

PROPOSAL SISTEM TEMU KEMBALI INFORMASI
APLIKASI PEMENGALAN KATA BERIMBUHAN DALAM BAHASA
INDONESIA DENGAN METODE ALGORITMA STEMMING



Disusun Oleh:

Ardyan Wahyu Anggoro

(A11.2022.14729)

JURUSAN TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS DIAN NUSWANTORO

Ringkasan

Bahasa Indonesia kaya akan imbuhan. Kurang lebih ada sekitar 35 imbuhan resmi yang disebutkan dalam kamus Besar Bahasa Indonesia. Imbuhan-imbuhan ini dapat berupa prefix (awalan), sufiks (akhiran), konfiks maupun infiks (sisipan) yang diserap dari Bahasa Jawa. Algoritma stemming yaitu suatu teknik pencarian bentuk dasar dari suatu term. Yang dimaksud dengan term itu sendiri adalah tiap kata yang berada pada suatu dokumen teks.

Di dalam algoritma stemming terdapat algoritma Nazief dan Adriani yaitu salah satu algoritma yang digunakan dalam tahap stemming. Algoritma Nazief et Adriani merupakan algoritma untuk mengubah kata yang mempunyai sufiks, prefiks, dan imbuhan menjadi bentuk kata dasar. Algoritma Nazief dan Adriani digunakan dalam penelitian ini sebagai algoritma pendukung dalam proses identifikasi kata dasar dari kata majemuk dalam sebuah kalimat. Penentuan kata dasar kata majemuk ini dengan menghilangkan seluruh imbuhan termasuk awalan, sisipan, akhiran dan konfiks pada kata majemuk tersebut menggunakan algoritma radikal Nazief dan Adriani. Seperti diketahui kelas kata dasar mempunyai akhiran.

Kata Kunci : Bahasa Indonesia, Algoritma Stemming, Nazief dan Adriani

1. PENDAHULUAN

1.1 Latar Belakang

Bahasa Indonesia kaya akan imbuhan. Ada sekitar 35 imbuhan resmi yang disebutkan dalam Kamus Besar Bahasa Indonesia. Imbuhan tersebut dapat berupa awalan (prefixes), sufiks (akhiran), imbuhan atau sisipan (insertions) yang diserap dari bahasa Jawa. Satu hal yang unik di Indonesia adalah kecenderungan penggunaan imbuhan secara bebas. Tentunya dapat dikatakan bahwa afiks dalam bahasa Indonesia dapat digunakan pada semua kata dan afiks tersebut dapat digabungkan satu sama lain.

Dalam dunia pemrograman, terdapat algoritma stemming yang dapat digunakan untuk menentukan basis suatu imbuhan. Adapun pengertian dari *Algoritma Stemming* yaitu suatu teknik pencarian bentuk dasar dari suatu *term*. Yang dimaksud dengan *term* itu sendiri adalah tiap kata yang berada pada suatu dokumen teks. Stemming dilakukan pada saat pembuatan indeks dari suatu dokumen. Pembuatan indeks dilakukan karena suatu dokumen tidak dapat dikenali langsung oleh suatu sistem. Dokumen tersebut terlebih dahulu perlu dipetakan kedalam suatu representasi dengan menggunakan teks yang berada di dalamnya.

Pada algoritma derivasi terdapat algoritma Nazief dan Adriani yang merupakan salah satu algoritma yang digunakan pada tahap derivasi. Algoritma Nazief dan Adriani merupakan algoritma untuk mengubah kata yang memiliki sufiks, prefiks, dan konjungsi menjadi bentuk kata dasar. Algoritma Nazief dan Adriani digunakan dalam penelitian ini sebagai algoritma bantuan dalam proses identifikasi kata dasar dari penugasan kata dalam kalimat. Teknik stemming diperlukan tidak hanya untuk mengurangi jumlah indeks yang berbeda dalam suatu dokumen, tetapi juga untuk mengelompokkan kata lain yang mempunyai akar dan makna yang sama, namun mempunyai bentuk yang berbeda karena mempunyai imbuhan yang berbeda.

1.2 Rumusan Masalah

Adapun rumusan yang di uraikan pada penelitian ini yaitu:

1. Bagaimana proses pembentukan kata dasar menjadi kata berimbuhan pada kalimat Bahasa Indonesia
2. Bagaimana menerapkan algoritma stemming dalam penentuan kata dasar dan kelas kata dari kata berimbuhan pada kalimat Bahasa Indonesia?
3. Bagaimana merancang aplikasi penentuan kata dasar dari kata berimbuhan pada kalimat Bahasa Indonesia dengan algoritma stemming?

1.3 State Of The Art

NO	Nama dan Tahun Penelitian	Metode Penelitian	Hasil Penelitian
1.	ANALISIS PEMAKAIAN BAHASA INDONESIA PADA LAPORAN PENELITIAN DOSEN DI	Metode Kualitatif	Hasil penelitian menunjukkan bahwa dalam laporan penelitian dosen Universitas Pendidikan Ganesha tahun 2012 masih ditemukan aneka ragam penyimpangan pemakaian bahasa, baik yang menyangkut tata kalimat, tata bentukan, tata makna/diksi/ logika, maupun tata tulis. Secara kuantitatif,

	<p>LINGKUNGAN UNIVERSITAS PENDIDIKAN GANESHA (2 Oktober 2014)</p>		<p>jumlah penyimpangan tertinggi terdapat pada ragam tata tulis atau ejaan, yaitu sebanyak 591 buah (74.90%), sedangkan jumlah penyimpangan terendah ditemukan pada ragam penyimpangan tata bentukan, yaitu sebanyak 17 buah (2.15%), Sementara itu, ragam penyimpangan tata kalimat, tata makna/diksi/logika, dan campuran masing-masing ditemukan 38 buah (4.82%), 31 buah (3.93%), dan 112 buah (14.20%). Secara kualitatif, dalam kaitannya dengan tata kalimat, ditemukan adanya penyimpangan kalimat tanpa subjek dan atau predikat, kalimat tanpa paralelisme, kalimat dengan frasa berpola Menerangkan Diterangkan, dan kalimat pleonastis</p>
2.	<p>ANALISIS KESALAHAN KATA BERIMBUHAN DALAM TEKS NEGOSIASI SISWA KELAS X SMA NEGERI 1 KOTA TANGERANG SELATAN SEMESTER GENAP TAHUN PELAJARAN 2015/2016</p>	<p>Metode kualitatif</p>	<p>SMA Negeri 1 Kota Tangerang Selatan merupakan sekolah yang berdiri pada tanggal 4 Januari 1977. Sekolah ini dulunya merupakan SMA N 27 Jakarta kelas jauh Ciputat. Sekolah ini memiliki visi yaitu unggul dalam IPTEKS, bermartabat, berkarakter, dan berbudaya lingkungan berlandaskan IMTAQ, sedangkan misi sekolah sebagai berikut:</p> <ul style="list-style-type: none"> - Menumbuhkan penghayatan terhadap ajaran agama yang dianut dan budaya bangsa sehingga menjadi sumber kearifan dalam bertindak.

			- Meningkatkan pengetahuan dan membina perilaku peserta didik dalam meningkatkan martabat pribadi dan lembaga
--	--	--	---

2. DATASET

Data set yang saya dapat ialah bersifat publik dan saya dapatkan dari kaggle.com

kaggle datasets download -d jackdaoud/marketing-data

3. METODE

3.1 Analisa Kata Dasar

Dalam aplikasi ini dibuat kata dasar sebagai database kata dasar yang merupakan pusat dari daftar kata dasar yang selanjutnya akan dijadikan referensi untuk mengidentifikasi kata dasar dari kata akibat radikalisi kemajuan. Hasil penentuan kata dasar ini tergantung apakah kata tersebut ada di database kata dasar atau tidak. Dengan demikian, batasan pada bagian kata dasar adalah seluruh kata yang melalui proses pencarian akar tidak berhasil dicari kata dasarnya. Keberhasilan identifikasi kata dasar dipengaruhi oleh daftar kata dasar yang ada pada database kata dasar beserta proses derivasinya sendiri.

Basis data kata dasar disini berperan sebagai daftar dari kumpulan kata dasar yang akan digunakan pada saat pencarian kata dasar oleh menggunakan algoritma stemming. Dalam database, kata dasar ini mencakup sejumlah bagian fungsional yang membantu dalam mengidentifikasi bentuk kata. Misalnya, kata makan memiliki kata dasar makan.

3.2 Penerapan Algoritma Stemming

Algoritma ini menggunakan algoritma asli yang telah dibahas pada bagian sebelumnya dari algoritma Nazief dan Adriani. Selain itu, ada langkah tambahan dalam proses pencarian stemming kata untuk mendapatkan kata dasar dari sinonimnya. melibatkan pembagian kata-kata dengan imbuhan yang dikelompokkan menjadi empat bagian tergantung pada imbuhan yang dimiliki imbuhan tersebut. Adapun rumusan dari algoritma stemming ini:

DP + DP + ROOT WORD + DS + PP

Keterangan :

DP: derivational prefix

DS: derivational suffix

PP: possessive pronoun Mem-pe-kerja-kan-nya

Keterangan:

1. Derivation prefiks I = Mem
2. Derivation prefiks II = pe
3. Root word = kerja
4. Derivation sufiks = kan
5. Possessive pronoun = nya

Adapun proses dari tahap pemotongan kata pada kata berimbuhan diatas dapat dilihat dari penjelasan pada table dibawah ini :

Algoritma penentuan kata dasar dari kata berimbuhan

DP + DP + ROOT WORD + DS + PP

Input : I : Input teks.

Output :k : Kata

Rw : Root word (Kata dasar)

kk : kelas kata

Dp :Derivation Prefiks (imbuhan awal)

Ds : Derivation Surfiks (imbuhan akhir)

Dp = Dp + Root Word

Ds = Root Word + Ds

Proses : Do while i

If i= Root word then

Rw=i

Elseif i <> Root word then

If i termasuk kelas kata 1 then

If i terdapat Dp then

i= hapus Dp

Elseif

i = hapus Ds

End if

If i termasuk Rw then

Rw = i

End if

Elseif i termasuk kelas kata 2 then

If i terdapat Dp then

Rw = sisa pemotongan

```

Dp = hasil pemotongan
Elseif i terdapat Ds then
Rw = sisa pemotongan
Ds = hasil pemotongan
    End if
Elseif i termasuk kelas kata 3 then
Rw =sisa pemotongan
Dp= hasil pemotongan Dp

Ds = hasil pemotongan Ds
End if
If Rw=Root word then
    Rw=Rw
End if
k = i
End while

```

4. DESAIN

1. Halaman Form Masuk Halaman form masuk ini adalah halaman yg akan tampil pertama kali apabila aplikasi dijalankan.

form Stemming	
Masukkan kata berimbuhan :	
Bekerjalah	
Proses	Hapus
keluar	
Hasil Stemming :	
Kata	= bekerjalah
kata dasar	= kerja
Kelas kata	= Kerja
Imbuhan awal	= be
Imbuhan akhir	= lah

2. Form Aplikasi Stemming Halaman ini menampilkan form aplikasi penentuan kata dasar dari kata berimbuhan dari sistem yang dirancang.

Format Masuk

Aplikasi Penentuan Kata Dasar dan Kata Berimbuhan pada kalimat Bahasa Indonesia

Masuk Keluar

5. JADWAL PENELITIAN

MINGGU	PROGRESS PROJECT
8	Mencari dataset untuk melengkapi project
9	Membuat desain UI/UX aplikasi web
10	Membuat tampilan dengan codingan
11	Membuat tampilan dengan codingan dan menambahkan beberapa variasi
12	Melihat kembali codingan dan tampilan jika ada yang kurang atau yang salah
13	Evaluasi project dari awal sampai akhir agar lebih maksimal
14	Menghosting web yang telah saya buat

DAFTAR PUSTAKA

Afuan, L. (2013). Stemming Dokumen Teks Bahasa Indonesia. *Jurnal Telematika*, 6(2), 34–40.

Khuzaemah, Emah, D. (2022). *Buku Ajar Sejarah , Kedudukan , Dan Fungsi*.

Nuraeni, R., Mulyati, S., Putri, T. E., Rangkuti, Z. R., Pratomo, D., Ak, M., Ab, S., Soly, N., Wijaya, N., Operasi, S., Ukuran, D. A. N., Terhadap, P., Sihaloho, S., Pratomo, D., Nurhandono, F., Amrie, F., Fauzia, E., Sukarmanto, E., Partha, I. G. A., ... Abyan, M. A. (2017). No 主観的健康感を中心とした在宅高齢者における 健康関連指標に関する 分散構造分析Title. *Diponegoro Journal of Accounting*, 2(1), 2–6. http://i-lib.ugm.ac.id/jurnal/download.php?dataId=2227%0A???%0Ahttps://ejournal.unisba.ac.id/index.php/kajian_akuntansi/article/view/3307%0Ahttp://publicacoes.cardiol.br/portal/ijcs/portugues/2018/v3103/pdf/3103009.pdf%0Ahttp://www.scielo.org.co/scielo.php?scri

Simarangkir, M. S. H. (2017). Studi Perbandingan Algoritma - Algoritma Stemming Untuk Dokumen Teks Bahasa Indonesia. *Jurnal Inkofar*, 1(1), 40–46. <https://doi.org/10.46846/jurnalinkofar.v1i1.2>

Wahyuni, F., & Herlinda, H. (2021). Paradigma Pembelajaran Efektif Bahasa Dan Sastra Indonesia. *Gurindam: Jurnal Bahasa Dan Sastra*, 1(2), 40. <https://doi.org/10.24014/gjbs.v1i2.12786>