

**IMPLEMENTASI ALGORITMA JARO WINKLER DISTANCE UNTUK  
SISTEM PENDETEKSI SIMILARITY PADA DOKUMEN  
NASKAH PUBLIKASI**

**SKRIPSI**



disusun oleh

**Ihsan Almunawaroh**

**15.11.8846**

**PROGRAM SARJANA  
PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA  
2019**

**IMPLEMENTASI ALGORITMA JARO WINKLER DISTANCE UNTUK  
SISTEM PENDETEKSI SIMILARITY PADA DOKUMEN  
NASKAH PUBLIKASI**

**SKRIPSI**

untuk memenuhi sebagian persyaratan  
mencapai gelar Sarjana  
pada Program Studi Informatika



disusun oleh

**Ihsan Almunawaroh**

**15.11.8846**

**PROGRAM SARJANA  
PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA  
2019**

## PERSETUJUAN

### SKRIPSI

#### IMPLEMENTASI ALGORITMA JARO WINKLER DISTANCE UNTUK SISTEM PENDETEKSI SIMILARITY PADA DOKUMEN NASKAH PUBLIKASI

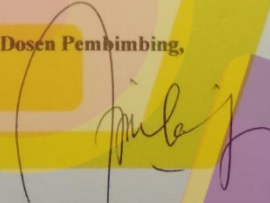
yang dipersiapkan dan disusun oleh

**Ihsan Almunawaroh**

**15.11.8846**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 21 Oktober 2019

Dosen Pembimbing,

  
**Nila Feby Puspitasari, S.Kom, M.Cs**  
**NIK. 190302161**

**PENGESAHAN**  
**SKRIPSI**  
**IMPLEMENTASI ALGORITMA JARO WINKLER DISTANCE UNTUK**  
**SISTEM PENDETEKSI SIMILARITY PADA DOKUMEN**  
**NASKAH PUBLIKASI**

yang dipersiapkan dan disusun oleh

**Ihsan Almunawaroh**

**15.11.8846**

telah dipertahankan di depan Dewan Penguji  
pada tanggal 18 November 2019

**Susunan Dewan Penguji**

**Nama Penguji**

**Tanda Tangan**

Erni Seniwati, S. Kom, M. Cs  
NIK. 190302231

Wiwi Widayani, M.Kom  
NIK. 190302272

Nila Feby Puspitasari, S.Kom, M.Cs  
NIK. 190302161



Skripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
pada tanggal 30 November 2019

  
**DEKAN FAKULTAS ILMU KOMPUTER**  
Kusnawati, S.Si, M.T.  
NIK. 190302038

## PERNYATAAN

Saya yang bertandatangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggung jawab saya pribadi.

Yogyakarta 7 Desember 2019

METERAI  
TEMPEL  
NO CD1AHF152670183  
6000  
ENAM RIBU RUPIAH  
Husan Almunawaroh  
15.11.8874

## MOTTO

“Bekerja keras dan bersikap baiklah. Hal luar biasa akan terjadi.”

(Conan O’Brien)

”Jangan Kau kerjakan besok, bila kau dapat mengerjakannya hari ini.”

(Bulek Endang)

”Aku akan perintahkan diriku dan mengatakan bahwa aku mampu! Aku akan mengalahkan keraguan, rasa takut, perasaan minder, dan menukarnya dengan keberanian.”

(Merry Riana)

”Kekuatan sejati dari umat manusia adalah bahwa kita memiliki kuasa penuh untuk mengubah diri kita sendiri.”

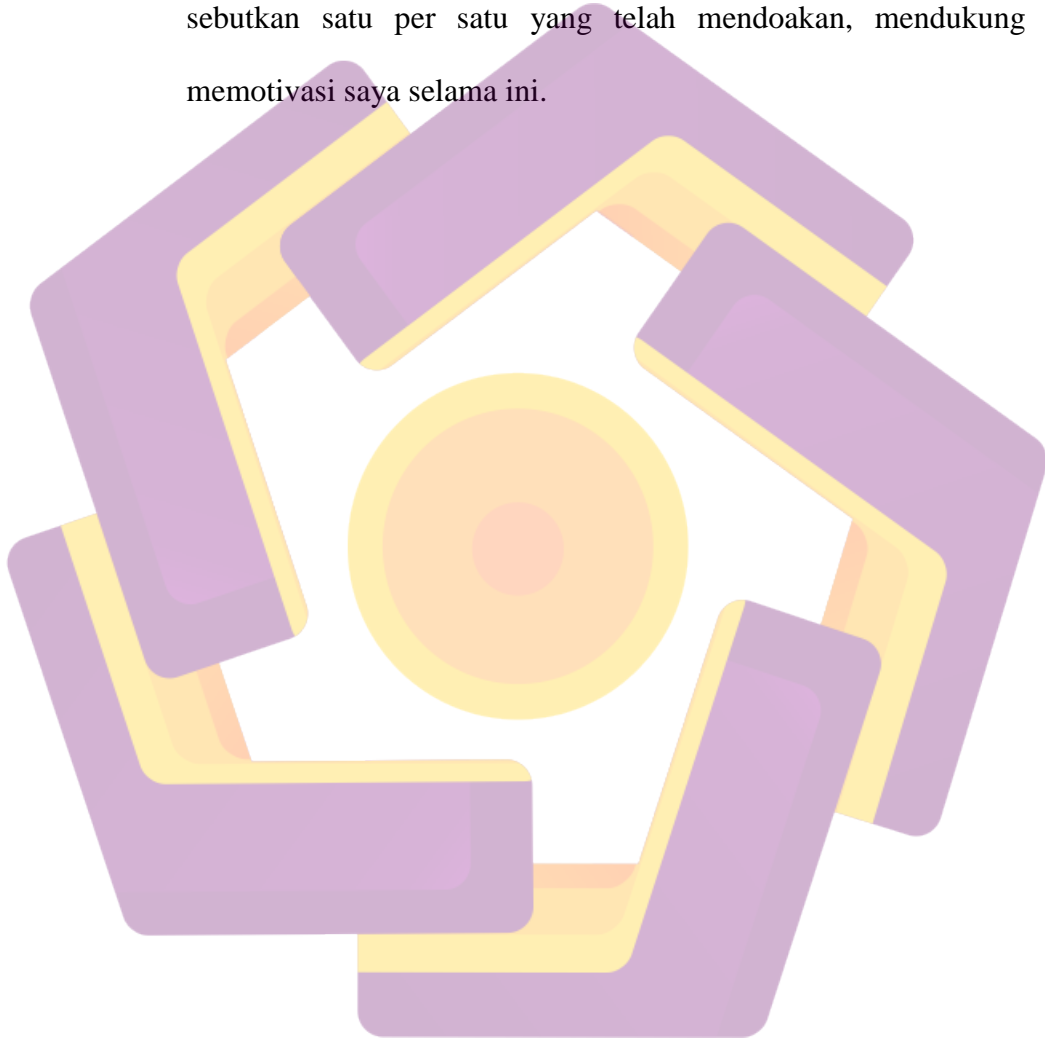
(Saitama)

## PERSEMBAHAN

Puji dan syukur saya ucapkan puji syukur kepada Tuhan Yang Maha Esa atas segala anugerah dan nikmat yang tak terkira sehingga saya dapat menyelesaikan skripsi ini. Pada kesempatan ini saya ingin menyampaikan ucapan terima kasih kepada:

1. Kedua orang tua saya yang senantiasa memberikan doa dan semangat. Terimakasih untuk segala curahan kasih sayang yang tulus dan ikhlas segala pengorbanan dan doa yang tak pernah berhenti mengalir.
2. Ibu Nila Feby Puspitasari, S. Kom, M.Cs selaku dosen pembimbing yang telah memberikan bimbingan, masukan, arahan dan motivasi selama proses penyusunan skripsi ini.
3. Bapak dan Ibu dosen Universitas Amikom Yogyakarta yang telah memberikan banyak ilmu dan pengetahuan kepada saya selama perkuliahan.
4. Moses Asi Oktovianus dan Akhid Bunayari sudah menjadi tim support dalam pengerjaan skripsi ini.
5. Sahabat-sahabat kontrakan, Rio, Kamang, Yongki, Rasid, Arya, dan Andyka yang sudah bersama dari awal perkuliahan hingga akhir perkuliahan, terimakasih telah memberikan banyak cerita dan pengalaman kepada saya.
6. Teman-Teman Student Staff UPT Amikom, terimakasih sudah menemani selama 9 (sembilan) bulan lamanya.

7. Teman-teman kelas IF06 2015 yang sudah menemani masa perkuliahan. Semoga kita selalu berbahagia, selalu bersahabat dan berproses menjadi lebih baik lagi.
8. Serta semua teman-teman dan pihak-pihak yang tidak dapat saya sebutkan satu per satu yang telah mendoakan, mendukung dan memotivasi saya selama ini.





## KATA PENGANTAR

Dengan mengucapkan syukur kepada Tuhan Yang Maha Esa atas berkat, rahmat serta karunia-Nya, Penulis dapat menyelesaikan skripsi berjudul “Implementasi Algoritma Jaro Winkler Distance Untuk Sistem Pendeteksi Similarity Pada Dokumen Naskah Publikasi”

Skripsi ini disusun dalam rangka memenuhi salah satu syarat kelulusan perguruan tinggi Program Studi Strata 1 Informatika di Universitas AMIKOM Yogyakarta dan meraih gelar Sarjana Komputer (S.Kom). Pembuatan skripsi ini tidak lepas dari berbagai pihak yang telah membantu baik dari segi material dan spiritual. Oleh karena itu Penulis ingin mengucapkan banyak rasa terimakasih kepada :

1. Bapak Prof. Dr. Suyanto, M.M., selaku rektor Universitas AMIKOM Yogyakarta.
2. Ibu Nila Feby Puspitasari, S.Kom, M.Cs., selaku dosen pembimbing, berkat bimbingan, bantuan dan saran beliau, skripsi ini dapat terselesaikan dengan hasil terbaik.
3. Ibu Krisnawati, S.Si., M.T., selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
4. Bapak Sudarmawan, M.T., selaku Ketua Program Studi S1 Informatika Universitas Amikom Yogyakarta.
5. Dosen Universitas Amikom Yogyakarta yang telah memberikan ilmu dan pengalaman, terimakasih atas semua jasa Bapak dan Ibu sekalian.

6. Kedua Orang tua saya yang tidak pernah lelah dalam memberikan dukungan, motivasi dan doanya.
7. Teman-teman dan sahabat yang telah memberikan semangat, bantuan, motivasi dan saran dalam pengerjaan skripsi ini.
8. Semua pihak yang telah membantu sampai selesainya penyusunan skripsi ini yang tentunya sangat berharga dan tidak bisa disebutkan satu per satu.

Penulis menyadari sepenuhnya, bahwa dalam penulisan skripsi ini masih jauh dari sempurna, baik dalam metode penulisan maupun cara penyajian materi. Sehingga dengan rendah hati penulis mengharapkan saran dan kritik yang bersifat membangun sehingga dikemudian hari dapat memperbaiki segala kekurangannya

Yogyakarta, 31 Oktober 2019

Ihsan Almunawaroh

15.11.8846

## DAFTAR ISI

JUDUL .....	ii
PERSETUJUAN .....	iii
PENGESAHAN .....	iv
PERNYATAAN.....	v
MOTTO .....	vi
PERSEMBAHAN .....	vii
KATA PENGANTAR .....	ix
DAFTAR ISI .....	xi
DAFTAR TABEL.....	xiv
DAFTAR GAMBAR .....	xv
INTISARI.....	xvii
ABSTRACT .....	xviii
BAB I.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Batasan masalah.....	3
1.4 Maksud dan Tujuan Penelitian.....	4
1.4.1 Maksud Penelitian.....	4
1.4.2 Tujuan Penelitian .....	4
1.5 Manfaat Penelitian .....	4
1.6 Metode Penelitian .....	5
1.6.1 Metodologi Pengumpulan Data .....	5
1.6.2 Analisis Sistem.....	6
1.6.3 Metode Pengembangan Aplikasi.....	6
1.7 Sistematika Penulisan .....	7
BAB II.....	9
2.1 Tinjauan Pustaka.....	9
2.2 Dasar Teori.....	12
2.2.1 Kecerdasan Buatan.....	12
2.2.2 Similarity.....	12

2.2.3	Plagiarisme.....	12
2.2.4	Text Mining .....	15
2.2.5	Algoritma Jaro Winkler Distance.....	20
2.2.6	K-Grams.....	22
2.2.7	Algoritma Stemming Nazief-Adriani.....	22
2.2.8	Metode Waterfall .....	25
2.2.9	Perancangan UML .....	28
2.2.10	PHP .....	33
2.2.11	Javascript.....	34
2.2.12	MySQL .....	34
2.2.13	XAMPP.....	35
<b>BAB III</b>	.....	<b>37</b>
3.1	Identifikasi Masalah.....	37
3.2	Analisis Kebutuhan.....	37
3.2.1	Analisis kebutuhan Fungsional .....	37
3.2.1.1	Definisi Fitur.....	37
3.2.1.2	Aktor .....	38
3.2.2	Analisis kebutuhan Non-Fungsional .....	39
3.3	Analisis Sistem.....	40
3.3.1	Analisis Data.....	40
3.3.2	Analisis Pemodelan Sistem.....	41
3.4	Rancangan Sistem.....	43
3.4.1	Rancangan Perangkat Lunak.....	43
3.4.2	Perhitungan Manual Pengujian Similarity Dua Naskah Publikasi.....	43
3.4.3	Perancangan Aplikasi.....	50
<b>BAB IV</b>	.....	<b>61</b>
4.1	Implementasi Interface.....	61
4.2	Pembahasan Source Code .....	61
4.2.1	Case Folding .....	62
4.2.2	Number Removal .....	62
4.2.3	Filtering.....	63
4.2.4	Stemming .....	64
4.2.5	Space Removal.....	65

4.2.6	Panjang String ( $S1$ ) & ( $S2$ ).....	65
4.2.7	Jumlah Karakter Sama ( $m$ ).....	66
4.2.8	Transposisi ( $t$ ) .....	68
4.2.9	Jaro Distance ( $dj$ ) .....	69
4.2.10	Prefix Length ( $l$ ).....	70
4.2.11	Prefix Scale ( $p$ ).....	71
4.2.12	Jaro Winkler Distance ( $jw$ ) .....	71
4.3	Pengujian Sistem.....	72
4.3.1	White Box Testing .....	72
4.3.2	Black Box Testing.....	73
4.4	Hasil dan Analisis Pengujian .....	73
4.4.1	Pengujian Dalam Sistem .....	73
4.4.2	Hasil Pengujian Naskah Publikasi .....	77
<b>BAB V</b>	.....	<b>84</b>
5.1	Kesimpulan .....	84
5.2	Saran .....	85
<b>DAFTAR PUSTAKA</b>	.....	<b>86</b>
<b>LAMPIRAN</b>	.....	<b>89</b>

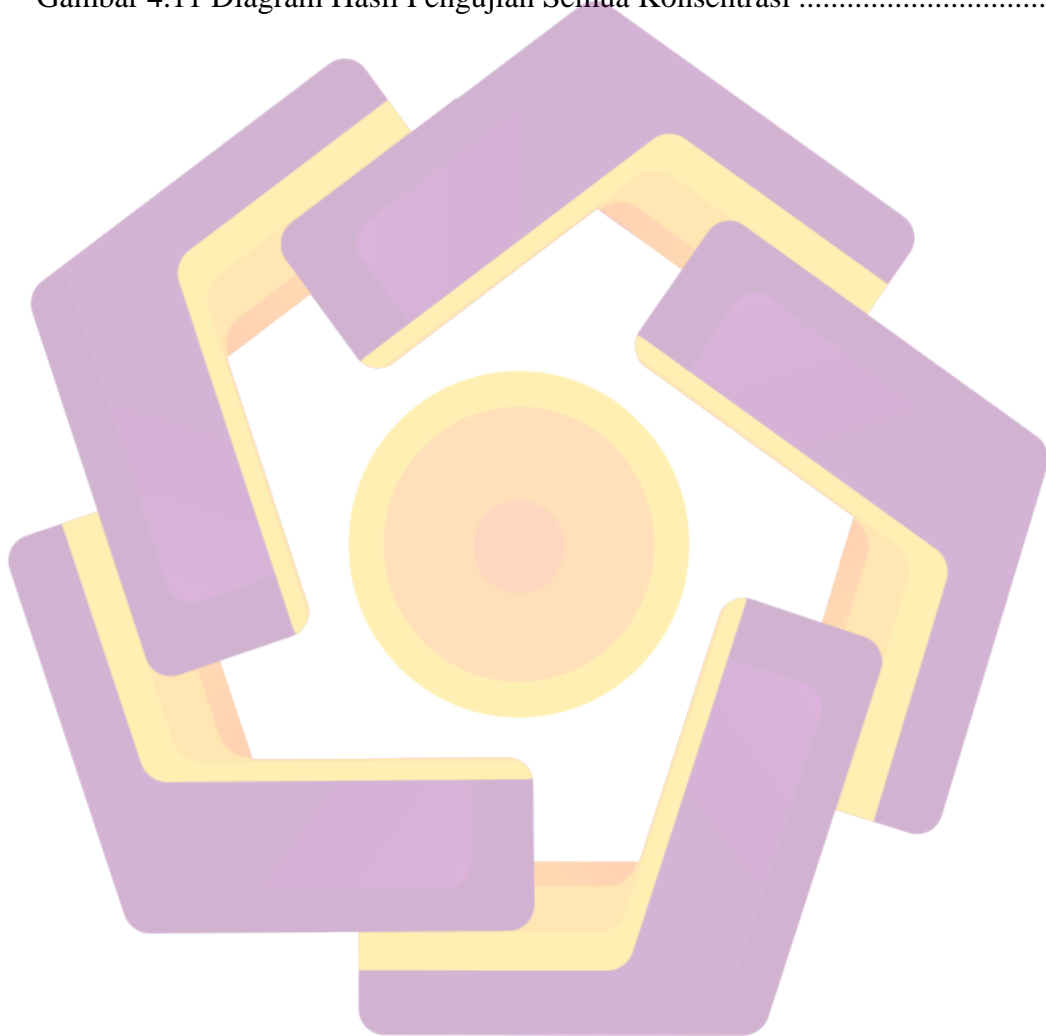
## DAFTAR TABEL

Tabel 2.1 <i>Use Case Diagram</i> .....	29
Tabel 2.2 <i>Sequence Diagram</i> .....	31
Tabel 2.3 <i>Activity Diagram</i> .....	32
Tabel 3.1 Definisi Fitur .....	38
Tabel 3.2 Aktor .....	38
Tabel 3.3 Kebutuhan Perangkat Lunak .....	39
Tabel 3.4 Kebutuhan Perangkat Keras .....	40
Tabel 3.5 Input Abstraksi1 dan Abstraksi2 .....	44
Tabel 3.6 <i>Case Folding</i> .....	45
Tabel 3.7 <i>Filtering</i> .....	46
Tabel 3.8 <i>Stemming</i> .....	46
Tabel 3.9 Hasil <i>Text Preprocessing</i> .....	47
Tabel 3.10 Jumlah String .....	47
Tabel 4.1 Potongan Kode <i>Case Folding</i> .....	62
Tabel 4.2 Potongan Kode <i>Number Removal</i> .....	63
Tabel 4.3 Potongan Kode <i>Filtering</i> .....	63
Tabel 4.4 Potongan Kode <i>Stemming</i> .....	64
Tabel 4.5 Potongan Kode <i>Space Removal</i> .....	65
Tabel 4.6 Potongan Kode <i>String 1&amp;2</i> .....	65
Tabel 4.7 Potongan Kode Perhitungan Karakter Sama.....	66
Tabel 4.8 Potongan Kode Jumlah Karakter Sama .....	67
Tabel 4.9 Potongan Kode Perhitungan Transposisi .....	67
Tabel 4.10 Potongan Kode Perhitungan <i>Jaro Distance</i> .....	69
Tabel 4.11 Potongan Kode Perhitungan <i>Prefix Length</i> .....	70
Tabel 4.12 Potongan Kode Perhitungan <i>Jaro Winkler Distance</i> .....	71
Tabel 4.13 Hasil Pengujian Naskah Publikasi Jaringan.....	77
Tabel 4.14 Hasil Pengujian Naskah Publikasi Multimedia.....	79
Tabel 4.15 Hasil Pengujian Naskah Publikasi Pemrograman .....	81

## DAFTAR GAMBAR

Gambar 2.1 Metode Pendeteksi Plagiarisme .....	14
Gambar 2.2 Tahap <i>Preprocessing</i> .....	16
Gambar 2.3 Proses <i>Case Folding</i> .....	17
Gambar 2.4 Tahap <i>Number Removal</i> .....	17
Gambar 2.5 Tahap <i>Filtering</i> .....	18
Gambar 2.6 Tahap <i>Stemming</i> .....	18
Gambar 2.7 Tahap <i>Space Removal</i> .....	19
Gambar 2.8 Proses <i>Stemming</i> .....	22
Gambar 2.9 Flowchart Algoritma <i>Stemming Nazief-Adriani</i> .....	24
Gambar 2.10 Gambar Metode <i>Waterfall</i> .....	26
Gambar 3.1 Flowchart Alur Uji Abstraksi Naskah Publikasi .....	41
Gambar 3.2 <i>Use Case</i> Diagram Input Abstraksi Naskah Publikasi .....	50
Gambar 3.3 <i>Activity</i> Diagram Membandingkan Abstraksi .....	51
Gambar 3.4 <i>Activity</i> Diagram <i>Text Preprocessing</i> .....	52
Gambar 3.5 <i>Activity</i> Diagram <i>Jaro Distance</i> .....	53
Gambar 3.6 Gambar <i>Activity</i> Diagram <i>Jaro Winkler Distance</i> .....	54
Gambar 3.7 <i>Activity</i> Diagram Perhitungan Nilai <i>Similarity</i> .....	55
Gambar 3.8 <i>Sequence</i> Diagram Membandingkan Abstraksi .....	56
Gambar 3.9 <i>Gambaran Halaman Depan Uji Plagiarisme</i> .....	57
Gambar 3.10 Input Teks dan Proses <i>Case Folding</i> .....	58
Gambar 3.11 Proses <i>Number Removal</i> dan <i>Filtering</i> .....	58
Gambar 3.12 Proses <i>Stemming</i> dan Proses <i>Space Removal</i> .....	59
Gambar 3.13 Proses Perhitungan <i>Jaro Distance</i> .....	59
Gambar 3.14 <i>Proses Perhitungan Jaro Winkler Distance dan Hasil Uji</i> .....	60
Gambar 4.1 Form Input Abstraksi Naskah Publikasi .....	74
Gambar 4.2 Teks Asli yang Akan Diproses .....	74
Gambar 4.3 Hasil <i>Case Folding</i> .....	74
Gambar 4.4 Hasil <i>Number Removal</i> .....	75
Gambar 4.5 Hasil <i>Filtering</i> .....	75

Gambar 4.6 Hasil <i>Stemming</i> .....	75
Gambar 4.7 Hasil <i>Space Removal</i> .....	76
Gambar 4.8 Hasil Perhitungan <i>Jaro Distance</i> .....	76
Gambar 4.9 Hasil Perhitungan <i>Jaro Winkler Distance</i> .....	76
Gambar 4.10 Hasil Persentase Kemiripan .....	77
Gambar 4.11 Diagram Hasil Pengujian Semua Konsentrasi .....	83





## INTISARI

Similarity atau biasa kita kenal dengan kemiripan ialah, mencari kesamaan antara kata yang di inputkan dengan kata pada dokumen teks sumber atau dokumen teks yang diuji. Kemiripan yang dimaksud bukan sinonim, melainkan kata yang mendekati atau yang hampir sama dengan kata pada dokumen yang diuji. Penggunaan alat deteksi similarity atau dapat disebut dengan alat deteksi plagiarisme secara umum dapat menjadi langkah yang cukup efektif dalam menangani plagiarisme di era globalisasi ini.

Sebenarnya sudah jelas ditulis didalam Undang-Undang No. 20 tahun 2003 yang mengatur tentang sanksi bagi orang-orang yang melakukan plagiat, khususnya yang terjadi dalam lingkungan akademik. Kasus pagiarisme sangat bisa ditanggulangi dengan cara pencegahan dan penanganan. Dapat digunakan alat deteksi plagiarisme dengan membandingkan tingkat similarity dari dua buah dokumen uji untuk mengetahui apakah dokumen tersebut terdeteksi sebagai dokumen plagiat atau tidak.

Aplikasi ini menggunakan algoritma Jaro Winkler Distance karena algoritma ini dapat memeriksa kemiripan hanya dengan melewati sedikit proses sehingga meminimalkan waktu dan proses yang relatif lebih cepat. Dengan menggunakan pendekatan string metrik, dimana dengan melakukan proses perbandingan string dan memasukkan dalam fungsi matematis, algoritma Jaro-winkler sangat akurat dan cocok untuk digunakan dalam perbandingan string pendek, seperti abstraksi pada naskah publikasi.

**Kata Kunci:** Algoritma Jaro Winkler Distance, Similarity, Artificial Intelligence

## ABSTRACT

*Similarity is looking for same word that are input with words in the source text document or text document being tested. The similarity in question is not a synonym, but a word that is close to or almost the same as the word in the document being tested. The use of similarity detection tools or can be called plagiarism detection tools in general can be a fairly effective step in dealing with plagiarism in this globalization era.*

*Actually it was clearly written in Law No. 20 of 2003 which regulates sanctions for people who commit plagiarism, especially those that occur in an academic environment. Plagiarism cases can be overcome by prevention and treatment. Plagiarism detection tool can be used by comparing the level of similarity of two test documents to determine whether the document is detected as a plagiarism document or not.*

*This application uses the Jaro Winkler Distance algorithm because this algorithm can check the similarity by only passing through a few processes so as to minimize time and relatively faster processes. By using the metric string approach, which is by carrying out the string comparison process and incorporating it in mathematical functions, the Jaro-winkler algorithm is very accurate and suitable for use in short string comparisons, such as abstractions in publication texts.*

*Keywords: Jaro Winkler Distance Algorithm, Similarity, Artificial Intelligence*