

**PERHITUNGAN FREKUENSI REFERENSI KARYA ILMIAH TUGAS
AKHIR DALAM BENTUK *VECTOR SPACE MODEL* MENGGUNAKAN
COSINE SIMILARITY DAN *HAMMING DISTANCE*
Studi Kasus: Program Studi S1 Informatika UNS**

SKRIPSI

Diajukan untuk Memenuhi Salah Satu Syarat Mencapai Gelar Strata Satu



Disusun oleh :
Lydia Permata Sari
NIM. M0510031

**PROGRAM STUDI INFORMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS SEBELAS MARET
SURAKARTA
2017**

SKRIPSI

Perhitungan Frekuensi Referensi Karya Ilmiah Tugas Akhir Dalam Bentuk
Vector Space Model Menggunakan *Cosine Similarity* Dan *Hamming Distance*
Studi Kasus: Program Studi S1 Informatika UNS

Disusun oleh :

Lydia Permata Sari

NIM. M0510031

Skripsi ini telah disetujui untuk dipertahankan di hadapan dewan penguji,
Pada tanggal : 26 Juli 2017

PERSETUJUAN PEMBIMBING

Tugas Akhir ini telah disetujui oleh :

Pembimbing 1,



Ristu Saptono, S.Si, M.T.

NIP. 19790210 200212 1001

Pembimbing 2,



Esti Suryani, S.Si, M.Kom.

NIP. 19761129 20081 2 2001

SKRIPSI

Perhitungan Frekuensi Referensi Karya Ilmiah Tugas Akhir Dalam Bentuk
Vector Space Model Menggunakan Cosine Similarity Dan Hamming Distance

Studi Kasus: Program Studi S1 Informatika UNS

Disusun oleh :

Lydia Permata Sari

NIM. M0510031

Telah dipertahankan di hadapan Dewan Penguji

Pada tanggal : 26 Juli 2017

Susunan Dewan Penguji

1. Ristu Saptono, S.Si, M.T.
NIP. 19790210 200212 1 001
2. Esti Suryani, S.Si., M.Kom.
NIP. 19761129 200812 2 001
3. Afrizal Doewes, S.Kom., M.Sc.
NIP. 19850831 201212 1 004
4. Sari Widya Sihwi, S.Kom., M. T. I
NIP. 19830412 200912 2 003

()
()
()
()

Disahkan oleh:

Kepala Program Studi Informatika,



Drs. Bambang Harjito, M.App.Sc., Ph.D

NIP. 19621130 199103 1 002

MOTTO

“Sesuatu yang belum dikerjakan, seringkali tampak mustahil. Kita baru yakin kalau kita telah berhasil melakukannya dengan baik”

(Evelyn Underhill)

“Learn from the past, live for the today, and plan for tomorrow”



PERSEMBAHAN

Karya ini Penulis persembahkan kepada:

“Bapak Djoko Suprpto dan Ibu Lies Marliana, kedua orang tua yang tercinta”

“Devi Nurmala Sari, kakak yang selalu mendukung”

“Farrel Viendra Syahputra, keponakan yang memberikan keceriaan dalam kesedihan”

“B’tara Linggamurti yang selalu menemani dan memberi dukungan”

“Teman – teman seluruh Informatika, khususnya Hedik, Aji, Cerren, Taufik, Praditio, Ashar, Miftah, Viko dan April”

KATA PENGANTAR

Puji dan syukur penulis ucapkan kepada Allah SWT, yang hanya karena rahmat dan karunia-Nya, penulis dapat menyelesaikan Tugas Akhir dengan judul “Perhitungan Frekuensi Referensi Karya Ilmiah Tugas Akhir Dalam Bentuk *Vector Space Model* Menggunakan *Cosine Similarity* dan *Hamming Distance* pada studi kasus Program Studi S1 Informatika UNS”. Laporan Skripsi ini disusun untuk memenuhi salah satu syarat dalam memperoleh gelar Sarjana Komputer pada Program Studi Informatika Universitas Sebelas Maret.

Dalam melakukan penelitian dan penyusunan laporan Skripsi ini penulis telah mendapatkan banyak dukungan dan bantuan dari berbagai pihak. Penulis mengucapkan terimakasih yang tak terhingga kepada:

1. Bapak, ibu, dan kakak yang senantiasa selalu memberikan motivasi, doa dan dukungannya.
2. Bapak Prof. Ir. Ari Handono Ramelan, M.Sc.(Hons), Ph.D. selaku Dekan Fakultas MIPA Universitas Sebelas Maret.
3. Bapak Drs. Bambang Harjito, M.APP.Sc, Ph.D. selaku Kepala Program Studi Informatika Universitas Sebelas Maret.
4. Bapak Ristu Saptono, S.Si., M.T. dan Ibu Esti Suryani, S.Si.,M.Kom. selaku dosen pembimbing yang telah dengan penuh kesabaran dan ketulusan memberikan ilmu dan bimbingan terbaik kepada Penulis.
5. Para Dosen Program Studi Informatika Universitas Sebelas Maret yang telah memberikan bekal ilmu kepada Penulis.

Semoga Tugas Akhir ini bermanfaat dan memberikan inspirasi bagi semua pihak yang berkepentingan.

Surakarta, 26 Juli 2017

Penulis

***FREQUENCY CALCULATION OF SCIENTIFIC WORK OF THESIS IN THE
FORM OF VECTOR SPACE MODEL USING COSINE SIMILARITY AND
HAMMING DISTANCE***

CASE STUDY: DEPARTMENT OF INFORMATICS UNS

Lydia Permata Sari

Department of Informatics. Faculty of Mathematics and Natural Sciences.

Universitas Sebelas Maret

ABSTRACT

Bibliography on final project contains references that are used by writer. References cited in the form of books, journals, magazines, newspapers or internet. Scientific works in the form of journals and proceedings used as references also are listed in the bibliography. Calculation of scientific works that have been noted in the bibliography can be used to calculate the frequency of the use of scientific works used by undergraduate students of Informatics UNS in the form of Vector Space Model using Cosine Similarity and Hamming Distance.

Calculation of the frequency of use of the work of the scientific method using Vector Space Model to find the value of the weight of the title scientific papers where the results will be used to identify the title similarity search papers by using the method of Cosine Similarity. Hamming Distance method used to calculate the distance of the similarity of author names and year. The data used for this research is 100 final project documents. There are 47% of eligible documents Harvard and APA.

The result of the calculation has been done i.e. There are 27 same scientific works, and produces 12 scientific papers that have a frequency of occurrence of twice or more. There are 5 scientific papers originating from within the country, and 7 scientific papers which came from international.

Keywords: *Cosine Similarity, Frequency of Scientific Papers, Hamming Distance, Vector Space Model.*

**PERHITUNGAN FREKUENSI REFERENSI KARYA ILMIAH TUGAS
AKHIR DALAM BENTUK *VECTOR SPACE MODEL* MENGGUNAKAN
COSINE SIMILARITY DAN *HAMMING DISTANCE***

Studi Kasus: Program Studi S1 Informatika UNS

Lydia Permata Sari

Program Studi Informatika. Fakultas Matematika dan Ilmu Pengetahuan Alam.

Universitas Sebelas Maret

ABSTRAK

Daftar pustaka pada tugas akhir berisi sumber pustaka yang digunakan oleh penulis. Sumber pustaka yang dikutip berupa buku, jurnal, majalah, surat kabar atau internet. Karya ilmiah berupa jurnal dan *proceeding* yang digunakan sebagai sumber pustaka juga dicantumkan di dalam daftar pustaka. Perhitungan karya ilmiah yang telah dicantumkan di daftar pustaka dapat digunakan untuk menghitung frekuensi penggunaan karya ilmiah yang digunakan oleh mahasiswa S1 Informatika UNS dalam bentuk *Vector Space Model* menggunakan metode *Cosine Similarity* dan *Hamming Distance*.

Perhitungan frekuensi penggunaan karya ilmiah menggunakan metode *Vector Space Model* untuk mencari nilai bobot judul karya ilmiah dimana hasilnya akan digunakan untuk mengidentifikasi pencarian kesamaan judul karya ilmiah dengan menggunakan metode *Cosine Similarity*. Metode *Hamming Distance* digunakan untuk menghitung jarak kemiripan nama pengarang dan tahun. Data yang dijadikan untuk bahan penelitian ini sebanyak 100 dokumen tugas akhir. Terdapat 47% dokumen yang memenuhi syarat penulisan Harvard / APA.

Hasil dari perhitungan yang telah dilakukan yaitu terdapat 27 judul karya ilmiah yang sama, dan menghasilkan 12 karya ilmiah yang memiliki frekuensi kemunculan sebanyak 2 kali atau lebih. Terdapat 5 karya ilmiah yang berasal dari dalam negeri, dan 7 karya ilmiah yang berasal dari internasional.

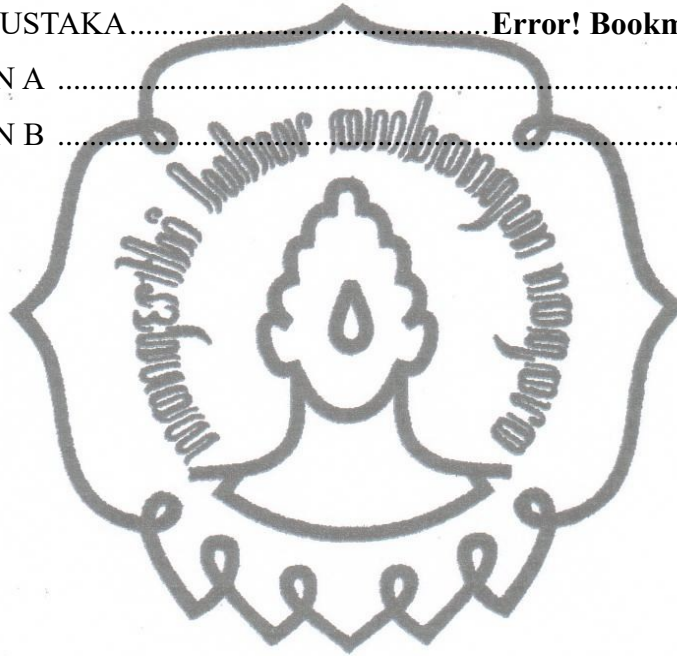
Kata kunci : Cosine Similarity, Frekuensi Karya Ilmiah, Hamming Distance, Vector Space Model.

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
MOTTO	i
PERSEMBAHAN.....	ii
KATA PENGANTAR.....	iii
ABSTRACT.....	iv
ABSTRAK.....	v
DAFTAR ISI.....	vi
DAFTAR TABEL	ix
DAFTAR GAMBAR	x
DAFTAR LAMPIRAN.....	xi
BAB I PENDAHULUAN.....	Error! Bookmark not defined.
1.1 Latar Belakang	Error! Bookmark not defined.
1.2 Rumusan Masalah.....	Error! Bookmark not defined.
1.3 Batasan Masalah.....	Error! Bookmark not defined.
1.4 Tujuan Penelitian	Error! Bookmark not defined.
1.5 Manfaat Penelitian.....	Error! Bookmark not defined.
1.6 Sistematika Penulisan	Error! Bookmark not defined.
BAB II TINJAUAN PUSTAKA.....	Error! Bookmark not defined.
2.1 Dasar Teori	Error! Bookmark not defined.
2.1.1. Tata-tata Cara Penulisan Daftar Pustaka.....	Error! Bookmark not defined.
2.1.2. <i>Preprocessing Process</i>	Error! Bookmark not defined.
2.1.3. <i>Naïve Bayes Classifier</i>	Error! Bookmark not defined.
2.1.4. <i>Vector Space Model</i>	Error! Bookmark not defined.
2.1.5. <i>Hamming Distance</i>	Error! Bookmark not defined.
2.1.6. <i>Regular Expression</i>	Error! Bookmark not defined.
2.2 Penelitian Terkait	Error! Bookmark not defined.

2.2.1.	Perbandingan Algoritma Stemming Porter dengan Algoritma Nazief & Adriani untuk Stemming Dokumen Teks Bahasa Indonesia	Error! Bookmark not defined.
2.2.2.	Klasifikasi Status Gizi Menggunakan <i>Naïve Bayes Classification</i>	Error! Bookmark not defined.
2.2.3.	Mengukur Tingkat Kesamaan Paragraf Menggunakan Vector Space Model untuk Mendeteksi Plagiarism	Error! Bookmark not defined.
2.3	Kerangka Pemikiran	Error! Bookmark not defined.
BAB III METODOLOGI PENELITIAN		Error! Bookmark not defined.
3.1.	Pengumpulan Data	Error! Bookmark not defined.
3.2.	Penerapan Metode	Error! Bookmark not defined.
3.3.1.	<i>Input Data</i>	Error! Bookmark not defined.
3.3.2.	Pemisahan Rangkaian Kata	Error! Bookmark not defined.
3.3.3.	Judul Karya Ilmiah	Error! Bookmark not defined.
3.3.3.1.	Identifikasi Bahasa	Error! Bookmark not defined.
3.3.3.2.	Penghapusan kata-kata yang tidak digunakan	Error! Bookmark not defined.
3.3.3.3.	Pemilahan Kata Dasar	Error! Bookmark not defined.
3.3.3.4.	Pembobotan Kata	Error! Bookmark not defined.
3.3.3.5.	Pengidentifikasian Judul Karya Ilmiah	Error! Bookmark not defined.
3.3.4.	Nama Pengarang dan Tahun	Error! Bookmark not defined.
3.3.4.1.	Penghapusan kata-kata yang tidak digunakan	Error! Bookmark not defined.
3.3.4.2.	Pengidentifikasian nama pengarang dan tahun ..	Error! Bookmark not defined.
3.3.5.	Perhitungan Frekuensi	Error! Bookmark not defined.
BAB IV HASIL DAN PEMBAHASAN		Error! Bookmark not defined.
4.1.	Deskripsi Data	Error! Bookmark not defined.
4.2.	Penerapan Metode	Error! Bookmark not defined.
4.2.1.	<i>Input Data</i>	Error! Bookmark not defined.
4.2.2.	Pemisahan Rangkaian Kata	Error! Bookmark not defined.
4.2.3.	Judul Karya Ilmiah	Error! Bookmark not defined.
4.2.3.1.	Identifikasi Bahasa	Error! Bookmark not defined.
4.2.3.2.	Penghapusan kata-kata yang tidak digunakan	Error! Bookmark not defined.

4.2.3.3. Pemilahan kata dasar	Error! Bookmark not defined.
4.2.3.4. Pembobotan Kata.....	Error! Bookmark not defined.
4.2.3.5. Pengidentifikasi Judul Karya Ilmiah	Error! Bookmark not defined.
4.2.4. Nama Pengarang dan Tahun	Error! Bookmark not defined.
4.2.5. Perhitungan Frekuensi	Error! Bookmark not defined.
BAB V.....	Error! Bookmark not defined.
5.1. Kesimpulan.....	Error! Bookmark not defined.
5.2. Saran.....	Error! Bookmark not defined.
DAFTAR PUSTAKA.....	Error! Bookmark not defined.
LAMPIRAN A	49
LAMPIRAN B	56



DAFTAR TABEL

Tabel 2. 1 Contoh penulisan daftar pustaka APA Style.....	7
Tabel 2. 2 Langkah Algoritma Porter.....	9
Tabel 2. 3 Contoh Perhitungan Hamming Distance pada himpunan kata	19
Tabel 2. 4 Meta-karakter <i>RegEx</i> pada PHP	19
Tabel 2. 5 Keterkaitan penelitian dengan penelitian sebelumnya.....	22
Tabel 4. 1 Rincian Jumlah Data.....	27
Tabel 4. 2 Contoh penulisan karya ilmiah.....	28
Tabel 4. 3 <i>RegEx</i> yang digunakan.....	29
Tabel 4. 4 Pemisahan Rangkaian Kata.....	30
Tabel 4. 5 Hasil Tokenizing Judul Karya Ilmiah.....	31
Tabel 4. 6 Term Frequency.....	31
Tabel 4. 7 Perhitungan Prior	32
Tabel 4. 8 Perhitungan Likelihood.....	33
Tabel 4. 9 Perhitungan Posterior.....	33
Tabel 4. 10 Hasil Identifikasi Bahasa.....	34
Tabel 4. 11 Proses Penghapusan Kata dan Stemming.....	35
Tabel 4. 12 Contoh Hasil Text Preprocessing	36
Tabel 4. 13 Perhitungan Document Frequency (df).....	37
Tabel 4. 14 Perhitungan invers document frequency (idf).....	37
Tabel 4. 15 Perhitungan Nilai Bobot (W)	38
Tabel 4. 16 Tabel Perhitungan Panjang Vektor	39
Tabel 4. 17 Tabel Perhitungan Cosine Similarity.....	40
Tabel 4. 18 Nama Pengarang dan Tahun.....	41
Tabel 4. 19 Hasil Preprocessing Nama Pengarang	41
Tabel 4. 20 Perhitungan Hamming Distance untuk Nama Pengarang.....	42
Tabel 4. 21 Perhitungan Hamming Distance untuk Tahun	42
Tabel 4. 22 Tabel Frekuensi Hasil Perhitungan	43

DAFTAR GAMBAR

Gambar 2. 1 Gambar Matriks Dokumen-Term..... **Error! Bookmark not defined.**

Gambar 3. 1 Flowchart Penerapan Metode Pada Sistem**Error! Bookmark not defined.**



DAFTAR LAMPIRAN

LAMPIRAN A	49
LAMPIRAN B	56



