

**IMPLEMENTASI *TERM FREQUENCY-INVERS DOCUMENT*
FREQUENCY (TF-IDF) DAN *COSINE SIMILARITY*
DALAM PENENTUAN *REVIEWER* PENELITIAN
DOSEN UNIVERSITAS PENDIDIKAN INDONESIA**

SKRIPSI

diajukan untuk memenuhi bagian dari
Syarat memperoleh gelar Sarjana Komputer
pada Departemen Pendidikan Ilmu Komputer



Oleh :

HAZMI RAMADHAN ADLI

1301478

**PROGRAM STUDI ILMU KOMPUTER
DEPARTEMEN PENDIDIKAN ILMU KOMPUTER
FAKULTAS PENDIDIKAN MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS PENDIDIKAN INDONESIA**

2020

Hazmi Ramadhan Adli, 2020

**IMPLEMENTASI *TERM FREQUENCY INVERS DOCUMENT FREQUENCY (TF-IDF)* DAN *COSINE SIMILARITY*
DALAM PENENTUAN *REVIEWER* PENELITIAN DOSEN UNIVERSITAS PENDIDIKAN INDONESIA**

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

**IMPLEMENTASI *TERM FREQUENCY-INVERS DOCUMENT FREQUENCY*
(*TF-IDF*) DAN *COSINE SIMILARITY* DALAM PENENTUAN *REVIEWER*
PENELITIAN DOSEN UNIVERSITAS PENDIDIKAN INDONESIA**

Oleh :

Hazmi Ramadhan Adli

1301478

Sebuah Skripsi yang diajukan untuk Memenuhi Salah Satu Syarat Memperoleh
Gelar Sarjana Komputer di Fakultas Pendidikan Matematika dan Ilmu Pengetahuan Alam

© Hazmi Ramadhan Adli 2020

Universitas Pendidikan Indonesia

Agustus 2020

Hak Cipta Dilindungi Undang-Undang

Skripsi ini tidak boleh diperbanyak seluruhnya atau sebagian, dengan dicetak ulang, difoto
kopi atau cara lainnya tanpa izin dari penulis

Hazmi Ramadhan Adli, 2020

**IMPLEMENTASI *TERM FREQUENCY INVERS DOCUMENT FREQUENCY (TF-IDF)* DAN *COSINE SIMILARITY*
DALAM PENENTUAN *REVIEWER* PENELITIAN DOSEN UNIVERSITAS PENDIDIKAN INDONESIA**

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

LEMBAR PENGESAHAN

**IMPLEMENTASI *TERM FREQUENCY-INVERS DOCUMENT FREQUENCY*
(*TF-IDF*) DAN *COSINE SIMILARITY* DALAM PENENTUAN *REVIEWER*
PENELITIAN DOSEN UNIVERSITAS PENDIDIKAN INDONESIA**

Oleh :

HAZMI RAMADHAN ADLI

1301478

DISETUJUI DAN DISAHKAN OLEH :

Pembimbing I,

Prof. Dr. Munir, M.IT.

NIP. 1966032520001121001

Pembimbing II,

Dr. Rani Megasari, M.T.

NIP. 19870524014042002

Mengetahui,

Ketua Departemen Pendidikan Ilmu Komputer,

Lala Septem Riza, MT., Ph.D.

NIP. 197809262008121001

**IMPLEMENTASI *TERM FREQUENCY-INVERS DOCUMENT*
FREQUENCY (TF-IDF) DAN *COSINE SIMILARITY*
DALAM PENENTUAN *REVIEWER* PENELITIAN
DOSEN UNIVERSITAS PENDIDIKAN INDONESIA**

Oleh

Hazmi Ramadhan Adli — hazmi.ramadhan.adli@student.upi.edu
1301478

ABSTRAK

Penelitian merupakan salah satu kegiatan yang wajib dilaksanakan oleh Dosen Universitas Pendidikan Indonesia. LPPM selaku pengelola kegiatan penelitian di UPI, sebelum penelitian dilaksanakan, tahapan yang harus dilalui yaitu pengujian atau *review* proposal penelitian oleh Dosen Penguji (*reviewer*). Dalam penentuan *reviewer* dilakukan dengan cara melihat kemiripan antara variabel peneliti dengan *reviewer* dan *availabilitas* kuota *me-reviewer* dari seorang *reviewer*. Penentuan *reviewer* ini dilakukan dengan menerapkan metode *Term Frequency Invers Document Frequency (TF-IDF)* dan *Cosine Similarity*, untuk menghitung kemiripan antara *query* dan *document* dalam hal ini abstrak dan bidang keilmuan. Sehingga dengan dapat menghasilkan rekomendasi *reviewer* dengan akurasi 83,3% yang sudah divalidasi oleh ahli.

Kata kunci: Penelitian, *reviewer*, *TF-IDF*, *Cosine Similarity*

Hazmi Ramadhan Adli, 2020

IMPLEMENTASI *TERM FREQUENCY INVERS DOCUMENT FREQUENCY (TF-IDF)* DAN *COSINE SIMILARITY* DALAM PENENTUAN *REVIEWER* PENELITIAN DOSEN UNIVERSITAS PENDIDIKAN INDONESIA

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

**IMPLEMENTATION OF TERM FREQUENCY-INVERS DOCUMENT
FREQUENCY (TF-IDF) AND COSINE SIMILARITY
IN THE DETERMINATION OF THE LECTURER'S RESEARCH
REVIEWER OF INDONESIA UNIVERSITY OF EDUCATION**

Arranged by

Hazmi Ramadhan Adli --- hazmi.ramadhan.adli@student.upi.edu

1301478

ABSTRACT

Research is one of the activities that must be carried out by lecturers at the Indonesia University of Education. LPPM as the manager of research activities at UPI, before the research is carried out, the stages that must be passed are testing or reviewing research proposals by examining lecturers (reviewers). In determining reviewers, it is done by looking at the similarities between the variables of the researcher and the reviewer and the availability of the reviewer quota from a reviewer. This reviewer was determined by applying the Term Frequency Invers Document Frequency TF-IDF and Cosine Similarity methods, to calculate the similarity between the query and document, in this case abstract and scientific fields. So that by being able to produce reviewer recommendations with an accuracy of 83.3% which have been validated by experts.

Keywords : Research, reviewers, TF-IDF, Cosine Similarity

Hazmi Ramadhan Adli, 2020

IMPLEMENTASI TERM FREQUENCY INVERS DOCUMENT FREQUENCY (TF-IDF) DAN COSINE SIMILARITY DALAM PENENTUAN REVIEWER PENELITIAN DOSEN UNIVERSITAS PENDIDIKAN INDONESIA

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

DAFTAR ISI

PERNYATAAN.....	i
ABSTRAK	ii
ABSTRACT.....	iii
KATA PENGANTAR	iv
UCAPAN TERIMA KASIH.....	v
DAFTAR ISI.....	vii
DAFTAR TABEL.....	x
DAFTAR GAMBAR	xi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	8
1.3 Tujuan Penelitian.....	8
1.4 Batasan Masalah.....	8
1.5 Manfaat Penelitian.....	9
1.6 Sistematika Penulisan.....	10
BAB II TINJAUAN PUSTAKA.....	11
2.1 Lembaga Penelitian dan Pengabdian Kepada Masyarakat UPI	11
2.2 Sistem Pendukung Keputusan (<i>Descision Support System</i>).....	12
2.2.1 Definisi SPK	12
2.2.2 Tujuan SPK.....	14
2.2.3 Karakteristik dan Nilai Guna Sistem	17
2.3 Sistem Rekomendasi	17
2.4 <i>Website</i>	18
2.5 <i>Markup Languange</i>	19
2.5.1 HTML (<i>Hypertext Markup Languange</i>)	19
2.5.2 XML.....	19
2.6 PHP.....	20
2.7 Framework Yii.....	22
2.7.1 Konsep MVC pada Framework Yii.....	23
2.7.2 Fitur-fitur Framework Yii.....	23
2.8 Library	25

Hazmi Ramadhan Adli, 2020

IMPLEMENTASI TERM FREQUENCY INVERS DOCUMENT FREQUENCY (TF-IDF) DAN COSINE SIMILARITY DALAM PENENTUAN REVIEWER PENELITIAN DOSEN UNIVERSITAS PENDIDIKAN INDONESIA

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

2.9	MySQL.....	25
2.10	XAMPP	26
2.11	<i>Text Mining</i>	27
2.11.1	Information Extraction	28
2.11.2	Text Preprocessing	29
2.12	Pembobotan Kata (<i>Term Weighting</i>).....	33
2.13	<i>Cosine Similarity</i> (Kesamaan Kosinus).....	34
BAB III METODOLOGI PENELITIAN.....		36
3.1	Desain Penelitian	36
3.2	Metode Penelitian.....	38
3.2.1	Metode Pengumpulan Data	38
3.2.2	Metode Pengembangan Perangkat Lunak	38
3.3	Alat dan Bahan Penelitian	39
3.3.1	Alat Penelitian	39
3.3.2	Bahan Penelitian.....	40
BAB IV HASIL DAN PEMBAHASAN		41
4.1	Pengumpulan Data	41
4.2	Alur Model Sistem Penentuan Reviewer	44
4.3	Data Uji	46
4.4	Tahap <i>Text Preprocessing</i>	46
4.4.1	Tahap <i>Case Folding</i>	47
4.4.2	Tahap <i>Tokenizing</i>	48
4.4.3	Tahap <i>Stopwords Removal</i>	48
4.4.4	Tahap <i>Stemming</i>	49
4.5	Tahap Pembobotan Kata (<i>Term Weighting</i>).....	50
4.6	Tahap <i>Cosine Similarity</i>	55
4.7	Pembangunan Perangkat Lunak	58
4.9.1	Analisis Kebutuhan	58
4.9.2	Desain.....	59
4.9.3	Implementasi	61
4.9.4	Pengujian Perangkat Lunak.....	61
4.8	Desain Eksperimen.....	62

4.9	Hasil dan Analisis Eksperimen.....	63
4.9.1	Hasil Eksperimen	63
4.9.2	Analisis Hasil Eksperimen	78
BAB V KESIMPULAN DAN SARAN.....		96
5.1	Kesimpulan.....	96
5.2	Saran	97
DAFTAR PUSTAKA		98
LAMPIRAN		

DAFTAR PUSTAKA

- Agrawal, R., & Batra, M. (2013). A Detailed Study on Text Mining Techniques. *International Journal of Soft Computing and Engineering (IJSCE)*, 118-121.
- Appelt, D. E., & Israel, D. J. (1999). Introduction to Information Extraction Technology. *International Joint Conferences on Artificial Intelligence Organization*.
- Asanov, D. A. (2011). Algorithm and Method in Recommender System.
- Asfi, M., & Sari, R. P. (2010). Sistem Penunjang Keputusan Seleksi Mahasiswa Berprestasi Menggunakan Metode AHP (Studi Kasus: STMIK CIC Cirebon). *Jurnal Informatika*, 131-144.
- Batubara, F. A. (2012). Perancangan Website Pada PT Ratu Enim Palembang. *Jurnal Ilmu Pengetahuan dan Teknologi Terapan*, 15-27.
- Christina, S. (2014). Kinerja Cosine Similarity dan Semantic Similarity dalam Pengidentifikasian Relevansi Nomor Halaman pada Daftar Indeks Istilah. *SENTIKA*, 35-40.
- Dang, S., & Ahmad, P. H. (2015). A Review of Text Mining Techniques Associated with Various Application Areas. *International Journal of Science and Research (IJSR)*, 2461-2466.
- Destiningrum, M., & Adrian, Q. J. (2017). Sistem Informasi Penjadwalan Dokter Berbasis Web Dengan Menggunakan Framework Codeigniter (Studi Kasus : Rumah Sakit Yukum Medical Centre). *Jurnal Teknoinfo*, 30-37.
- Deviana, H. (2011). Penerapan XML Web service Pada Sistem Distribusi Barang. *Jurnal Generic*, 61-70.
- Djamal, R. A., Maharani, W., & Kurniati, A. P. (2010). Analisis dan Implementasi Metode Item-Based Clustering Hybrid Pada Recommender System. *Konferensi Nasional Sistem dan Informatika*, (hal. 216-222). Bali.
- Dwijaya, I. F. (2010). Sistem Pendukung Keputusan Kenaikan Jabatan pada PT Sysmex Menggunakan Metode Profile Matching. *Universitas Komputer Indonesia*.
- Eniyati, S. (2011). Perancangan Sistem Pendukung Pengambilan Keputusan untuk Penerimaan Beasiswa dengan Metode SAW (Simple Additive Weighting). *Jurnal Teknologi Informasi DINAMIK*, 171-176.
- Fachrurrozi, M., & Manik, A. A. (2015). Perbaikan Ejaan Kata pada Dokumen Bahasa Indonesia dengan Metode Cosine Similarity. *Seminar Nasional Rekayasa Komputer dan Aplikasinya*. Padang: Universitas Andalas.
- Februariyanti, H., & Zuliarso, E. (2012). Rancang Bangun Sistem Perpustakaan untuk Jurnal Elektronik. *Jurnal Teknologi Informasi DINAMIK*, 124-132.
- Firman, A., Wowor, H. F., & Najooan, X. (2016). Sistem Informasi Perpustakaan Online Berbasis Web. *E-journal Teknik Elektro dan Komputer*, 29-36.

- Fitri, M. (2013). Perancangan Sistem Temu Balik Informasi Dengan Metode Pembobotan Kombinasi Tf-Idf Untuk Pencarian Dokumen Berbahasa Indonesia. *Jurnal Sistem dan Teknologi Informasi*, 80-85.
- Habert, B., Adda, G., Decker, M. A., & de Mar'euil, P. B. (1998). Towards Tokenization Evaluation. *Proceedings of LREC*. limsi.fr.
- Hidayat, R. (2010). *Cara Praktis Membangun Website Gratis*. Jakarta: PT Elex Media Komputindo.
- Imbar, R. V., Adelia, Ayub, M., & Rehatta, A. (2014). Implementasi Cosine Similarity dan Algoritma Smith-Waterman untuk Mendeteksi Kemiripan Teks. *Jurnal Informatika*, 31-42.
- Indonesia, R. (2005). *Undang-Undang Republik Indonesia Nomor 14 Tahun 2005*. Jakarta.
- Indonesia, R. (2012). *Undang-undang Republik Indonesia No. 12 Tahun 2012 Tentang Pendidikan Tinggi*. Jakarta: Sekretariat Negara.
- Irianto, W. A. (2013). *Penentuan Tingkat Plagiarisme Dokumen Penelitian menggunakan Centroid Linkage Hierarchical Method (CLHM)*. . Malang: Universitas Brawijaya.
- Kadhim, A. I. (2018). An Evaluation of Preprocessing Techniques for Text Classification. *International Journal of Computer Science and Information Security (IJCSIS)*, 22-32.
- Karhendana, A. (2008). *Pemanfaatan Document Clustering Pada Agregator Berita*. Bandung.
- Kohila, R., & Arunesh, D. K. (2016). TEXT MINING: TEXT SIMILARITY MEASURE FOR NEWS ARTICLES BASED ON. *Global Journal of Engineering Science and Research Management*, 35-42.
- Konchady, M. (2006). *Text Mining Application Programming*. Boston: Charles River Media.
- Kurniawan, A., Solihin, F., & Hastarita, F. (2014). Perancangan Dan Pembuatan Aplikasi Pencarian Informasi Beasiswa Dengan Menggunakan Cosine Similarity. *Jurnal SimanteC*, 115-124.
- Kurniawan, B., Effendi, S., & Sitompul, S. (2012). Klasifikasi Konten Berita Dengan Metode Text Mining. *Jurnal Dunia Teknologi Informasi*, 14-19.
- Kustiyahningsih, Y., & Anamisa, D. R. (2011). *Pemrograman Basis Data Berbasis Web Menggunakan PHP dan MySQL*. Yogyakarta: Graha Ilmu.
- Lavarino, D., & Yustanti, W. (2016). Rancang Bangun E-Voting Berbasis Website di Universitas Negeri Surabaya. *Jurnal Manajemen Informatika*, 72-81.
- LPPM. (2018). *Laporan Akhir Tahun*. Bandung: LPPM UPI.
- Lusiana, V. (2014). Teknik Kompresi Citra Digital untuk Penyimpanan File Menggunakan Format Data XML. *Jurnal Teknologi Informasi DINAMIK*, 112-119.
- Mariel, W. C., Mariyah, S., & Pramana, S. (2018). Sentiment analysis: a comparison of deep learning neural network algorithm with SVM and naïve Bayes for Indonesian Text. *Journal of Physics*, 1-8.

- Menristekdikti. (2015). *Peraturan Menteri Riset, Teknologi, Dan Pendidikan Tinggi Republik Indonesia Nomor 44 Tahun 2015*. Jakarta: Menteri Riset, Teknologi, Dan Pendidikan Tinggi Republik Indonesia.
- Muttaqien, M. Z., & Kusmayadi, E. (2009). *Dasar-dasar teknologi informasi*. Jakarta: Universitas Terbuka.
- Noor, I. H. (2010). Penelitian dan Pengabdian Masyarakat pada Perguruan Tinggi. *Jurnal Pendidikan dan Kebudayaan*, 285-297.
- Nurdiana, O., Jumadi, & Nursantika, D. (2016). Perbandingan Metode Cosine Similarity Dengan Metode Jaccard Similarity Pada Aplikasi Pencarian Terjemah Al-Qur'an Dalam Bahasa Indonesia. *Jurnal Online Informatika*, 59-63.
- Nurmaya, S. D. (2012). *Eksplorasi Framework Yii Sebagai Pendukung Pembuatan Perangkat Lunak Berbasis Web*. Bandung: Universitas Pasundan.
- Palit, R. V., Rindengan, Y. D., & Lumenta, A. S. (2015). Rancangan Sistem Informasi Keuangan Gereja Berbasis Web Di Jemaat GMIM Bukit Moria Malalayang. *E-Journal Teknik Elektro dan Komputer*.
- Polettini, N. (2004). The Vector Space Model in Information Retrieval-Term Weighting Problem. *Department of Information and Communication Technology*.
- Prasetyo, B., Pattiasina, T. J., & Soetarmono, A. N. (2015). Perancangan dan Pembuatan Sistem Informasi Gudang (Studi Kasus : PT. PLN (Persero) Area Surabaya Barat). *Teknika*, 12-16.
- Pratama, F. (2014). *Sistem Pendukung Keputusan Menentukan Siswa Berprestasi Pada Sma Institut Indonesia Semarang Menggunakan Metode Profile Matching*. Semarang.
- Pressman, R. S. (2005). *Software engineering: a practitioner's approach*. New York: Palgrave Macmillan.
- Ricci, F., Rokach, L., & Shapira, B. (2010). *Recommender Systems Handbook*.
- Sidik, B. (2011). *PHP Framework Codeigniter*. Bandung: Penerbit Informatika.
- Siregar, R. R., Sinaga, F. A., & Arianto, R. (2017). Aplikasi Penentuan Dosen Penguji Skripsi Menggunakan Metode Tf-Idf Dan Vector Space Model. *Journal of Computer Science and Information Systems*, 171-186.
- Solichin, A. (2016). *Pemrograman Web dengan PHP dan MySql*. Jakarta: Penerbit Budi Luhur.
- Steinberger, R. (2011). A survey of methods to ease the development of highly multilingual text mining applications. *Language Resources and Evaluation*, 155-176.
- Sugiyono. (2012). *Metode Penelitian Pendidikan Pendekatan Kuantitatif, Kualitatif dan R & D*. Bandung: Alfabeta.
- Sulistyo, M. E., Saptono, R., & Asshidiq, A. (2015). Penilaian Ujian Bertipe Essay Menggunakan Metode Text Similarity. *Telematika*, 146-158.
- Supranto, J. (1992). *Statistika Dan Sistem Informasi Untuk Pimpinan*. Jakarta: Erlangga.

- Susanti, Y., Siswanto, & Yupianti. (2015). Sistem Pelayanan On-Line Pada Asosiasi INKINDO Bengkulu. *Jurnal Media Infotama*, 191-200.
- Triwijoyo, B. K., & Kartarina. (2019). Analysis of Document Clustering based on Cosine Similarity and K-Main Algorithms. *Journal of Information Systems and Informatics*, 164-177.
- Turban, E. (1995). *Decision Support and Expert System : Management Support System. Fourth Edition*. New Jersey: Prentice Hall International Inc.
- Turban, E., Aronson, J. E., & Liang, T. P. (2005). *Decision Support System and Intelligent Systems 7th. Edition*. Yogyakarta: Penerbit Andi.
- Vijayarani, D. S., Ilamathi, J., & Nithya. (2014). Preprocessing Techniques for Text Mining - An Overview. *International Journal of Computer Science & Communication Networks*, 7-16.
- Wali, M., & Ahmad, L. (2017). Perancangan Aplikasi Source code library Sebagai Solusi Pembelajaran Pengembangan Perangkat Lunak. *Journal JTJK (Jurnal Teknologi Informasi dan Komunikasi)*, 39-47.
- Widjaja, M., & Hansun, S. (2015). Implementation of Modified Porter Stemming Algorithm to Indonesian Word Error Detection Plugin Application. *International Journal Of Technology*, 139-150.
- Yasni, L., Subroto, I. M., & Haviana, S. C. (2018). IMPLEMENTASI COSINE SIMILARITY MATCHING DALAM PENENTUAN DOSEN PEMBIMBING TUGAS AKHIR. *Transmisi*, 22-28.
- Yuan, M., Ouyang, Y., & Xiong, Z. (2013). A Text Categorization Method using Extended Vector Space Model by Frequent Term Sets. *Journal of Information Science and Engineering*, 99-114.
- Yuana, R. A., & Harjunowibowo, D. (2018). Data Similarity Filtering of Wartegg Personality Test Result using Cosine-Similarity. *International Journal of Recent Contributions from Engineering, Science & IT*, 19-28.
- Zaware, P. N., Gautam, A., Nashte, S., & Khanuja, P. (2015). AN EFFECTUAL APPROACH FOR CALCULATING COSINE SIMILARITY. *International Journal of Advance Engineering and Research*, 13-18.