

LAPORAN SKRIPSI

**ANALISIS ALGORITMA WINNOWER DAN RABIN-KARB
UNTUK SIMILARITY DETECTION PADA DOCUMENT
BAHASA INDONESIA**



Disusun Oleh :

EKA JAYA NAGARA

2017230072

PROGRAM STUDI TEKNOLOGI INFORMASI

FAKULTAS TEKNIK

UNIVERSITAS DARMA PERSADA

JAKARTA

2022



UNIVERSITAS DARMA PERSADA

Jl. Taman Malaka Selatan, Pondok Kelapa, Jakarta Timur, Indonesia 13450


Telp. (021) 8649051, 8649053, 8649057 Fax. (021) 8649052

E-mail : humas@unsada.ac.id Home page : <http://www.unsada.ac.id>

LEMBAR PERBAIKAN

SIDANG SKRIPSI

Nama : Eka Jaya Nagara
Nim : 2017230072
Fakultas / Jurusan : Teknik / Teknologi Informasi
Tanggal : Jumat, 25 Februari 2022

No	Keterangan	Dosen
1	<ol style="list-style-type: none">1. point point, pola numbering, bukan bullet, kesimpulan diperbaiki.2. hasil uji aplikasi di ditampilkan lebih jelas pada laporan, ukuran file, jumlah document.3. latar belakang di perbaiki	<p>pak Aji Setiawan, MMSI</p> <p>09/03/2022</p> 

Mengetahui,

Kajur Teknologi Informasi

MONOZUKURI • TRILINGUAL • ENERGI TERBARUKAN



Lembaga Layanan
Pendidikan Tinggi



TERAKREDITASI B
SK BAN-PT NO. 603/SK/BAN-PT/AKRRED/VI/2015



ASOSIASI PERGURUAN TINGGI SWASTA
INDONESIA
APTISI



LEMBAR BIMBINGAN SKRIPSI TEKNOLOGI INFORMASI DARMA PERSADA

NIM : 2017230072
NAMA : Eka Jaya Nagara
DOSEN PEMBIMBING : Adam Arif Budiman M.kom
JUDUL : Analisis Algoritma Winoing dan Rabin Karb untuk Similarity Detection pada Document Bahasa Indonesia

no	Tanggal	Mat eri	Paraf Dosen Pembimbing
1	Sabtu, 20 November 2021	Laporan Bab 1	
2	Rabu, 24 November 2021	Revisi Bab 1	
3	Senin, 22 November 2021	Laporan Bab 2	
4	Selasa, 22 November 2021	Revisi Bab 2	
5	Jumat, 31 Desember 2021	Laporan Bab 3	
6	Sabtu, 8 Januari 2022	Laporan Bab 4	
7	Sabtu, 8 Januari 2022	Laporan Bab 5	
8	Sabtu, 8 Januari 2022	Demo Aplikasi	
9.	Senin, 10 Januari 2022	Revisi Bab 2	
10.	Rabu, 12 Januari 2022	Revisi Aplikasi	

Jakarta, 12 Januari 2022
Dosen Pembimbing,

Adam Arif Budiman, M.kom

LEMBAR PERNYATAAN

Saya yang bertanda tangan dibawah ini :

Nama : Eka Jaya Nagara

NIM : 2017230072

Fakultas : Teknik

Jurusan : Teknologi Informasi

Menyatakan bahwa Laporan Skripsi ini saya susun sendiri berdasarkan hasil peninjauan, penelitian lapangan, wawancara serta memadukannya dengan buku-buku, literature atau bahan-bahan referensi lain yang terkait dan relevan di dalam penyelesaian Laporan Skripsi ini.

Demikian pernyataan ini penulis buat dengan sesungguhnya.

Jakarta, 9 Januari 2022



Eka Jaya Nagara

LEMBAR PENGESAHAN

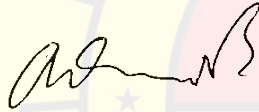
: Analisis Algoritma Winnowing dan Rabin Karb untuk Similarity Detection pada Document

Bahasa Indonesia

Disusun oleh :

Nama : Eka Jaya Nagara

NIM : 2017230072



Adam Arif Budiman, M. Kom

Pembimbing Lapangan



Adam Arif Budiman, M. Kom.

Pembimbing Laporan



Adam Arif Budiman, M. Kom

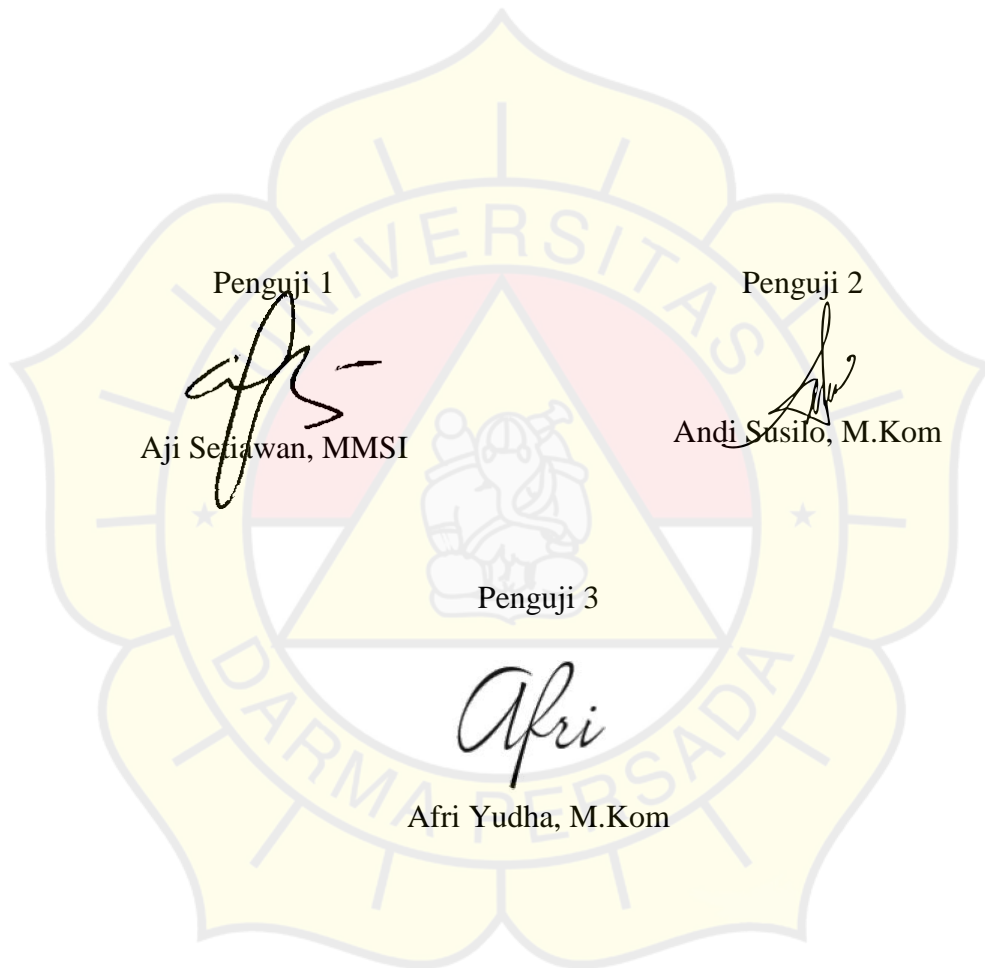
Ketua Jurusan Teknologi Informasi

LEMBAR PENGUJI SKRIPSI

Laporan Skripsi yang berjudul : “Analisis Algoritma Rabin Karb dan
Winnowing untuk Similarity Detection pada Document Bahasa Indonesia”

ini telah diajukan pada tanggal

27 September 2021



KATA PENGANTAR

Alhamdulillah rabbil alamin. Segala puji syukur kehadirat Allah SWT, karena dengan rahmad dan hidayahnya sehingga penulis sampai di tahap akhir perjalanan pendidikan dalam meraih gelar sarjana, dengan membawakan topik “Analisis algoritma winnowing dan rabin karb untuk similarity detection pada document Bahasa Indonesia”.

Penulis menyadari bahwa banyak yang membahas topik yang mirip namun banyak topik yang hanya terbatas pada syarat untuk penyelesaian studinya saja. Penulis berusaha mewujudkan agar tidak hanya sekedar menjadi syarat namun bisa digunakan sebagai layanan *freemium* terutama di tempat penulis memulai petualangan di dunia pendidikan.

Tanpa Doa orang tua, segenap motivasi, kerja keras, krisis ekonomi, cobaan berat yang dilalui, dan semangat untuk lulus, penulis berhasil menjalani petualangan di dunia akademis fakultas teknik darma persada kurang lebih 5 (lima) tahun lamanya.

Dalam perjalanan penulisan skripsi ini, penulis juga ingin mengucapkan terima kasih kepada pihak-pihak berikut ini :

1. Keluarga, Ayah, Ibu, Adik, Nene yang selalu mendukung segala kegiatan akademis saya.
2. Bapak Adam Arif Budiman M.Kom selaku kakur Prodi Teknologi Informasi dan sebagai dosen pembimbing saya. Yang sering memberikan informasi tentang beasiswa, workshop IT kepada mahasiswanya, dan saya merasa terbantu sekali. terimakasih
3. Bapak Herianto, M.T selaku dosen pemrograman web aplikasi pertama saya
4. Ibu Timor Setyaningsih selaku dosen pemrograman berbasis object yang mengenalkan saya magic OOP.
5. Bapak Sandhika Galih, sebagai dosen Online “UNPAS” . terimakasih atas ilmu yang diberikan dalam dasar PHP hingga saya bisa survive sendiri.

6. Dicoding Academy, lembaga pendidikan online terimakasih banyak telah memberikan beasiswa kepada saya.
7. PT SPS Divisi OrangT, terimakasih kepada Iko, Oki, david dan pak jimmy yang memberikan pengalaman 1 tahun 3 bulan dalam dunia industry IT.
8. Checkmarx company IT Sec, terimakasih telah memberikan kelas gratis basic security application melalui PT SPS, kepada saya.
9. Siti Nurmala “mhs IPB”, Erik Kuswanto, Sapitri Anggraini, Nita Agustine,”mhs unsada”.
10. Ni made anggraini “orang berharga sekaligus tetangga saya” Terimakasih untuk supportnya.
11. Taylor otwell, terimakasih telah menginspirasi saya dalam membangun kerangka kerja sendiri.
12. Youtube, terimakasih youtube yang telah menjadi platform belajar gratis saya selama ini
13. Semua pihak yang telah banyak membantu dalam penyusunan skripsi ini yang tidak bisa penulis sebutkan semuanya.

Akhir kata. Walaupun sudah berusaha semaksimal mungkin, tentunya masih banyak kekurangan dan keterbatasan yang dimiliki, oleh karena itu maaf dan terimakasih semoga tulisan ini bermanfaat.

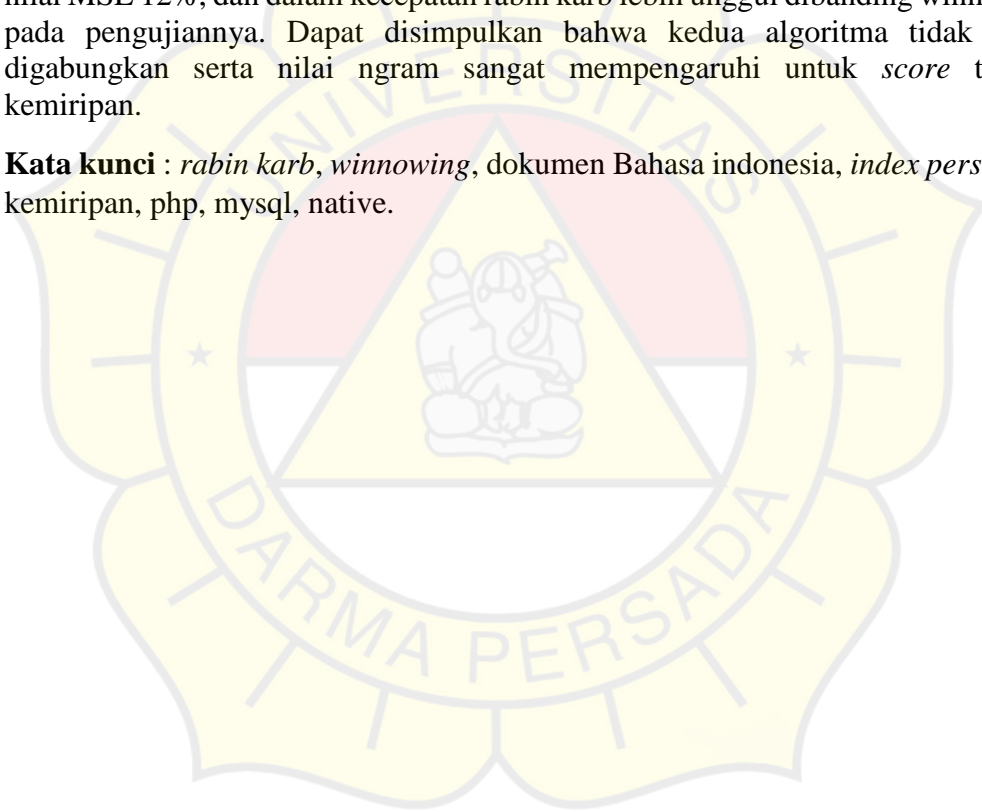
Bekasi, November 2021

Penulis

ABSTRAK

Tiap tahunnya karya ilmiah terus bertambah di Universitas Darma persada untuk mengoreksi satu persatu akan memakan waktu. Oleh karena itu perlu adanya alat bantu *scoring* yang memuat nilai *index persentase* kemiripan pada *document*. Dalam penyelesaian masalah ini penulis memilih menggunakan algoritma *rabin karb* dan *winnowing* yang dapat melakukan proses pencarian kesamaan kata melalui pola. Berdasarkan hal tersebut penelitian ini dilakukan perbandingan antara *rabin karb* dan *winnowing* dalam proses pencarian kesamaan kata atau *string*. Berdasarkan pengujian algoritma *winnowing* merupakan ekstensi dari *rabin karb* dengan penambahan *window* dan diketahui bahwa akurasi algoritma *winnowing* lebih baik yaitu untuk laporan full sebesar 100% dari 3 document yang dibanding dan pada *rabin karb* sebesar 100% dari 3 document yang dibanding. Untuk laporan bab 2 sebesar 18% dari 3 document yang dibanding kan untuk *winnowing* dan untuk *rabin karb* 24% dari 3 document yang dibandingkan dengan format file pdf dengan nilai MSE 12%, dan dalam kecepatan *rabin karb* lebih unggul dibanding *winnowing* pada pengujiannya. Dapat disimpulkan bahwa kedua algoritma tidak dapat digabungkan serta nilai ngram sangat mempengaruhi untuk *score* tingkat kemiripan.

Kata kunci : *rabin karb*, *winnowing*, dokumen Bahasa Indonesia, *index persentase* kemiripan, php, mysql, native.



Daftar Isi

LEMBAR BIMBINGAN SKRIPSI	2
LEMBAR PERNYATAAN	4
LEMBAR PENGESAHAN	5
LEMBAR PENGUJI SKRIPSI	6
KATA PENGANTAR	vii
ABSTRAK	ix
Daftar Isi	x
Daftar Gambar	xiii
Daftar Table	xv
Daftar Kode Program	xvi
BAB 1 Pendahuluan	1
1.1 Latar belakang	1
1.2 Rumusan masalah	2
1.3 Tujuan dan manfaat	2
1.3.1 Tujuan	2
1.3.2 Manfaat	2
1.4 Batasan masalah	3
1.5 Metodologi penulisan	4
1.6 Metodologi pengembangan.....	4
1.7 Sistematika penulisan	4
BAB 2 Landasan Teori	7
2.1 Tinjauan terhadap penelitian terkait	7

2.2 Similarity Index.....	9
2.3 Plagiarisme.....	9
2.4 Jenis Plagiarisme.....	10
2.5 Algoritma.....	11
2.5.1 Rabin Karb.....	11
2.5.2 Winnowing.....	11
2.6 Web Scrapping Data.....	13
2.7 Arsitekture Program.....	13
2.7.1 PHP.....	13
2.7.2 Javascript.....	14
2.7.3 MVC.....	14
2.7.4 ENV.....	16
2.7.5 Composer.....	17
2.8 UML Diagram.....	18
2.9 Agile.....	19
BAB 3 Analisis dan perancangan.....	23
3.1 Analisa.....	23
3.1.1 Analisa Masalah.....	23
3.1.2 Analisa kebutuhan system.....	23
3.2 Explorasi Agile Xstream Programing.....	24
3.3 Analisa Algoritma.....	29
3.3.1 Winnowing Algoritma.....	29
3.3.2 Rabin Karb Algoritma.....	31
3.4 Populasi dan Sample.....	32
3.5 Perancangan system.....	32
3.5.1 Use case diagram orion scanner.....	32
3.5.2 Activity diagram title scanner.....	34
3.5.3 Activity diagram text scanner.....	35
3.5.4 Activity diagram document scanner.....	36
3.5.5 Sequence diagram title scanner.....	37
3.5.6 Sequence diagram text scanner.....	37
3.5.7 Sequence diagram document scanner.....	38
3.6 Perancangan Antar muka.....	39

3.6.1 Halaman scanner title	39
3.6.2 Halaman scanner text	39
3.6.3 Halaman scanner document	40
3.6.4 Halaman report scanner document	40
3.7 Perancangan Database	41
BAB 4 Hasil dan Pembahasan	48
4.1 Spesifikasi Sistem	48
4.1.1 Hardware dan Software untuk Development	48
4.1.2 Tester Deployment serve	49
4.2 Hasil User Interface	49
4.2.1 Halaman Login	49
4.2.2 Halaman Profile	50
4.2.3 Halaman Title Scanner	50
4.2.4 Halaman Text Scanner	51
4.2.5 Halaman Document Scanner	51
4.2.6 Halaman Report Document Scanner	52
4.3 Pengujian metode	53
4.3.1 Winnowing	53
4.3.2 Rabin Karb	59
4.4 Kesimpulan pengujian	64
4.5 Hasil pengujian	64
BAB 5 Penutup	67
5.1 Kesimpulan	67
5.2 Saran	68
Daftar Pustaka	69
Lampiran	72

Daftar Gambar

Gambar 2.1. Model View Controller Design Pattern	14
Gambar 2.2. Composer Dependency manager untuk PHP.....	17
Gambar 3.1 Algoritma winnowing	30
Gambar 3.2 Algoritma rabin karb.....	31
Gambar 3.3 use case diagram orion scanner 1	33
Gambar 3.4 use case diagram orion scanner 2	33
Gambar 3.5 diagram activity scanner title.....	34
Gambar 3.6 diagram activity scanner text.....	35
Gambar 3.7 diagram activity document scanner	36
Gambar 3.8 sequence diagram titile scanner	37
Gambar 3.9 sequence diagram text scanner	37
Gambar 3.10 sequence diagram document scanner	38
Gambar 3.11 perancangan halaman title scanner	39
Gambar 3.12 perancangan halaman text scanner	39
Gambar 3.13 Perancangan halaman document scanner	40
Gambar 3.14 Perancangan halaman report document scanner.....	40
Gambar 4.1 User Interface Halaman Login.....	50
Gambar 4.2 User Interface Halaman Profile	50
Gambar 4.3 User Interface Halaman Title Scanner.....	51
Gambar 4.4 User Interface Halaman Text Scanner	51
Gambar 4.5 User Interface Halaman Document Scanner – 1.....	52
Gambar 4.6 User Interface Halaman Document Scanner - 2	52
Gambar 4.7 User Interface Halaman Report Document Scanner.....	52

Gambar 4.8 gambar hasil pengujian document laporan skripsi full 57

Gambar 4.9 hasil report pengujian document bab 2 58

Gambar 4.10 formula MSE 65



Daftar Table

Table 2.1 Penelitian Terdahulu.....	8
Table 2.2 tabel uml diagram.....	19
Table 2.3 contoh table user stories pada XP.....	21
Table 3.1 tabel user stories	24
Table 3.2 tabel hak akses stories	25
Table 3.3 Tabel fitur stories.....	25
Table 3.4 Desain Tabel Users.....	41
Table 3.5 Desain Tabel Documents	42
Table 3.6 Desain Table Skripsi	43
Table 3.7 Desain Table Jurnal	44
Table 3.8 Desain Tabel Intership.....	44
Table 3.9 Desain Table temp Scanner.....	45
Table 3.10 Desain Tabel Scanner.....	45
Table 4.1 hasil pengujian metode winnowing pada judul skripsi	53
Table 4.2 tabel hasil pengujian document algoritma winnowing.....	58
Table 4.3 table hasil pengujian string.....	59
Table 4.4 table hasil pengujian string.....	63
Table 4.5 tabel hasil pengujian document algoritma winnowing.....	63
Table 4.6 hasil analisis pengujian algoritma dalam bentuk persentase	64

Daftar Kode Program

Kode Program 2.1 contoh penggunaan .env pada konfigurasi program computer	16
Kode Program 2.2 contoh pemanggilan file .env pada php	16
Kode Program 2.3 contoh penggunaan composer.....	18

