

**PENERAPAN METODE LEVENSHTTEIN *DISTANCE*
BERBASIS N-GRAM UNTUK MENINGKATKAN
AKURASI PADA PENGEMBANGAN *REST WEB*
SERVICE DETEKSI PLAGIARISME**

TESIS

**Diajukan kepada
Universitas Pendidikan Ganesha
Untuk Memenuhi sebagai Persyaratan
Memperoleh Gelar Magister Ilmu Komputer
Program Studi Ilmu Komputer**

**Oleh
I WAYAN ADI SPARTA
NIM : 1729101068**

**PROGRAM STUDI ILMU KOMPUTER
PROGRAM PASCASARJANA
UNIVERSITAS PENDIDIKAN GANESHA**

2019

Tesis oleh I Wayan Adi Sparta ini telah diperiksa dan disetujui untuk Ujian Tesis.

Singaraja, Juli 2019

Pembimbing I,



Dr. Drs. I Nyoman Sukajaya, M.I
NIP. 19671115 199303 1 001

Pembimbing II,



Dr. I Gede Aris Gunadi, S.Si. M.Kom.
NIP. 19770318 200812 1 004

Tesis oleh I Wayan Adi Sparta ini telah berhasil dipertahankan di depan tim penguji dan dinyatakan diterima sebagai sebagian persyaratan untuk memperoleh gelar Magister Komputer di Program Studi Ilmu Komputer Program Pascasarjana, Universitas Pendidikan Ganesha.

Disetujui pada tanggal:

oleh Tim Penguji

	Ketua	<u>Dr. Gede Indrawan, S.T., M.T.</u> NIP 197601022003121001
	Anggota	<u>Made Windu Antara Kesiman, S.T., M.Sc., Ph.D.</u> NIP. 198211112008121001
	Anggota	<u>Dr. Gede Rasben Dantes, S.T., M.T.I.</u> NIP 197502212003121001
	Anggota	<u>Dr. I Nyoman Sukajaya, M.T.</u> NIP 196711151993031001
	Anggota	<u>Dr. I Gede Aris Gunadi, S.Si, M.Kom.</u> NIP 197703182008121004

Mengetahui Direktur
Program Pascasarjana UNDIKSHA,

Prof. Dr. I Gusti Putu Suharta, M.Si.
NIP 196212151988031002

LEMBAR PERNYATAAN

Saya menyatakan dengan sesungguhnya, bahwa tesis yang saya susun sebagai syarat untuk memperoleh gelar Magister Ilmu Komputer dari Program Pascasarjana Universitas Pendidikan Ganesha seluruhnya merupakan hasil karya saya sendiri. Adapun bagian-bagian tertentu dalam penulisan tesis yang saya kutip dari hasil karya orang lain telah dituliskan sumbernya secara jelas dan sesuai dengan norma, kaidah, serta etika akademis.

Apabila di kemudian hari ditemukan seluruh atau sebagian tesis ini bukan hasil karya saya sendiri atau adanya plagiat dalam bagian-bagian tertentu, saya bersedia menerima sanksi pencabutan gelar akademik yang saya sandang dan sanksi-sanksi lainnya sesuai dengan peraturan perundang-undangan yang berlaku di wilayah Negara Kesatuan Republik Indonesia.

Singaraja, Juli 2019

Yang membuat pernyataan



I Wayan Adi Sparta

PRAKATA

Puji dan syukur penulis panjatkan kehadapan Ida Sang Hyang Widhi Wasa karena atas asung waranugraha-Nya, tesis yang berjudul “*Penerapan Metode Levenshtein Distance Berbasis N-gram untuk Meningkatkan Akurasi pada Pengembangan Rest Web Service Deteksi Plagiarisme*” dapat diselesaikan sesuai dengan yang direncanakan.

Tesis ini ditulis untuk memenuhi sebagian persyaratan dalam menyelesaikan studi di Program Pascasarjana Universitas Pendidikan Ganesha, Program Studi Ilmu Komputer. Pada lembar-lembar awal tesis ini, ijinakan penulis menyampaikan terimakasih dan penghargaan yang sebesar-besarnya kepada:

1. Dr. Drs. I Nyoman Sukajaya, M.T, sebagai pembimbing I yang dengan sabar membimbing, mengarahkan, dan memberikan motivasi yang demikian bermakna, sehingga penulis mampu menyelesaikan tesis ini.
2. Dr. I Gede Aris Gunadi, S.Si. M.Kom., sebagai pembimbing II yang telah membimbing, mengarahkan, dan memberikan motivasi selama peyusunan tesis, sehingga tesis ini dapat terwujud dengan baik sesuai harapan.
3. Rektor Universitas Pendidikan Ganesha, yang telah memberikan bantuan secara moril dan memfasilitasi berbagai kepentingan studi, selama penulis menempuh perkuliahan di Program Pascasarjana Undiksha.
4. Direktur Program Pascasarjana Undiksha dan staf, yang telah banyak membantu selama penulis mengikuti dan menyelesaikan penulisan tesis ini.
5. Ketua Program Studi Ilmu Komputer dan staf dosen pengajar di program studi Ilmu Komputer yang telah banyak membantu dan memotivasi penulis selama perjalanan studi dan penyusunan tesis ini.
6. Seluruh peneliti yang jurnalnya telah terpublikasi pada *e-journal* Universitas Pendidikan Ganesha, yang penulis gunakan sebagai bahan uji dalam tesis ini, serta penulis lainnya yang memberikan informasi baik mengenai materi maupun referensi yang membantu peneliti dalam mengembangkan penelitian ini.

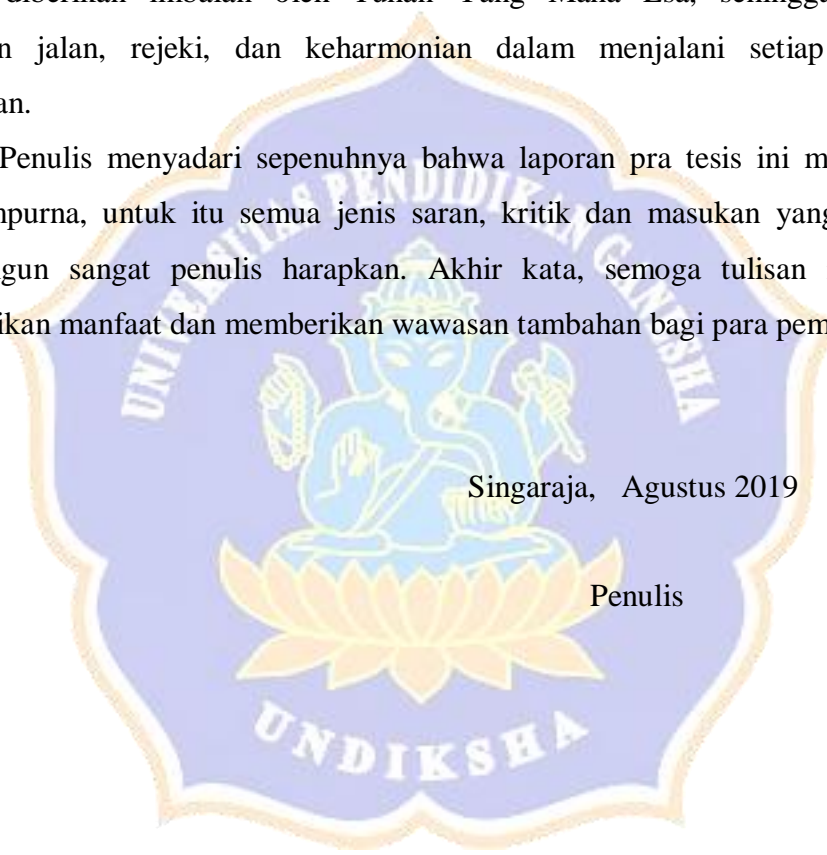
7. Rekan-rekan seangkatan di Program Studi Ilmu Komputer yang dengan karakternya masing-masing telah banyak berkontribusi membentuk kemandirian penulis selama menjalani studi dan menyelesaikan tesis ini.
8. I Nyoman Seken dan Ni Made Lendri selaku orang tua penulis, serta adik dan keluarga yang telah banyak membantu secara material dan moral selama perjalanan studi yang penulis lakoni di Program Studi Ilmu Komputer.

Semoga semua bantuan yang telah mereka berikan dalam perjalanan studi penulis diberikan imbalan oleh Tuhan Yang Maha Esa, sehingga mereka diberikan jalan, rejeki, dan keharmonian dalam menjalani setiap langkah kehidupan.

Penulis menyadari sepenuhnya bahwa laporan pra tesis ini masih jauh dari sempurna, untuk itu semua jenis saran, kritik dan masukan yang bersifat membangun sangat penulis harapkan. Akhir kata, semoga tulisan ini dapat memberikan manfaat dan memberikan wawasan tambahan bagi para pembaca.

Singaraja, Agustus 2019

Penulis



DAFTAR ISI

	HALAMAN
LEMBAR PERNYATAAN	v
PRAKATA	vi
ABSTRAK	viii
DAFTAR ISI	x
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xiv
DAFTAR LAMPIRAN	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	6
1.3 Ruang Lingkup dan Keterbatasan Penelitian	7
1.4 Tujuan Penelitian	7
BAB II KAJIAN TEORI	9
2.1 Plagiat	9
2.2 Levenshtein <i>Distance</i>	10
2.3 N-Gram	14
2.4 Jaro-Winkler <i>Distance</i>	15
2.5 Levenshtein <i>Distance</i> Berbasis N-gram	18
2.6 Penelitian Relevan	22
BAB III METODE PENELITIAN	26
3.1 Fase Pemodelan Bisnis (<i>Bussiness Modeling</i>)	27
3.2 Fase Pemodelan Data (<i>Data Modeling</i>)	28
3.3 Fase Pemodelan Proses (<i>Process Modeling</i>)	29

3.4	Fase Pengembangan Aplikasi (<i>Application Generation</i>).....	32
3.5	Fase Pengujian dan Pergantian (<i>Testing and Turnover</i>)	32
BAB IV HASIL PENELITIAN DAN PEMBAHASAN		34
4.1	<i>Harvesing</i>	34
4.2	Struktur Data Perangkat Lunak	35
4.3	Implementasi Perangkat Lunak	36
4.4	<i>Preprocessing</i>	38
4.4.1	<i>Case Folding</i>	40
4.4.2	<i>Tokenizing</i>	41
4.4.3	<i>Stop Word Removal</i>	41
4.4.4	<i>Stemming</i>	42
4.5	Hasil Eksperimen Pengujian Operator <i>Double Sama Dengan</i>	49
4.6	Levenshtein <i>Distance</i>	50
4.7	Jaro-Winkler <i>Distance</i>	52
4.8	Levenshtein <i>Distance</i> Berbasis N-gram	53
4.9	Hasil Pengujian	54
4.9.1	Hasil Pengujian Akurasi.....	55
4.9.2	Hasil Pengujian Kecepatan Metode.....	56
4.10	Pembahasan.....	57
4.10.1	Nilai Persentase Plagiarisme Levenshtein <i>Distance</i> Identik Lebih Tinggi Dibandingkan Levenshtein <i>Distance</i> Berbasis N-gram.....	57
4.10.2	Nilai Persentase yang Diberikan Jaro-Winkler <i>Distance</i> Relatif Lebih Tinggi.....	61
4.10.3	Pembahasan Kecepatan Metode Levenshtein <i>Distance</i> Berbasis N-gram	68

BAB V PENUTUP	70
5.1 Simpulan	70
5.2 Saran	72
DAFTAR RUJUKAN.....	74



DAFTAR TABEL

Tabel	Halaman
Tabel 4. 1 Daftar Artikel untuk Uji Coba	35
Tabel 4. 2 Tabel Hasil Pengujian Operator "="	49
Tabel 4. 3 Rata-rata Nilai Plagiarisme	55
Tabel 4. 4 Hasil Uji dengan MAPE Jaro-Winkler <i>Distance</i> sebagai Patokan	55
Tabel 4. 5 Perbandingan Rata-rata Kecepatan Metode.....	56
Tabel 4. 6 Hasil pengujian sebelum dikategorisasi	65
Tabel 4. 7 Hasil setelah dikategorisasi dan perhitungan akurasi.....	65
Tabel 4. 8 Rekapitulasi Uji Akurasi dengan data <i>Prepostse</i>	66
Tabel 4. 9 Hasil Uji Validitas	67



DAFTAR GAMBAR

Gambar	Halaman
Gambar 2. 1 Alur Proses Levenshtein <i>Distance</i>	18
Gambar 2. 2 Alur Proses Levenshtein <i>Distance</i> Berbasis N-gram.....	18
Gambar 3. 1 Skema Model RAD.....	27
Gambar 3. 2 Rancang ERD Sistem.....	28
Gambar 3. 3 <i>Flowchart</i> Alur Penggunaan <i>Rest Web Service</i>	30
Gambar 4. 1 Implementasi Tabel dan Relasinya	36
Gambar 4. 2 Implementasi Halaman Depan Sistem.....	37
Gambar 4. 3 Implementasi Halaman dokumentasi API.....	37
Gambar 4. 4 Implementasi Sistem Pengecekan Plagiarisme Artikel Mahasiswa .	38
Gambar 4. 5 <i>Source Code Indexing</i> tahap Normalisasi	40
Gambar 4. 6 <i>Source Code Function Case Folding</i>	41
Gambar 4. 7 <i>Source Code</i> Tahap pembersihan tanda baca	41
Gambar 4. 8 <i>Source Code Stop Word Removal</i>	42
Gambar 4. 9 <i>Source Code Stemming</i>	48
Gambar 4. 10 <i>Source Code Levenshtein Distance</i>	51
Gambar 4. 11 Implementasi Souce Code Jaro-Winkler <i>Distance</i>	53
Gambar 4. 12 <i>Source Code Levenshtein Distance</i> Berbasis N-gram.....	54
Gambar 4. 13 Simulasi Matrik Levenshtein <i>Distance</i> level karakter	58
Gambar 4. 14 Simulasi Matrik Levenshtein <i>Distance</i> Berbasis 2-gram level karakter.....	59
Gambar 4. 15 Simulasi Matrik Levenshtein <i>Distance</i> Level Kata	60
Gambar 4. 16 Simulasi Matrik Levenshtein <i>Distance</i> Berbasis 2-gram Level Kata	60
Gambar 4. 17 Simulasi Matrik Jaro-Winkler <i>Distance</i>	62
Gambar 4. 18 Simulasi Levenshtein <i>Distance</i> Level Karakter	63

DAFTAR LAMPIRAN

Lampiran	Halaman
Lampiran 1. Daftar <i>Stop word</i>	78
Lampiran 2. Daftar Kata Dasar.....	80
Lampiran 3. Daftar Artikel.....	81
Lampiran 4. Daftar Kata Kerja Bahasa Indonesia	83
Lampiran 5. Hasil Proses <i>Case Folding</i>	85
Lampiran 6. Hasil Proses Tokenizing	86
Lampiran 7. Hasil Proses <i>Stop Word Removal</i>	87
Lampiran 8. Hasil Proses Stemming.....	88
Lampiran 9. Impementasi Tabel.....	89
Lampiran 10. Impementasi Tampilan Antarmuka Perangkat Lunak.....	93
Lampiran 11. Hasil Pendeteksian dengan Metode Levenshtein <i>Distance</i>	103
Lampiran 12. Hasil Pendeteksian Dengan Metode Levenshtein <i>Distance</i> 2 Gram	114
Lampiran 13. Hasil Pendeteksian Dengan Metode Levenshtein <i>Distance</i> 3 Gram	124
Lampiran 14. Hasil Pendeteksian Dengan Metode Jaro-Winkler <i>Distance</i>	134
Lampiran 15 Hasil Pengujian Levenshtein <i>Distance</i> dengan Jaro-Winkler <i>Distance</i>	144
Lampiran 16 Hasil Pengujian Levenshtein <i>Distance</i> 1-Gram dengan Jaro-Winkler <i>Distance</i>	153
Lampiran 17 Hasil Pengujian Levenshtein <i>Distance</i> 2-Gram dengan Jaro-Winkler <i>Distance</i>	161
Lampiran 18 Hasil Pengujian Levenshtein <i>Distance</i> 3-gram dengan Jaro-Winkler <i>Distance</i>	171
Lampiran 19 Hasil Pengujian dengan Data <i>Prepostseo</i>	180
Lampiran 20 Perbandingan Waktu Ketiga Metode	186
Lampiran 21 Data Uji Validitas Metode.....	194