

**STUDI KOMPARASI ALGORITMA KONVENSIONAL
DAN *VECTOR SPACE MODEL* PADA PERINGKAT
HASIL PENELUSURAN JURNAL**

SKRIPSI

Untuk memenuhi sebagian persyaratan mencapai derajat S-1
Program Studi Teknik Informatika



Disusun oleh:

AZIS ALVRIYANTO

16650036

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
YOGYAKARTA**

2020

HALAMAN PENGESAHAN



KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
FAKULTAS SAINS DAN TEKNOLOGI
Jl. Marsda Adisucipto Telp. (0274) 540971 Fax. (0274) 519739 Yogyakarta 55281

PENGESAHAN TUGAS AKHIR

Nomor : B-1141/Un.02/DST/PP.00.9/06/2020

Tugas Akhir dengan judul : STUDI KOMPARASI ALGORITMA KONVENSIONAL DAN VECTOR SPACE
MODEL PADA PERINGKAT HASIL PENELITIAN JURNAL

yang dipersiapkan dan disusun oleh:

Nama : AZIS ALVRIYANTO
Nomor Induk Mahasiswa : 16650036
Telah diujikan pada : Selasa, 19 Mei 2020
Nilai ujian Tugas Akhir : A-

dinyatakan telah diterima oleh Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta

TIM UJIAN TUGAS AKHIR



Valid ID: Seead6d194025

Ketua Sidang/Penguji I

Muhammad Taufiq Nuruzzaman, S.T. M.Eng.
SIGNED



Valid ID: Seead72d49a8c

Penguji II

Maria Ulfah Siregar, S.Kom. MIT., Ph.D.
SIGNED



Valid ID: Seead5c6ac78d

Penguji III

Rahmat Hidayat, S.Kom., M.Cs.
SIGNED



Valid ID: Seeadecce785f5

Yogyakarta, 19 Mei 2020

UIN Sunan Kalijaga
Dekan Fakultas Sains dan Teknologi

Dr. Murtono, M.Si.
SIGNED

SURAT PERSETUJUAN SKRIPSI



Universitas Islam Negeri Sunan Kalijaga



FM-UINSK-BM-05-03/R0

SURAT PERSETUJUAN SKRIPSI/TUGAS AKHIR

Hal : Persetujuan Skripsi

Lamp :

Kepada

Yth. Dekan Fakultas Sains dan Teknologi

UIN Sunan Kalijaga Yogyakarta

di Yogyakarta

Assalamu 'alaikum wr. wb.

Setelah membaca, meneliti, memberikan petunjuk dan mengoreksi serta mengadakan perbaikan seperlunya, maka kami selaku pembimbing berpendapat bahwa skripsi Saudara:

Nama : Azis Alvriyanto

NIM : 16650036

Judul Skripsi : Studi Komparasi Algoritma Konvensional dan *Vector Space Model* pada Peringkat Hasil Penelusuran Jurnal

sudah dapat diajukan kembali kepada Program Studi Teknik Informatika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta sebagai salah satu syarat untuk memperoleh gelar Sarjana Strata Satu dalam Program Studi Teknik Informatika

Dengan ini kami mengharap agar skripsi/tugas akhir Saudara tersebut di atas dapat segera dimunaqsyahkan. Atas perhatiannya kami ucapkan terima kasih.

Wassalamu 'alaikum wr. wb.

Yogyakarta, 18 Juni 2020
Pembimbing

M. Taufiq Nuruzzaman, S.T., M.Eng.
NIP. 19791118 200501 1 003

PERNYATAAN KEASLIAN SKRIPSI

PERNYATAAN KEASLIAN SKRIPSI

Saya yang bertanda tangan dibawah ini:

Nama : Azis Alvriyanto
NIM : 16650036
Program Studi : Teknik Informatika
Fakultas : Sains dan Teknologi

Menyatakan bahwa skripsi saya yang berjudul “**Studi Komparasi Algoritma Konvensional dan *Vector Space Model* pada Peringkat Hasil Penelusuran Jurnal**” merupakan hasil penelitian saya sendiri, tidak terdapat pada karya yang pernah di ajukan untuk memperoleh gelar kesarjana di suatu perguruan tinggi, dan bukan plagiasi karya orang lain kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, 18 Juni 2020



Azis Alvriyanto
NIM. 16650036

KATA PENGANTAR

Puji syukur kehadirat Allah SWT yang telah melimpahkan rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan penelitian yang berjudul **Studi Komparasi Algoritma Konvensional dan *Vector Space Model* pada Peringkat Hasil Penelusuran Jurnal** sebagai salah satu syarat untuk mencapai gelar sarjana program studi Teknik Informatika Universitas Islam Negeri Sunan Kalijaga Yogyakarta. Sholawat serta salam selalu tucurahkan kepada junjungan kita Nabi Muhammad saw. beserta seluruh keluarga dan sahabat beliau. Penulis menyadari bahwa apa yang dilakukan dalam penyusunan laporan penelitian ini masih terlalu jauh dari kesempurnaan.

Oleh karena itu, penulis mengharapkan kritik dan saran yang berguna dalam penyempurnaan analisis ini di masa yang akan datang. Semoga apa yang telah penulis lakukan dapat bermanfaat bagi pembaca.

Tidak lupa penulis juga mengucapkan terima kasih kepada pihak-pihak yang telah membantu dalam penyelesaian tugas akhir ini, baik secara langsung maupun tidak langsung. Ucapan terima kasih penulis sampaikan kepada:

1. Bapak Dr. Phil. Sahiron, M.A., selaku Rektor UIN Sunan Kalijaga.
2. Bapak Dr. H. Waryono, M.Ag., selaku Wakil Rektor Bidang Kemahasiswaan dan Kerjasama UIN Sunan Kalijaga.
3. Bapak Dr. Murtono, M.Si., selaku Dekan Fakultas Sains dan Teknologi UIN Sunan Kalijaga.
4. Bapak Sumarsono, S.T., M.Kom., selaku Ketua Program Studi Teknik Informatika UIN Sunan Kalijaga.
5. Bapak M. Taufiq Nuruzzaman, S.T., M.Eng., selaku dosen pembimbing skripsi yang telah sabar dan meluangkan waktunya untuk memberikan motivasi, koreksi dan kritik saran kepada penulis sehingga skripsi ini dapat diselesaikan.
6. Bapak Dr. Agung Fatwanto, S.Si., M.Kom., selaku Dosen Pembimbing Akademik.

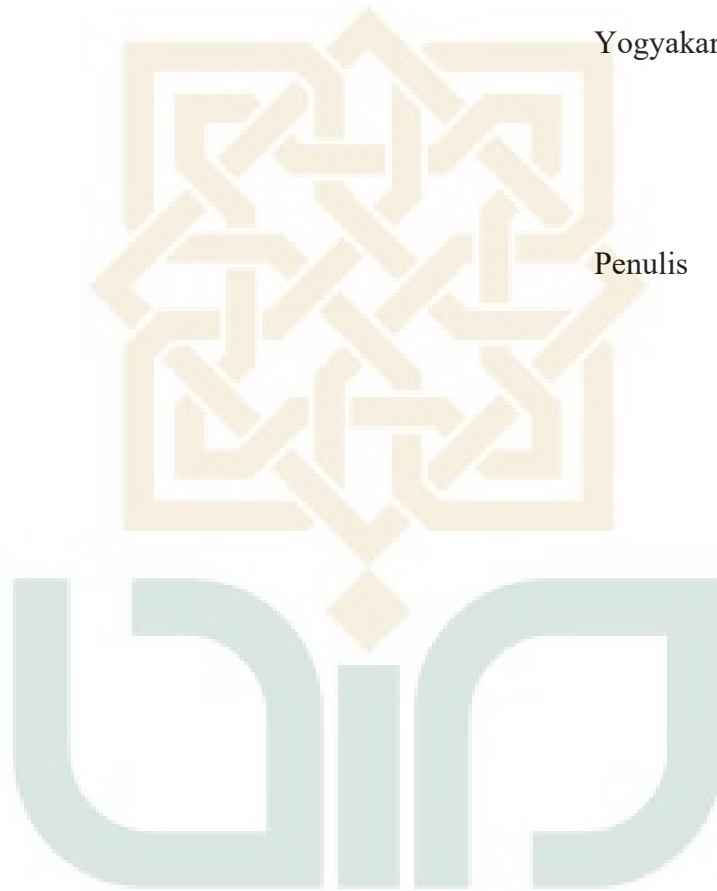
7. Agus Mulyanto, S.Si., M.Kom., Ph.D., Aulia Faqih Rifa'i, M.Kom., Muhammad Didik Rohmad Wahyudi, S.T., MT., Maria Ulfah Siregar, S.Kom. MIT., Ph.D., Nurochman, S.Kom., M.Kom., Rahmat Hidayat, S.Kom., M.Cs., Dr. Shofwatul 'Uyun, S.T., M.Kom., selaku dosen pengampu mata kuliah program studi Teknik Informatika UIN Sunan Kalijaga yang telah banyak membantu sehingga penulis dapat menyusun tugas akhir.
8. Seluruh staf dan karyawan Fakultas Sains dan Teknologi UIN Sunan Kalijaga.
9. Ucapan terima kasih yang terdalam untuk kedua orang tua saya, ibu Rini Astuti dan bapak Wakmad yang selalu memberikan doa, perhatian, kasih sayang dan segala dukungan yang tak terhingga.
10. Adik-adikku, Isnenni Oktivia dan Ajeng Novita yang tidak pernah henti memberikan doa, kasih sayang dan hiburan kepada penulis selama mengerjakan skripsi ini.
11. Sahabat-sahabatku Riswan, Wahid, Fadqur, Nabil, Amrul, Juang, Husni, Aat, Lina, Nelly, Sekar, Ulfa, Ari, Adit, Irfan yang telah senantiasa menemani penulis serta memberikan semangat, dorongan, dan inspirasi dalam proses pengerjaan skripsi ini.
12. Teman-teman yang pernah menjadi satu kelompok selama kuliah dan seluruh teman-teman Teknik Informatika 2016 yang tidak dapat penulis sebutkan satu per satu.
13. Teman-teman Himpunan Mahasiswa Kupu-Kupu dan Kontrakan SWAG yang telah memberikan warna hari-hari penulis selama kuliah.
14. Teman-teman KKN Angkatan 99 Pulau Raas yang telah banyak menghibur dan memberi dukungan kepada penulis, secara khusus Latif, Nabil, Humam, Fajril, Rahma, Clara, Anri, Kiki dan Diyah.
15. Seluruh teman-teman dan warga desa Ketupat, Kec. Raas, Kab. Sumenep, Jawa timur yang senantiasa mendoakan dan memberikan dukungan kepada penulis.

16. Serta semua pihak yang tidak dapat penulis sebutkan satu persatu dan telah memberikan banyak doa dan dukungan sehingga penelitian ini dapat terselesaikan.

Semoga dukungan dan bantuan yang telah diberikan kepada penulis, menjadi amal baik dan mendapatkan pahala dari Allah swt. Akhir kata penulis ucapkan terima kasih.

Yogyakarta, 18 Juni 2020

Penulis



HALAMAN PERSEMBAHAN



Skripsi ini saya persembahkan untuk ibu, ayah dan semesta ini.

HALAMAN MOTTO



“...Kemerdekaan itu ialah hak segala bangsa.”

—Pembukaan UUD 1945



DAFTAR ISI

HALAMAN DEPAN.....	i
HALAMAN PENGESAHAN.....	ii
SURAT PERSETUJUAN SKRIPSI.....	iii
PERNYATAAN KEASLIAN SKRIPSI.....	iv
KATA PENGANTAR.....	v
HALAMAN PERSEMBAHAN	viii
HALAMAN MOTTO	ix
DAFTAR ISI.....	x
DAFTAR GAMBAR.....	xiii
DAFTAR TABEL	xv
DAFTAR RUMUS	xviii
INTISARI	xix
ABSTRACT	xx
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	2
1.3. Tujuan Penelitian.....	2
1.4. Batasan Masalah.....	2
1.5. Manfaat Penelitian.....	3
1.6. Keaslian Penelitian	3
1.7. Sistematika Penulisan.....	3
BAB II DAFTAR PUSTAKA DAN LANDASAN TEORI.....	5
2.1 Tinjauan Pustaka	5
2.2 Landasan Teori	16
2.2.1 <i>Text Mining</i>	16
2.2.2 <i>Scraping</i>	16

2.2.3	<i>Text Preprocessing</i>	18
2.2.4	<i>Term Frequency-Inverse Document Frequency (TF-IDF)</i>	19
2.2.5	Algoritma Konvensional	21
2.2.6	<i>Vector Space Model</i>	21
2.2.7	<i>Cosine Similarity</i>	22
2.2.8	PHP	23
2.2.9	JavaScript	24
BAB III	METODE PENELITIAN	26
3.1	Metode Penelitian.....	26
3.2	Objek dan Lokasi Penelitian.....	26
3.3	Tahap Penelitian	26
3.3.1	Pengambilan Data	26
3.3.2	<i>Text Preprocessing</i>	27
3.3.3	Algoritma Konvensional	27
3.3.4	Algoritma <i>Vector Space Model</i>	27
3.3.5	Hasil dan Analisis	27
BAB IV	HASIL DAN PEMBAHASAN.....	28
4.1	Deskripsi Data	28
4.2	Uji Algoritma	103
4.3	Hasil Penelitian.....	106
4.4	Pembahasan Penelitian	113
BAB V	PENUTUP	121
5.1	Kesimpulan.....	121
5.2	Saran.....	121
DAFTAR PUSTAKA		123
LAMPIRAN		127

CURRICULUM VITAE.....	139
------------------------------	------------



DAFTAR GAMBAR

Gambar 2.1 Alur <i>Scraping</i>	17
Gambar 2.2 Representasi Dokumen dan <i>Query</i> pada Ruang Vektor	22
Gambar 2.3 Representasi Grafis Sudut Dokumen dan <i>Query</i>	23
Gambar 4.1 Daftar Tahun Jurnal yang Tersedia pada Institutional Repository UIN Sunan Kalijaga	28
Gambar 4.2 Daftar Jurnal Tahun 2013 Institutional Repository UIN Sunan Kalijaga	29
Gambar 4.3 <i>Code Style Scraping</i> Jurnal	30
Gambar 4.4 Data Jurnal pada <i>Database MySQL</i>	30
Gambar 4.5 <i>Code Style</i> Tahap <i>Text Preprocessing</i> dan Pembobotan	102
Gambar 4.6 Data Hasil <i>Text Preprocessing</i> dan Pembobotan Kata pada Firebase	103
Gambar 4.7 Program Algoritma Konvensional	104
Gambar 4.8 Program Algoritma <i>Vector Space Model</i>	104
Gambar 4.9 <i>Code Style</i> Kueri Algoritma Konvensional	105
Gambar 4.10 <i>Code Style</i> Algoritma <i>Vector Space Model</i>	105
Gambar 4.11 Peringkat Jurnal “Implementasi Nilai-Nilai Islam dalam Pendidikan” pada Algoritma Konvensional dan Diurutkan Berdasarkan Nama	106
Gambar 4.12 Peringkat Jurnal “Implementasi Nilai-Nilai Islam dalam Pendidikan” pada Algoritma Konvensional dan Diurutkan Berdasarkan Judul	107
Gambar 4.13 Peringkat Jurnal “Implementasi Nilai-Nilai Islam dalam Pendidikan” pada Algoritma Konvensional dan Diurutkan Berdasarkan Tahun Terbit	107
Gambar 4.14 Peringkat Jurnal “Implementasi Nilai-Nilai Islam dalam Pendidikan” pada Algoritma <i>Vector Space Model</i>	108
Gambar 4.15 Peringkat Jurnal Kontribusi Islam atas Perkembangan Peradaban	114
Gambar 4.16 Peringkat Jurnal Implementasi Nilai-Nilai Islam dalam	114
Gambar 4.17 Peringkat Jurnal Kearifan Lingkungan Masyarakat Lereng Gunung Merapi	115

Gambar 4.18 Peringkat Jurnal Perkembangan Teknologi Komunikasi dan Dampak Terhadap Kehidupan	115
Gambar 4.19 Peringkat Jurnal Melihat Potret Harmonisasi Hubungan Antar Umat Beragama di Indonesia	116
Gambar 4.20 Peringkat Jurnal Perkembangan Islam di Amerika Serikat	117
Gambar 4.21 Peringkat Jurnal Melihat Potret Harmonisasi Hubungan Antar Umat Beragama di Indonesia	117
Gambar 4.22 Peringkat Jurnal Teologis atau Bahasa?	117
Gambar 4.23 Peringkat Jurnal Teknologi Informasi untuk Masyarakat Pedesaan	118
Gambar 4.24 Peringkat Jurnal Sejarah Keuangan Islam	118
Gambar 4.25 Peringkat Jurnal <i>E-Commerce</i> dalam Hukum Islam	119



DAFTAR TABEL

Tabel 2.1 Tinjauan Pustaka	10
Tabel 2.2 Tinjauan Pustaka (Lanjutan)	11
Tabel 2.3 Tinjauan Pustaka (Lanjutan)	12
Tabel 2.4 Tinjauan Pustaka (Lanjutan)	13
Tabel 2.5 Tinjauan Pustaka (Lanjutan)	14
Tabel 2.6 Tinjauan Pustaka (Lanjutan)	15
Tabel 4.1 Data Jurnal dengan ID Dokumen 9016	31
Tabel 4.2 Data Jurnal dengan ID Dokumen 9016 (Lanjutan)	32
Tabel 4.3 Data Jurnal dengan ID Dokumen 9016 (Lanjutan)	33
Tabel 4.4 Data Jurnal dengan ID Dokumen 9016 (Lanjutan)	34
Tabel 4.5 Data Jurnal dengan ID Dokumen 9016 (Lanjutan)	35
Tabel 4.6 Data Jurnal dengan ID Dokumen 9016 (Lanjutan)	36
Tabel 4.7 Data Jurnal dengan ID Dokumen 9016 (Lanjutan)	37
Tabel 4.8 Data Jurnal dengan ID Dokumen 9016 (Lanjutan)	38
Tabel 4.9 Data Jurnal dengan ID Dokumen 9016 (Lanjutan)	39
Tabel 4.10 Data Jurnal dengan ID Dokumen 9016 (Lanjutan)	40
Tabel 4.11 Data Jurnal dengan ID Dokumen 9016 (Lanjutan)	41
Tabel 4.12 Data Jurnal dengan ID Dokumen 9016 (Lanjutan)	42
Tabel 4.13 Data Jurnal dengan ID Dokumen 9016 (Lanjutan)	43
Tabel 4.14 Data Jurnal dengan ID Dokumen 9016 (Lanjutan)	44
Tabel 4.15 Data Jurnal dengan ID Dokumen 9016 (Lanjutan)	45
Tabel 4.16 Data Jurnal dengan ID Dokumen 9016 (Lanjutan)	46
Tabel 4.17 Hasil <i>Case Folding</i> Data	46
Tabel 4.18 Hasil <i>Case Folding</i> Data (Lanjutan)	47
Tabel 4.19 Hasil <i>Case Folding</i> Data (Lanjutan)	48
Tabel 4.20 Hasil <i>Case Folding</i> Data (Lanjutan)	49
Tabel 4.21 Hasil <i>Case Folding</i> Data (Lanjutan)	50
Tabel 4.22 Hasil <i>Case Folding</i> Data (Lanjutan)	51
Tabel 4.23 Hasil <i>Case Folding</i> Data (Lanjutan)	52

Tabel 4.24 Hasil <i>Case Folding</i> Data (Lanjutan)	53
Tabel 4.25 Hasil <i>Case Folding</i> Data (Lanjutan)	54
Tabel 4.26 Hasil <i>Case Folding</i> Data (Lanjutan)	55
Tabel 4.27 Hasil <i>Case Folding</i> Data (Lanjutan)	56
Tabel 4.28 Hasil <i>Case Folding</i> Data (Lanjutan)	57
Tabel 4.29 Hasil <i>Case Folding</i> Data (Lanjutan)	58
Tabel 4.30 Hasil <i>Case Folding</i> Data (Lanjutan)	59
Tabel 4.31 Hasil <i>Case Folding</i> Data (Lanjutan)	60
Tabel 4.32 Hasil <i>Case Folding</i> Data (Lanjutan)	61
Tabel 4.33 Hasil <i>Case Folding</i> Data (Lanjutan)	62
Tabel 4.34 Hasil <i>Tokenizing</i> Data	63
Tabel 4.35 Hasil <i>Tokenizing</i> Data (Lanjutan)	64
Tabel 4.36 Hasil <i>Tokenizing</i> Data (Lanjutan)	65
Tabel 4.37 Hasil <i>Tokenizing</i> Data (Lanjutan)	66
Tabel 4.38 Hasil <i>Tokenizing</i> Data (Lanjutan)	67
Tabel 4.39 Hasil <i>Tokenizing</i> Data (Lanjutan)	68
Tabel 4.40 Hasil <i>Tokenizing</i> Data (Lanjutan)	69
Tabel 4.41 Hasil <i>Tokenizing</i> Data (Lanjutan)	70
Tabel 4.42 Hasil <i>Tokenizing</i> Data (Lanjutan)	71
Tabel 4.43 Hasil <i>Tokenizing</i> Data (Lanjutan)	72
Tabel 4.44 Hasil <i>Tokenizing</i> Data (Lanjutan)	73
Tabel 4.45 Hasil <i>Tokenizing</i> Data (Lanjutan)	74
Tabel 4.46 Hasil <i>Tokenizing</i> Data (Lanjutan)	75
Tabel 4.47 Hasil <i>Tokenizing</i> Data (Lanjutan)	76
Tabel 4.48 Hasil <i>Tokenizing</i> Data (Lanjutan)	77
Tabel 4.49 Hasil <i>Tokenizing</i> Data (Lanjutan)	78
Tabel 4.50 Hasil <i>Tokenizing</i> Data (Lanjutan)	79
Tabel 4.51 Hasil <i>Tokenizing</i> Data (Lanjutan)	80
Tabel 4.52 Hasil <i>Tokenizing</i> Data (Lanjutan)	81
Tabel 4.53 Hasil <i>Tokenizing</i> Data (Lanjutan)	82
Tabel 4.54 Hasil <i>Stopwords Removal</i> Data	82

Tabel 4.55 Hasil <i>Stopwords Removal</i> Data (Lanjutan)	83
Tabel 4.56 Hasil <i>Stopwords Removal</i> Data (Lanjutan)	84
Tabel 4.57 Hasil <i>Stopwords Removal</i> Data (Lanjutan)	85
Tabel 4.58 Hasil <i>Stopwords Removal</i> Data (Lanjutan)	86
Tabel 4.59 Hasil <i>Stopwords Removal</i> Data (Lanjutan)	87
Tabel 4.60 Hasil <i>Stopwords Removal</i> Data (Lanjutan)	88
Tabel 4.61 Hasil <i>Stopwords Removal</i> Data (Lanjutan)	89
Tabel 4.62 Hasil <i>Stopwords Removal</i> Data (Lanjutan)	90
Tabel 4.63 Hasil <i>Stopwords Removal</i> Data (Lanjutan)	91
Tabel 4.64 Hasil <i>Stopwords Removal</i> Data (Lanjutan)	92
Tabel 4.65 Hasil <i>Stopwords Removal</i> Data (Lanjutan)	93
Tabel 4.66 Hasil <i>Stemming</i> Data	93
Tabel 4.67 Hasil <i>Stemming</i> Data (Lanjutan)	94
Tabel 4.68 Hasil <i>Stemming</i> Data (Lanjutan)	95
Tabel 4.69 Hasil <i>Stemming</i> Data (Lanjutan)	96
Tabel 4.70 Hasil <i>Stemming</i> Data (Lanjutan)	97
Tabel 4.71 Hasil <i>Stemming</i> Data (Lanjutan)	98
Tabel 4.72 Hasil <i>Stemming</i> Data (Lanjutan)	99
Tabel 4.73 Hasil <i>Stemming</i> Data (Lanjutan)	100
Tabel 4.74 Hasil <i>Stemming</i> Data (Lanjutan)	101
Tabel 4.75 Hasil <i>Stemming</i> Data (Lanjutan)	102
Tabel 4.76 Uji Penelitian.....	108
Tabel 4.77 Uji Penelitian (Lanjutan).....	109
Tabel 4.78 Uji Penelitian (Lanjutan).....	110
Tabel 4.79 Uji Penelitian (Lanjutan).....	111
Tabel 4.80 Uji Penelitian (Lanjutan).....	112
Tabel 4.81 Uji Penelitian (Lanjutan).....	113

DAFTAR RUMUS

Persamaan 2.1 <i>Term Frequency</i>	19
Persamaan 2.2 <i>Inverse Document Frequency</i>	20
Persamaan 2.3 <i>TF-IDF</i>	21
Persamaan 2.4 <i>Cosine Similarity</i>	23



Studi Komparasi
Algoritma Konvensional dan *Vector Space Model*
pada Peringkat Hasil Penelusuran Jurnal

Azis Alvriyanto
16650036

INTISARI

Semakin banyak dokumen pada database maka semakin lama waktu yang dibutuhkan untuk melakukan pencarian. Bahkan hasil pencarian tidak relevan, karena dokumen harus dipilah dan diteliti relevansinya dengan subjek yang dicari, sehingga mempengaruhi waktu dalam memperoleh dokumen yang sesuai.

Penelitian ini bertujuan mencari algoritma terbaik antara algoritma konvensional dan *Vector Space Model* terhadap hasil penelusuran dokumen. Kedua algoritma tersebut diuji dengan penggunaan kombinasi kata kunci berupa sinonim, frasa, kesalahan tipografi dan tambahan awalan kata.

Data yang berhasil diperoleh sebanyak 635 jurnal dari *digital library* UIN Sunan Kalijaga dengan melakukan 11 pengujian dokumen. Hasil penelitian ini membuktikan bahwa algoritma *Vector Space Model* lebih akurat dalam mengurutkan dokumen sesuai dengan kata kunci yang diberikan daripada algoritma konvensional.

Kata Kunci: *digital library, scraping, tf-idf, vector space model*

**Comparative Study of
Conventional and Vector Space Model Algorithms
on Journal Search Ranking**

Azis Alvriyanto

16650036

ABSTRACT

The more documents in the database, the longer it takes to search. Even the search results are not relevant, because the document must be sorted and examined its relevance to the subject sought, thus affecting the time in obtaining the appropriate document.

This study aims to find the best algorithm between the conventional algorithm and the Vector Space Model for document search results. Both algorithms are tested by using a combination of keywords in the form of synonyms, phrases, typographic errors and additional word prefixes.

Data obtained were 635 journals from digital library of UIN Sunan Kalijaga by conducting 11 document tests. The results of this study prove that the Vector Space Model algorithm is more accurate in sorting documents according to given keywords than conventional algorithms.

Keywords: digital library, scraping, tf-idf, vector space model

BAB I

PENDAHULUAN

1.1. Latar Belakang

Undang-Undang Nomor 43 tahun 2007 menyebutkan bahwa perpustakaan adalah institusi pengelola koleksi karya tulis, karya cetak, dan/atau karya rekam secara profesional dengan sistem yang baku guna memenuhi kebutuhan pendidikan, penelitian, pelestarian, informasi, dan rekreasi para pemustaka. Koleksi perpustakaan adalah semua informasi dalam bentuk karya tulis, karya cetak, dan/atau karya rekam dalam berbagai media yang mempunyai nilai pendidikan, yang dihimpun, diolah, dan dilayankan.

Oleh karena itu, perpustakaan perlu terus mengikuti perkembangan dan beradaptasi agar tetap memberikan pelayanan yang sesuai dengan kebutuhan. Perilaku pengunjung perpustakaan yang menginginkan informasi dengan cepat dan tepat berdampak pada pola pencarian informasi pada koleksi perpustakaan. Dengan menyesuaikan kebutuhan dan perilaku pengguna tersebut, saat ini perpustakaan telah masuk dalam *digital library* untuk menyimpan, mendapatkan, dan menyebarluaskan ilmu pengetahuan dalam format digital sehingga dapat diakses secara daring oleh semua orang melalui internet.

Salah satu tujuan adanya digitalisasi pada lingkungan perpustakaan adalah mempermudah pengunjung perpustakaan dalam mencari koleksi perpustakaan. Salah satu fitur yang dibutuhkan pada laman *digital library* adalah mesin telusur pustaka sesuai kata kunci atau menggunakan fitur seleksi subjek yang mereka cari. Hasil dari penelusuran tersebut dimanfaatkan untuk mendapatkan suatu informasi, bacaan, bahkan menjadi acuan dari tugas kuliah yang melibatkan informasi dari perpustakaan.

Akan tetapi, semakin banyak dokumen pada database maka semakin lama waktu yang dibutuhkan untuk melakukan pencarian. Bahkan hasil pencarian yang tidak relevan, karena dokumen harus dipilah dan diteliti relevansinya dengan subjek yang dicari. Urutan hasil pencarian yang tidak relevan dengan kata kunci akan mempengaruhi waktu dalam memperoleh pustaka yang sesuai.

Pada penelitian Annisa, Sikumbang, & Ananda (2014) yang berjudul “Implementasi Algoritma *Vector Space Model* dalam Pencarian *E-Book*”, informasi *e-book* yang didapatkan menggunakan algoritma *Vector Space Model* mengandung relevansi/keterkaitan dengan yang diharapkan sesuai dengan kata kunci yang telah dimasukkan. Algoritma *Vector Space Model* adalah salah satu metode pencarian yang mengukur relevansi antara kata kunci dengan dokumen yang ada di basis data. Data uji coba merupakan dokumen *e-book* dengan format PDF. Hasil pengujian terhadap pencarian *e-book* diperoleh nilai akurasi dari sistem ini memiliki nilai *recall* rata-rata sebesar 100 % dan nilai *precision* rata-rata sebesar 80%.

Mengacu dari permasalahan tersebut saya akan melakukan penelitian yang berjudul “Studi Komparasi Algoritma Konvensional dan *Vector Space Model* pada Peringkat Hasil Penelusuran Jurnal”. Penelitian ini akan membandingkan peringkat hasil penelusuran jurnal menggunakan algoritma konvensional dan *Vector Space Model*.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah dituturkan oleh penulis, maka rumusan masalah dalam penelitian ini adalah bagaimana peringkat dokumen yang dicari dari hasil penelusuran jurnal menggunakan algoritma konvensional dan *Vector Space Model*.

1.3. Tujuan Penelitian

Penelitian ini bertujuan untuk melakukan perbandingan antara algoritma konvensional dan *Vector Space Model* pada peringkat hasil penelusuran jurnal.

1.4. Batasan Masalah

Agar penelitian ini tidak keluar dari pokok permasalahan, maka batasan masalah dalam penelitian ini adalah:

1. Data yang digunakan adalah jurnal pada *website* Institutional Repository UIN Sunan Kalijaga yang beralamatkan <http://digilib.uin-suka.ac.id>.

2. Bahasa pemrograman yang digunakan dalam penelitian ini adalah bahasa pemrograman PHP dan JavaScript dengan penggunaan *library* di dalamnya.

1.5. Manfaat Penelitian

Manfaat dengan adanya penelitian ini adalah mengetahui perbandingan peringkat dokumen hasil penelusuran jurnal antara algoritma konvensional dan *Vector Space Model*. Dengan adanya penelitian ini, dapat mengetahui algoritma terbaik dan memberikan rekomendasi algoritma penelusuran dokumen kepada pengembang sistem informasi manajemen perpustakaan.

1.6. Keaslian Penelitian

Penelitian terkait algoritma *Vector Space Model* hingga saat ini sudah banyak dilakukan oleh peneliti sebelumnya. Namun, penelitian yang diajukan sebagai Tugas Akhir S1 program studi Teknik informatika Fakultas Sains dan Teknologi UIN Sunan Kalijaga mengenai “Studi Komparasi Algoritma Konvensional dan *Vector Space Model* pada Peringkat Hasil Penelusuran Jurnal” belum pernah dilakukan.

1.7. Sistematika Penulisan

Sebagai gambaran dan kerangka yang jelas mengenai pokok bahasan setiap bab dalam penelitian ini, maka diperlukan sistematika penulisan. Penyusunan laporan tugas akhir ini memiliki sistematika penulisan yang diawali dari BAB I dan diakhiri BAB V. Berikut adalah penjelasan pada tiap-tiap bab dalam laporan penelitian ini:

BAB I PENDAHULUAN

Bab pendahuluan berisikan penjelasan mengenai latar belakang dilakukannya penelitian, rumusan masalah penelitian, batasan masalah, tujuan penelitian, manfaat penelitian, keaslian penelitian, dan sistematika penulisan penelitian.

BAB II TINJAUAN PUSTAKA DAN LANDASAN TEORI

Bab tinjauan pustaka dan landasan teori berisikan mengenai penelitian terdahulu dan teori-teori dasar yang terkait dengan penelitian ini. Teori yang digunakan terdiri dari *text mining*, *text preprocessing*, TF-IDF, algoritma konvensional, algoritma *Vector Space Model*, PHP, dan Javascript.

BAB III METODE PENELITIAN

Bab metode penelitian berisi tentang penjelasan mengenai metode ataupun algoritma yang digunakan serta tahapan-tahapan yang dilakukan untuk mencapai tujuan dan kesimpulan tugas akhir.

BAB IV HASIL DAN PEMBAHASAN

Bab hasil dan pembahasan membahas analisis data dan hasil dari penelitian yang telah dilakukan.

BAB V PENUTUP

Bab penutup berisi tentang kesimpulan dari hasil penelitian yang telah dilakukan. Selanjutnya, kekurangan yang ada pada penelitian dituliskan pada saran untuk pengembangan penelitian dimasa yang akan datang.

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma konvensional dan *Vector Space Model* mampu diimplementasikan pada hasil penelusuran jurnal, sehingga menjadi pengetahuan yang bermanfaat bagi pengembang web *digital library*. Pengetahuan yang dibangun berdasarkan analisis perbandingan algoritma dengan memperhatikan aspek kombinasi sinonim, kesalahan tipografi, frasa dan penambahan awalan kata terhadap kata kunci.

Semua pengujian yang dilakukan, pada algoritma *Vector Space Model* menghasilkan dokumen yang dicari menempati urutan yang lebih tinggi daripada algoritma konvensional. Hal ini membuktikan bahwa algoritma *Vector Space Model* lebih akurat dalam mengurutkan dokumen sesuai dengan kata kunci yang diberikan.

Penggunaan kata kunci dengan kombinasi frasa dan penambahan awalan kata tidak mempengaruhi peringkat dokumen yang dicari. Pada kata kunci dengan kombinasi sinonim mempengaruhi peringkat dokumen yang dicari, namun perbedaan peringkat tidak terlalu signifikan. Sedangkan kesalahan tipografi pada kata kunci sangat mempengaruhi peringkat, bahkan terdapat pengujian menunjukan dokumen yang dicari tidak ditemukan.

5.2 Saran

Pada penelitian ini, tidak lepas dari kekurangan dan kesalahan. Oleh karena itu, penulis menyarankan beberapa hal guna kelanjutan penelitian ini ke depannya, diantaranya adalah sebagai berikut:

1. Penelitian selanjutnya dapat melakukan percobaan dengan menggunakan metode pembobotan kata yang lebih sesuai dengan bahasa dataset.

2. Menggunakan perbandingan algoritma selain perbandingan konvensional dan *Vector Space Model* yang memiliki tujuan sama, yaitu mengurutkan hasil penelusuran dokumen.
3. Penelitian selanjutnya dapat menggunakan selektif *query expansion* (QE) untuk melakukan prediksi terhadap *query* agar tidak merubah makna sesuai dengan kata kunci yang ketik. Sehingga relevansi dokumen yang ditampilkan dari hasil pencarian tetap terjaga.



DAFTAR PUSTAKA

- Amin, F. (2013). Sistem Temu Kembali Informasi dengan Pemeringkatan Metode Vector Space Model. *Jurnal Teknologi Informasi DINAMIK*, 18(2), 122–129. <https://doi.org/10.22441/fifo.v9i1.1444>
- Anhar. (2010). *PHP & MySql Secara Otodidak*. Jakarta: MediaKita. Retrieved from <https://books.google.co.id/books?id=J711efbP9LYC>
- Anna, & Hendini, A. (2018). Implementasi Vector Space Model Pada Sistem Pencarian Mesin Karaoke. *Evolusi : Jurnal Sains Dan Manajemen*, 6(1), 1–6. <https://doi.org/10.31294/evolusi.v6i1.3535>
- Annisa, N., Nengsih, W., & Ananda. (2014). Implementasi Algoritma Vector Space Model dalam Pencarian E-Book. *Jurnal Aksara Komputer Terapan*, 3(2), 1–7.
- Bania Amburika, Chrisnanto, Y. H., & Uriawan, W. (2016). Teknik Vector Space Model (VSM) dalam Penentuan Penanganan Dampak Game Online Pada Anak. *Prosiding SNST Ke-7 Tahun 2016*, 1(1), 10–27. <https://doi.org/10.1103/PhysRevC.6.1023>
- Betha, S. (2014). *Pemrograman Web dengan PHP : Edisi Revisi Kedua*. Bandung: Informatika. Retrieved from <https://openlibrary.telkomuniversity.ac.id/pustaka/98620/pemrograman-web-dengan-php-edisi-revisi-kedua.html>
- Bradji, L., & Boufaida, M. (2011). A Rule Management System for Knowledge Based Data Cleaning. *Intelligent Information Management*, 3, 230–239. <https://doi.org/10.4236/iim.2011.36028>
- Cios, K., Pedrycz, W., Swiniarski, R., & Kurgan, L. (2007). *Data Mining: A Knowledge Discovery Approach*. *Data Mining: A Knowledge Discovery Approach*. Springer. <https://doi.org/10.1007/978-0-387-36795-8>
- Dragut, E., Fang, F., Sistla, A., Yu, C., & Meng, W. (2009). Stop Word and Related Problems in Web Interface Integration. *PVLDB*, 2, 349–360. <https://doi.org/10.14778/1687627.1687667>
- Fauzi, M. A., Arifin, A. Z., & Yuniarti, A. (2014). Term Weighting Berbasis Indeks

- Buku dan Kelas untuk Perangkingan Dokumen Berbahasa Arab. *Lontar Komputer*, 5(2), 435–442.
- Garcia, E. (2006). *The Classic TF-IDF Vector Space Model*. Retrieved from <http://www.minerazzi.com/tutorials/term-vector-3.pdf>
- Hall, M. A., & Smith, L. A. (1999). Feature Selection for Machine Learning: Comparing a Correlation-Based Filter Approach to the Wrapper. In *Proceedings of the Twelfth International Florida Artificial Intelligence Research Society Conference* (pp. 235–239). AAAI Press.
- Indriani, A., Gunawan, & Novianto, E. (2013). Weight Adjusted K-Nearest Neighbor dan Minimum Spanning Tree untuk Information Retrieval System di Perpustakaan STMIK PPKIA Tarakanita Rahmawati Tarakan. *Seminar Nasional Aplikasi Teknologi Informasi 2013*, 18–22.
- Jain, N., Mangal, P., & Mehta, D. (2015). AngularJS: A Modern MVC Framework in JavaScript. *Journal of Global Research in Computer Science*, 5(12), 17–23. Retrieved from <http://www.jgrcs.info/index.php/jgrcs/article/view/952>
- Josi, A., Abdillah, L. A., & Suryayusra. (2014). Penerapan teknik web scraping pada mesin pencari artikel ilmiah. *ArXiv14105777 Cs*, 159–164. Retrieved from <http://arxiv.org/abs/1410.5777>
- Juliasari, N., & Sitompul, J. C. (2012). Aplikasi Search Engine Dengan Metode Depth First Search (DFS). *Jurnal Teknik Informatika Universitas Budi Luhur*. ISSN : 1693 -9166, 9(1), 9–12. <https://doi.org/10.1109/20.312267>
- Kim, B., Im, C., & Jung, H. (2011). Suspicious Malicious Web Site Detection with Strength Analysis of a JavaScript Obfuscation. *International Journal of Advanced Science and Technology*, 26(January), 19–32.
- Koesheryatin, & Suryana, T. (2014). *Aplikasi Internet Menggunakan HTML, CSS, dan JavaScript*. Elex Media Komputindo. Retrieved from <https://books.google.co.id/books?id=Mt1MDwAAQBAJ>
- Mandala, R. (2004). Bahan Kuliah Sistem Temu Kembali Informasi. *Institut Teknologi Bandung: Departemen Teknik Informatika*.
- Pasma, C. M., Rosiani, U. D., & Ariyanto, R. (2015). Pengembangan Sistem Pendeteksi Kemiripan Karya Pada Inaicta 2013. *Jurnal Informatika Polinema*,

- I(4)*, 14. <https://doi.org/10.33795/jip.v1i4.117>
- Prilianti, K. R., & Wijaya, H. (2014). Aplikasi Text Mining untuk Automasi Penentuan Tren Topik Skripsi dengan Metode K-Means Clustering. *Jurnal Cybermatika*, 2(1), 1–6.
- Salton, G. (1989). *Automatic text processing: the transformation, analysis, and retrieval of information by computer*. Addison-Wesley. Retrieved from <https://books.google.co.id/books?id=wb8SAQAAMAAJ>
- Sibero, A. F. K. (2011). *Kitab suci web programming*. Yogyakarta: Mediakom.
- Sukisno. (2019). Studi Komparatif Pemanfaatan Vector Space Model Pada Penerapan Algoritma Nazief Adriani , K-Nearest Neighbor dan Fungsi Jaccard: Kasus Prototipe Aplikasi Katagorisasi Teks Bahasa Indonesia. *Jurnal Teknik Informatika UNIS*, 4(2), 66–74.
- Sunyoto, A. (2007). *Ajax Membangun Web dengan Teknologi Asynchrone JavaScript&XML*. (D. Hardjono, Ed.). Yogyakarta: Penerbit Andi. Retrieved from <https://books.google.co.id/books?id=dM1ninTPmGYC>
- Sutarman. (2003). *Membangun Aplikasi Web dengan PHP dan MySQL*. Graha Ilmu, Yogyakarta.
- Tata, S., & Patel, J. (2007). Estimating the Selectivity of tf-idf based Cosine Similarity Predicates. *Sigmod Record*, 36. <https://doi.org/10.1145/1361348.1361351>
- Turland, M. (2010). *php|architect's Guide to Web Scraping with PHP*. Toronto: Marco Tabini & Associates, Inc.
- Wahib, A., Pasnur, Santika, P. P., & Arifin, A. Z. (2015). Perangkingan Dokumen Berbahasa Arab Menggunakan Latent Semantic Indexing. *Jurnal Buana Informatika*, 6(2), 83–92. <https://doi.org/10.24002/jbi.v6i2.411>
- Web Site Scraper - The Most Effective Tool for Web Data Extraction. (n.d.). Retrieved from <http://www.thecomputeradvisor.net/web-site-scraper-the-most-effective-tool-for-web-data-extraction/>
- Weiss, S., Indurkha, N., Zhang, T., & Damerau, F. (2004). *Text Mining: Predictive Methods for Analyzing Unstructured Information*. *Text Mining: Predictive Methods for Analyzing Unstructured Information*.

<https://doi.org/10.1007/978-0-387-34555-0>

Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques*. Elsevier Science. Retrieved from <https://books.google.co.id/books?id=bDtLM8CODsQC>

Zipf, G. K. (1932). *Selected Studies of the Principle of Relative Frequency in Language*. Harvard University Press. Retrieved from <https://books.google.co.id/books?id=Brp2XwAACAAJ>

