

Aria Hendrawan
Universitas Semarang, Indonesia
Email: ariahendrawan@usm.ac.id

Pulung Nurtanti
Universitas Dian
Email: pulung@r



METODE STEMMING DALAM PENERAPAN ARTIKEL BERBAHASA INDONESIA

Aria Hendrawan, M.Kom
Dr. Titin Winarti, S.Kom., M.M
Henny Indriyawati., M.Kom

Metode Stemming dalam Penerapan Artikel Berbahasa Indonesia

Penulis

Aria Hendrawan, M.Kom
Dr. Titin Winarti, S.Kom., M.M
Henny Indriyawati., M.Kom

ISBN 978-623-6121-85-6

Cetakan Pertama, Juni 2021

v, 69 hlm; 15.5 x 23

Penyunting

Umi Salamah, Misbahul Munir

Desain Sampul

Husni

Desain Layout

Ajizah

Penerbit :

CV. Pustaka Learning Center

Karya Kartika Graha A.9 Malang 65132

Whatsapp 08994458885

www.pustakalearningcenter.com

Hak Cipta Dilindungi Undang-Undang . Dilarang memperbanyak atau memindahkan Sebagian atau seluruh isi buku ini ke dalam bentuk apapun secara elektronik maupun mekanis tanpa izin Tertulis dari penulis dan Penerbit Pustaka Learning Center

KATA PENGANTAR

Puji syukur kami panjatkan kepada Allah SWT atas segala karunia nikmat serta hidayahnya sehingga kami dapat menyusun laporan penelitian yang berjudul “Pengembangan Stemming untuk Artikel Berbahasa Indonesia” dengan lancar dan tepat waktu.

Tujuan dari penyusunan laporan penelitian ini adalah tingginya dimensi ruang vektor akibat banyaknya kata yang digunakan dan adanya data yang tidak relevan serta redundan. Upaya mengurangi tingginya dimensi ruang vektor dapat ditempuh dengan melakukan tahap *Stemming*. Selesaiannya penyusunan laporan penelitian ini tidak lepas dari bantuan, support, arahan dan bimbingan banyak pihak. Oleh sebab itu peneliti ingin menyampaikan terima kasih kepada pihak-pihak yang telah membantu penelitian ini.

Meski demikian, peneliti merasa masih banyak kesalahan dalam penyusunan laporan penelitian ini. Oleh sebab itu peneliti sangat terbuka menerima kritik dan saran yang membangun untuk dijadikan sebagai bahan evaluasi.

Akhir kata, semoga laporan penelitian ini dapat diterima sebagai gagasan anak bangsa yang layak didukung untuk menjadi solusi atas permasalahan di kehidupan sehari-hari.

DAFTAR ISI

KATA PENGANTAR ~ iii

DAFTAR ISI ~ v

BAGIAN 1. PENDAHULUAN ~ 1

BAGIAN 2. KERANGKA TEORITIS ~ 9

BAGIAN 3. METODE PENELITIAN ~ 19

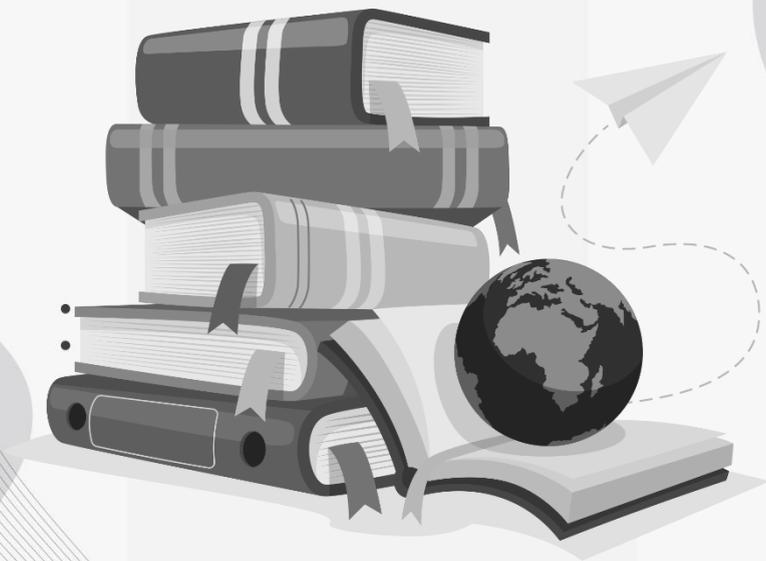
BAGIAN 4. HASIL DAN PEMBAHASAN ~ 27

BAGIAN 5. KESIMPULAN ~ 41

DAFTAR PUSTAKA ~ 45

BIOGRAFI PENULIS ~ 49

Pendahuluan



BAGIAN 1

PENDAHULUAN

Bahasa Indonesia adalah bahasa aglutinatif yang memungkinkan kata-kata baru akan dibentuk dengan menambahkan prefix dan sufiks untuk kata. Kata-kata baru juga dapat dibentuk dengan mengulang kata dan dengan memasukkan infiks menjadi sebuah kata. Paice menyatakan bahwa kata-kata biasanya Stemming karena bentuk yang berbeda sintaksis diasumsikan memiliki arti yang sama. Stemming membutuhkan pemahaman yang baik tentang bahasa yang bersangkutan. Seperti bahasa Inggris memiliki awalan "hiper-" dalam "hipertensi" dan "hiperaktif", "anti-" dalam "antisosial", dan "ultra-" dalam "ultraviolet". Awalan ini menciptakan makna baru yang berbeda dari makna aslinya, karena itu mereka tidak mempertimbangkan dalam Information Retrieval (IR). Dalam bahasa Inggris Stemming biasanya hanya menghapus akhiran. Implementasi pada penelitian ini didasarkan pada



Kerangka Teoritis

BAGIAN 2

KERANGKA TEORI

TEXT PROCESSING

Berdasarkan ketidak teraturan struktur data teks, maka proses sistem temu kembali informasi ataupun text mining memerlukan beberapa tahap awal yang pada intinya adalah mempersiapkan agar teks dapat diubah menjadi lebih terstruktur. Salah satu implementasi dari text mining adalah tahap Text Preprocessing. Tahap Text Preprocessing adalah tahapan dimana aplikasi melakukan seleksi data yang akan diproses pada setiap dokumen. Proses preprocessing ini meliputi (1) case folding, (2) tokenizing, (3) filtering, dan (4) stemming.

Metode



BAGIAN 3

METODE PENELITIAN

Dengan penerapan (algoritma) metode stemming maka proses pengambilan dan pengupasan kata imbuhan dokumen akan menjadi lebih cepat dan akurat dalam melakukan proses yang dilakukan. Proses Stemming merupakan pengelolaan keyword menjadi keyword yang utuh yaitu dengan menghilangkan imbuhan seperti diantaranya “yang”, “di”, “ke”, “me”, “meng”, “kan”. Penguraian dari suatu kata menjadi bentuk kata dasarnya (stem). Untuk lebih jelasnya tahapan proses Stemming ditunjukkan sebagai berikut: Token hasil tokenizing diperiksa apakah mengandung imbuhan atau tidak. Jika terdapat imbuhan maka akan dilakukan pembuangan imbuhan, terus berulang sampai tidak mengandung imbuhan. Jika tidak mengandung imbuhan maka akan ditampilkan.

Hasil & Pembahasan



C. Proses 3 (Stopword)

Tahapan ini mengambil kata-kata penting dari hasil token. Bisa menggunakan algoritma stop list (membuang kata yang kurang penting) atau word list (menyimpan kata penting). Sistem ini menggunakan metode stop list yaitu penghilangan kata tidak penting (stopword) pada deskripsi melalui pengecekan kata-kata hasil token deskripsi apakah termasuk didalam daftar kata tidak penting (stop list) atau tidak. Jika termasuk di dalam stoplist maka kata-kata tersebut akan di-remove dari deskripsi sehingga kata-kata yang tersisa di dalam deskripsi dianggap sebagai kata-kata penting atau keywords (pattern).

D. Proses 4 (Stemming Menggunakan

Pendekatan Porter, Sastrawi, dan Snowball)

Pada proses ini penelitian dilakukan dengan melakukan 3 algoritma stemming, yakni Porter, Sastrawi, dan Snowball. Dari hasil yang didapatkan pada proses stemming tersebut, selanjutnya akan mencari nilai akurasi hasil ketiga algoritma itu. Sehingga bisa didapatkan manakah algoritma yang memiliki akurasi terbaik

Kesimpulan



BAGIAN 5

KESIMPULAN

A. Kesimpulan

Stemming pada kalimat berbahasa Indonesia yang digunakan untuk meningkatkan performa Information Retrieval dengan cara mentransformasi kata-kata dalam sebuah dokumen teks ke kata dasarnya, pada penelitian ini mendapatkan hasil terbaik dari penggunaan algoritma sastrawi. Hal ini ditunjukkan dengan tingkat akurasi yang sebesar 98% dibandingkan algoritma porter yang memiliki 72% akurasi dan algoritma snowball dengan 88% akurasi.

B. Saran

Untuk tahapan berikutnya diperlukan proses klusterisasi pada dokumen-dokumen yang sudah distemming untuk melihat hasil dari proses stemming yang pernah dilakukan tersebut, apakah berpengaruh secara baik atau tidak.

DAFTAR PUSTAKA



DAFTAR PUSTAKA

- Hapsari, R. K., & Santoso, Y. J. (2015). Stemming Artikel Berbahasa Indonesia Dengan Pendekatan Confix-Stripping. In Prosiding Seminar Nasional Manajemen Teknologi XXII, Vols. ISBN (pp. 978-602).
- Asian J. (2007). Effective Techniques for Indonesian Text Retrieval. PhD Thesis
School of Computer Science and Information Technology RMIT University Australia.
- B. A. A. Nazief and M. Adriani, (1996). Confix-stripping: Approach to stemming algorithm for Bahasa Indonesia. Internal publication, Faculty of Computer Science, University of Indonesia, Depok, Jakarta.
- Elizabet N. S. C. P. (2013). Rancangan Bangunan Aplikasi ChatBot Informasi Objek Wisata Kota Bandung dengan pendekatan Natural Language Processing. Universitas Komputer Indonesia, Bandung.
- Ghazvini, A. and Ab Aziz, M.J., 2012, September. Stemming algorithm for different tenses to improve Persian dictionary. In Industrial

Electronics and Applications (ISIEA), 2012
IEEE Symposium on (pp. 50-53). IEEE.

I Putu Adhi Kerta Mahendra, Agus Zainal Arifin,
Henning Titi Ciptaningtyas. (2008).
Penggunaan Algoritma Semut dan Confix-
Stripping Stemmer untuk Klasifikasi Dokumen
teks Berbahasa Indonesia.

Kamus Besar Bahasa Indonesia, Edisi , 2005, Ketiga.
Jakarta: Balai Pustaka. Muhammed Yavuz
Nuzumlalı Arzucan. 2014. Analyzing
Stemming Approaches for Turkish

Navanath Saharia, U. Sharma, and J. Kalita. A suffix-
based noun and verb classifier for an
inflectional language. In International
Conference on Asian Language Processing,
pages 19–22. IEEE Computer Society, 2010.

BIOGRAFI PENULIS



BIOGRAFI PENULIS 1

1. Nama Lengkap	Aria Hendrawan, M.Kom.
2. Jenis Kelamin	Laki-Laki
3. Jabatan Fungsional	Asisten Ahli
4. NIS	06557003102159
5. NIDN	0621038501
6. Tempat dan	Semarang, 21 Maret 1985
7. E-mail	ariahendrawan@usm.ac.id
8. Nomor HP	08112790713
9. Alamat Kantor	Jl. Soekarno Hatta –
10. No Telepon	(024) 6702757 / (024)
11. Lulusan yang telah	S-1 = 10 Orang
12. Mata Kuliah yang diampu	<ol style="list-style-type: none"> 1. Pengolahan Citra 1. Kecerdasan Buatan 2. Open Source Systems 3. Grafika Komputer 4. Jaringan Syaraf Tiruan

❖ Riwayat Pendidikan

Uraian	S-1	S-2
Nama Perguruan Tinggi	Universitas Muhammadiyah Yogyakarta	Universitas Dian Nuswantoro
Bidang Ilmu	Teknik Elektro (TE)	Teknik Informatika (TI)
Tahun	2004	2012
Tahun Lulus	2009	2014

Judul	Analisis Pengaruh	Analisa
Skripsi/Thesis/Disertasi	Diameter Dan Bahan Terhadap Penerimaan Sinyal Antena Wajan Bolic	Peningkatan Kualitas Citra Bawah Air Berbasis Koreksi Gamma Untuk Pencocokan Gambar Pada Algoritma SIFT

Nama Pembimbing/Promotor	M. Helman, M.T. Ir. Tony K Hariadi, M.T.	Dr. Pulung Nurtantio A, S.T., M.Kom. Ricardus A P,
--------------------------	---	---

❖ **Pengalaman Penelitian dalam 5 Tahun Terakhir (bukan Skripsi, Tesis, maupun disertasi)**

No	Tahun	Judul Penelitian	Pendanaan Sumber	Jumlah
1.	2015	Raspberry Pi Dengan Modul Kamera Dan Motion Detection	LPPM	5.000.000

❖ **Pengalaman Pengabdian kepada Masyarakat dalam 5 Tahun Terakhir**

No	Tahun	Judul Pengabdian	Pendanaan Sumber	Jumlah
----	-------	------------------	------------------	--------

1.	2014	Peningkatan Kemampuan Animasi 3D Bagi Siswa SMK Walisongo	Mandiri	1.000.000
----	------	---	---------	-----------

❖ Pengalaman Penulisan Artikel Ilmiah dalam Jurnal dalam 5 Tahun Terakhir

No.	Tahun	Judul Artikel	Nama	Vol/ No/ Th
1.	2014	Analisa Peningkatan Kualitas Citra Bawah Air Berbasis Koreksi Gamma untuk Pencocokan Gambar pada	Jurnal Transformatika	ISSN : 1693-3656, Vol. 12 No. 1, Edisi Juli 2014, Hal. 27-33.

BIOGRAFI PENULIS 2

Nama Lengkap	Dr. Titin Winarti, S. Kom., M.M.
Jabatan Fungsional	Lektor Kepala/ IV C
Jabatan Struktural	Ka. Lab. Oracle
NIS/NIDN	06557003102049/0627116901
Tempat, Tanggal Lahir	Jakarta, 27 November 1969
Perguruan Tinggi	Universitas Semarang
Alamat	Jl. Soekarno Hatta Tlogosari
Telp./Faks	024-6702757 / 024-6702272
Alamat Rumah	Taman Suryo Kusumo IV No. 58
Telp./Faks	085640328082
Alamat e-mail	titin@usm.ac.id
Mata Kuliah yang Diampu	1. Basis Data 2. Algoritma dan Pemrograman

3. Struktur Data
4. Algoritma dan Struktur Data
5. Oracle WDP

❖ **Riwayat Pendidikan Perguruan Tinggi**

	S1	S2	S3
Nama PT	Univ Gunadarma	Univ Gunadarma	Univ Gunadarma
Bidang Ilmu	Sistem Informasi	Sistem Informasi Bisnis	Teknologi Informasi
Tahun Masuk- Lulus	1988	1994	2017

❖ **Pengalaman Penelitian dalam 5 Tahun Terakhir**

Tahun	Judul Penelitian/ Pengabdian Kepada Masyarakat	Tim	Dana
2014	Penerapan Sistem Informasi Layanan	Titin Winarti	USM

Perpustakaan
Berbasis Teknologi
Barcode sebagai
Pengembangan
Sistem Informasi
Perpustakaan di
Fakultas Teknologi
Informasi dan
Komunikasi
Universitas Semarang.
(Penelitian)

2015	Sistem Pendukung Keputusan Pemilihan Kelurahan Percontohan Dengan Metode Profile Matching Pada Kecamatan Mijen Kota Semarang.	Titin Winarti	USM
2015	Aplikasi Pengukuran Kemiripan Dokumen Menggunakan Cosine Similarity	Titin Winarti	USM

2018	Deteksi Kemiripan Dokumen Bahasa Indonesia Dengan Menggunakan Model Ruang Vektor	Titin Winarti	USM
2019	Pemilihan Fitur Bahasa Indonesia Untuk Pengelompokan Dokumen Dengan Metode K-means	Titin Winarti	USM

❖ **Pengalaman Penulisan Jurnal dan Prosiding dalam 5 Tahun Terakhir**

Tahun	Judul Penelitian	Nama Jurnal/ Prociding	Penerbit/ Penyelenggra
2015	Fuzzy Logic and Boston Consulting Group To Evaluate Products	ICCSE (International Conference Computer Science, Electronics and Instrumentation)	Universitas Gajah Mada

2015	Model of Document Similarity Measurement System by Fingerprint Method Utilization with Natural Language Processing Approach Page 205	The 5 th Uzbekistan – Indonesia International Joint Conference on Globalization, Economic Development, and Nation Character Building	Universitas Gunadarma
2016	Plagiarism Detection System For Indonesia Text Based Document by Fingerprint	The 4th International Conference on Internet Services Technology and Information Engineering 2016	Lembaga Telematika Indonesia
2017	Method and Natural	International Journal on	International Journal on

Language	Advanced	Advanced
Processing	Science on	Science on
Approach	Engineering	Engineering
Improving	Information	Information
Stemming	Technology	Technology
Algorithm		
Using		
Morphological		
Rules		

❖ **Pengalaman Pengabdian Kepada Masyarakat dalam 5 Tahun Terakhir**

Jenis	Tahun	Nama Kegiatan	Tempat
Pelatihan	2014	Pelatihan Mail Merge bagi Siswa SMU/ SMK se-Kota Semarang dan sekitarnya.	FTIK USM
Pelatihan	2015	Pelatihan Pembuatan Blog bagi Siswa SMU/SMK/Sederajat.	FTIK USM

Pelatihan	2016	Pengembangan Bahan Ajar Menggunakan Media Pembelajaran Interaktif (Mpi) Pada SMK NU 03 Kaliwungu Kabupaten Kendal	SMK NU 03 Kaliwungu Kabupaten Kendal
Pelatihan	2017	Peningkatan Kemampuan E-Commerce Sebagai Sarana Publikasi Dan Media Promosi Produk Kreatifitas Siswa Dan Guru Smk Grafika Bhakti Nusantara Semarang	FTIK USM
Pelatihan	2018	Peningkatan Pemahaman Dan Pemanfaatan Aplikasi Enterprise Resource Planning (Erp) Sebagai	FTIK USM

Penunjang
Kegiatan
Kewirausahaan
Siswa Pada Smk
Pelita Nusantara
Semarang

Pelatihan	2018	Sosialisasi Media Pembelajaran Bahasa Inggris Berbasis Macromedia Flash Untuk Guru Paud Di Kota Semarang	FTIK USM
Pelatihan	2019	Pemanfaatan Microsoft Excel Untuk Pembuatan Rapot Bagi Guru Paud Di Kota Semarang	FTIK USM

❖ Pengalaman Pembicara/Instruktur

Jenis	Tahun	Nama Kegiatan	Sebagai
Pelatihan	2014	Pelatihan Mail Merge bagi Siswa SMU/ SMK se-Kota Semarang dan sekitarnya.	Instruktur

Conference	2015	ICCSE (International Conference Computer Science, Electronics and Instrumentation)	Pemakalah
Conference	2015	The 5 th Uzbekistan – Indonesia International Joint Conference on Globalization, Economic Development, and Nation Character Building)	Pemakalah
Pelatihan	2015	Pelatihan Pembuatan Blog bagi Siswa	Instruktur

		SMU/SMK/Sede rajat.	
Conference	2016	The 4th International Conference on Internet Services Technology and Information Engineering 2016	Pemakalah
Pelatihan	2016	Pengembangan Bahan Ajar Menggunakan Media Pembelajaran Interaktif (Mpi) Pada SMK NU 03 Kaliwungu Kabupaten Kendal	Instruktur
Conference	2019	Feature Selection For Optimization	Pemakalah

Naive
Pemakalah
Bayes Algorithm

Pelatihan	2017	Peningkatan Kemampuan E- Commerce Sebagai Sarana Publikasi Dan Media Promosi Produk Kreatifitas Siswa Dan Guru Smk Grafika Bhakti Nusantara Semarang	Instruktur
-----------	------	--	------------

Pelatihan	2018	Peningkatan Pemahaman Dan Pemanfaatan Aplikasi Enterprise Resource Planning (Erp)	Instruktur
-----------	------	--	------------

Sebagai
Penunjang
Kegiatan
Kewirausahaan
Siswa Pada
Smk Pelita
Nusantara
Semarang

Pelatihan	2018	Sosialisasi Media Pembelajaran Bahasa Berbasis Macromedia Flash Untuk Guru Paud Di Kota Semarang Inggris	Instruktur
-----------	------	---	------------

Pelatihan	2019	Pemanfaatan Microsoft Excel Untuk Pembuatan Rapot Bagi	Instruktur
-----------	------	--	------------

Guru Paud Di
Kota Semarang

BIOGRAFI PENULIS 3

Nama Lengkap (dengan gelar)	Henny Indriyawati, M. Kom.
Jenis Kelamin	Perempuan
Jabatan Fungsional	Pengajar
NIP/NIK/Identitas lainnya	06557003102130
NIDN	0603068301
Tempat, Tanggal Lahir	Semarang , 3 Juni 1983
E-mail	henny@usm.ac.id
Nomor Telepon/HP	(024) 6709157/ 085740003306
Alamat Kantor	Jl. Soekarno Hatta Semarang

	Pemberian Kredit Komersial pada BRI Cabang Semarang Pattimura.	menggunakan Metode - Forward Chaining Berbasis SMS Gateway.
Nama Pembimbing/ Promotor	1. Dwi Agus Diartono M.Kom 2. Rina Candra, M.Kom	1. Bayu Suroso, P.hD 2. Aris Sugiharno, M.Kom

❖ **Pengalaman Penelitian Dalam 5 Tahun Terakhir**

No	Tahun	Judul Penelitian	Pendanaan
1	2013	Aplikasi Sistem Penilaian Dosen Teladan dengan Metode Fuzzy Multi Attribute Decition Making (MADM) pada Universitas Semarang	Universitas Semarang
2	2015	Fuzzy Multi-Atribute Decisionmaking	Dikti

		(Fuzzy Madm) Dengan Metode Saw Untuk Pemilihan Mahasiswa Berprestasi	
3	2016	Berprestasi Sistem Perkuliahan Online Berbasise-Learning Pada Fakultas Teknologi Informasi Dan Komunikasi Universitas Semarang	Universitas Semarang
4	2016	Mobile Information System Untuk Mendukung Pengelolaan Kegiatan Pelaksanaan Pendidikan Bagi Dosen	Dikti
5	2018	Aplikasi Sistem Inventaris Laboratorium Komputer	Universitas Semarang
6	2018	Penentuan Prediksi Stok Mobil Dengan	Universitas Semarang

Pendekatan Kepuasan
Kostomer
Menggunakan Metode
Multi- Objective
Optimization On The
Basis Of Ratio
Analysis (Moora) Di
PT. New Ratna Motor
Semarang

7	2019	Sistem Pendukung Keputusan Penentuan Lokasi Industri Berbasis Spasial Menggunakan Metode MOORA	Universitas Semarang
---	------	---	-------------------------