Algorithm Karp-Rabin at Thesis in Gunadarma University', pp. 3–10. Available at: <http://arxiv.org/abs/0811.4349>.
- Rachmawati, E., Khodra, M. L. and Supriana, I. (2016) 'Shape based recognition using freeman chain code and modified Needleman-Wunsch', in *2016 8th International Conference on Information Technology and Electrical Engineering (ICITEE)*. IEEE, pp. 1–6. doi: 10.1109/ICITEED.2016.7863307.
- Sanditya Hardaya, I. B. N., Dhini, A. and Surjandari, I. (2018) 'Application of text mining for classification of community complaints and proposals', *Proceeding - 2017 3rd International Conference on Science in Information Technology: Theory and Application of IT for Education, Industry and Society in Big Data Era, ICSITech 2017*, 2018-Janua, pp. 144–149. doi: 10.1109/ICSITech.2017.8257100.
- Singla, N. and Garg, D. (2012) 'String Matching Algorithms and their Applicability in various Applications', *Internationall Journal of Soft Computing and Engineering (IJSCE)*, 1(6), pp. 218–222.

- Slamet, C. *et al.* (2018) 'Automated Text Summarization for Indonesian Article Using Vector Space Model', *IOP Conference Series: Materials Science and Engineering*, 288(1). doi: 10.1088/1757-899X/288/1/012037.
- Sunarto, A. A. and Robial, S. M. (2019) 'PARALELISASI NEEDLEMAN-WUNSCH', *Jurnal Teknoinfo*, 13(1), p. 41. doi: 10.33365/jti.v13i1.230.
- Wibowo, A. T. *et al.* (2015) 'Analsis dan Implementasi Algoritma Winnowing dengan Synonym Recognition pada Deteksi Plagiarisme untuk Dokumen Teks Berbahasa Indonesia', 2(3), pp. 7674–7683.
- Yusuf, S., Fauzi, M. A. and Brata, K. C. (2018) 'Sistem Temu Kembali Informasi Pasal-Pasal KUHP (Kitab Undang-Undang Hukum Pidana) Berbasis Android Menggunakan Metode Synonym Recognition dan Cosine Similarity Sistem Temu Kembali Informasi Pasal-Pasal KUHP (Kitab Undang- Undang Hukum Pidana) Berbasis', (February).