

Perbandingan Algoritma Nazief & Adriani Dengan Algoritma Idris Untuk Pencarian Kata Dasar

Adhi Prasadhatama^[1], Kristien Margi Suryaningrum^[2]

Teknik Informatika, Fakultas Teknik dan Desain, Universitas Bunda Mulia
email: aprasidatama@gmail.com

Teknik Informatika, Fakultas Teknik dan Desain, Universitas Bunda Mulia
email: ksuryaningrum@bundamulia.ac.id

Abstract

Stemming is a process used to search root words. Stemming can be implemented as preprocessing in text processing. There are many stemming algorithms, one of them is Nazief & Adriani and Idris. Nazief & Adriani algorithm is an algorithm created by Bobby Nazief and Mirna Adriani to find the root word of Bahasa Indonesia. While the Idris algorithm is an algorithm created by Norisma Idris to find the root words of Malay language. Indonesian and Malay have similar language styles, so researchers want to find out whether the Idris algorithm can be used as one of the Indonesian stemming algorithms. To know which algorithm is better, then in this research will be explained about process speed and accuracy of result of each of those algorithm. After knowing the result of comparison of stemming algorithm, hopefully can help you in choosing which algorithm will be used for find root word of Bahasa Indonesia.

Keywords: *Perbandingan, Nazief & Adriani, Idris, Stemming*

1. PENDAHULUAN

Komunikasi merupakan suatu proses dimana seseorang atau beberapa orang menciptakan dan menggunakan informasi agar terhubung dengan lingkungan dan orang lain. Komunikasi pada umumnya dilakukan secara lisan atau verbal yang dapat dimengerti oleh kedua belah pihak. Dengan adanya teknologi kita dapat berkomunikasi dengan satu dan lainnya dengan mudah. Namun tanpa disadari bahwa berkomunikasi langsung dengan sebuah program masih dianggap sulit. Dikarenakan oleh perbedaan bahasa maupun tatanan bahasanya itu sendiri. Sehingga program tidak dapat dengan mudah mengerti dari apa yang dimaksudkan oleh manusia.

Untuk mengerti apa yang dimaksud oleh manusia sebuah program harus memiliki beberapa kemampuan seperti *text preprocessing*, ekstraksi fitur, pencarian kata dasar, hingga pelatihan atau pembelajaran untuk mengembangkan pola pikirnya. *Stemming* merupakan suatu proses untuk menemukan kata dasar dari sebuah kata dengan menghilangkan semua imbuhan (*affixes*) baik yang terdiri dari awalan (*prefixes*), sisipan (*infixes*), akhiran (*suffixes*) dan kombinasi dari awalan dan akhiran (*confixes*) pada kata turunan. *Stemming* adalah inti dari teknik pemrosesan *natural language* untuk mendapatkan informasi kembali (*Information Retrieval*) yang efektif dan efisien dan secara luas dapat diterima oleh pengguna (*users*). *Stemming* digunakan untuk mengganti bentuk dari suatu kata menjadi kata dasar dari kata tersebut yang sesuai dengan struktur morfologi bahasa Indonesia yang baik dan benar. Pada pencarian kata dasar (*stemming*) khususnya bahasa Indonesia terdapat beberapa algoritma seperti Jelita Asian tahun 2005, Arifin &

Setiono tahun 2002, Nazief & Adriani tahun 1996, Ahmad Yusoff Sembok tahun 1996, Vega tahun 2001, Idris tahun 2001 dan *Enhanced Confix Stripping Stemmer* tahun 2008.

Mengetahui ada berbagai macam algoritma untuk *stemming* bahasa Indonesia. Penulis memutuskan untuk mempersempit ruang penelitian ke algoritma Nazief & Adriani dan algoritma Idris. Dikarenakan algoritma Nazief & Adriani merupakan algoritma *stemming* yang populer untuk memproses teks berbahasa Indonesia, sedangkan algoritma Idris sendiri kurang populer untuk *stemming* bahasa Indonesia. Maka dari itu penelitian ini berjudul "Analisis Perbandingan Algoritma Nazief & Adriani dengan Algoritma Idris untuk Pencarian Kata Dasar pada Teks Bahasa Indonesia Berbasis *Website*". Pada penelitian ini akan membahas mengenai cara kerja, kelebihan & kekurangan, hingga hasil dari tiap algoritma yang akan dibandingkan.

Information retrievals System

Information retrievals System atau sistem temu balik informasi merupakan bagian dari *computer science* tentang pengambilan informasi dari dokumen-dokumen yang didasarkan pada isi dan konteks dari dokumen-dokumen itu sendiri.[1] Informasi yang diambil dapat berupa teks, gambar, audio, dan video yang berguna untuk melakukan penyimpanan, pencarian, dan pemeliharaan informasi.

Text preprocessing

Text preprocessing berfungsi mengubah data tekstual yang tidak terstruktur ke dalam data terstruktur dan disimpan dalam basis data. Tahap

preprocessing terdiri dari beberapa langkah yaitu : *case floding, tokenization, filtering dan stemming*. [5]

Pada proses *case floding* karakter selain huruf dihilangkan dan semua huruf diubah menjadi *lowercase*. *Tokenization* adalah proses dimana sebuah kalimat dipotong untuk menghasilkan kata-kata yang akan digunakan untuk proses selanjutnya. Pengambilan kata-kata penting dari hasil *tokenization* merupakan proses yang dilakukan dalam *filtering*. Proses *filtering* memiliki dua teknik yaitu *stop list* (membuang kata yang kurang penting) dan *word list* (menyimpan kata yang penting). Setelah mendapatkan kata-kata yang dirasa penting, selanjutnya dilakukan proses *stemming*. Proses pada tahap *stemming* ialah melakukan pengolahan dan pencarian kata dasar dari hasil *filtering* atau sekumpulan kata-kata yang sudah dianggap penting.

Algoritma Stemming Nazief & Adriani

Algoritma yang dibuat oleh Bobby Nazief dan Mirna Adriani memiliki tahapan sebagai berikut. [2]

1. Cari kata yang akan *distem* dalam kamus. Jika ditemukan maka diasumsikan bahwa kata tersebut adalah kata dasar. Maka algoritma berhenti.
2. Infleksi akhiran (“-lah”, “-kah”, “-ku”, “-mu”, atau “-nya”) dibuang. Jika berupa partikel (“-lah”, “-kah”, “-tah” atau “-pun”) maka langkah ini diulangi lagi untuk menghapus kata ganti posesif (“-ku”, “-mu”, atau “-nya”), jika ada.
3. Hapus penurunan akhiran (“-i”, “-an” atau “-kan”). Jika kata ditemukan di kamus, maka algoritma berhenti. Jika tidak maka ke langkah 3a.
 - a. Jika “-an” telah dihapus dan huruf terakhir dari kata tersebut adalah “-k”, maka “-k” juga ikut dihapus. Jika kata tersebut ditemukan dalam kamus maka algoritma berhenti. Jika tidak ditemukan maka lakukan langkah 3b.
 - b. Akhiran yang dihapus (“-i”, “-an” atau “-kan”) dikembalikan, lanjut ke langkah 4.
4. Hapus penurunan awalan. Jika pada langkah 3 ada akhiran yang dihapus maka pergi ke langkah 4a, jika tidak pergi ke langkah 4b.
 - a. Periksa tabel kombinasi awalan akhiran yang tidak diizinkan. Jika ditemukan maka algoritma berhenti, jika tidak pergi ke langkah 4b.
 - b. Pada langkah ini dilakukan perulangan sebanyak tiga kali. Tentukan tipe awalan kemudian hapus awalan. Jika kata dasar belum juga ditemukan lakukan langkah 5, jika sudah maka algoritma berhenti. Catatan: jika awalan kedua sama dengan awalan pertama algoritma berhenti.
5. Melakukan *Recoding*.

6. Jika semua langkah telah selesai tetapi tidak juga berhasil maka kata awal diasumsikan sebagai kata dasar lalu proses diakhiri.

Algoritma Stemming Idris

Berikut tahapan-tahapan mengenai *stemming* menggunakan algoritma Idris adalah sebagai berikut. [4]

1. a. Kata yang belum di-*stemming* dicari pada kamus umum atau Kamus Besar Bahasa Indonesia (KBBI). Jika kata itu langsung ditemukan, berarti kata tersebut adalah kata dasar. Kata tersebut dikembalikan dan algoritma dihentikan.
 - b. Kata yang belum di-*stemming* dicari pada kamus lokal. Jika kata itu langsung ditemukan, berarti kata tersebut adalah kata dasar. Kata tersebut dikembalikan dan algoritma dihentikan. Kamus lokal digunakan karena algoritma Idris dibuat untuk *stemming* bahasa melayu, yang dimana kamus lokal berisi kata-kata dasar bahasa melayu. Sedangkan *stemming* bahasa Indonesia tidak memerlukan kamus lokal.
2. Menghilangkan turunan awalan. Langkah ini terus dilakukan sampai tidak ada lagi turunan awalan. Jika tidak ada lagi, maka lanjutkan ke langkah berikutnya. Untuk beberapa kombinasi imbuhan, dilakukan penghilangan awalan terlebih dahulu. Yaitu pada kombinasi imbuhan “ber-”, “ber-”, “ber-an”, “men-i”, “di-i”, “pe-i”, “ter-i”.
3. Hilangkan ifleksi akhiran terlebih dahulu. Jika hal ini berhasil dan akhiran adalah partikel (“-lah”, “-pun” dan “-kah”), langkah ini dilakukan lagi untuk menghilangkan kata ganti akhiran posesif (“ku”, “mu” atau “nya”).
4. Hilangkan turunan akhiran. Langkah ini terus dilakukan sampai tidak ada lagi penurunan akhiran. Jika tidak ada lagi, maka lanjutkan ke langkah berikutnya.
5. Setelah setiap penghilangan imbuhan dilakukan, maka lakukan pengecekan menggunakan kamus. Jika kata ditemukan, maka algoritma berhenti dan tidak perlu dilakukan pengecekan terhadap imbuhan lainnya. Jika sampai selesai penghilangan imbuhan masih belum menemukan kata dasar, maka dilakukan *recoding*.
6. Jika semua langkah sudah dilakukan termasuk *recoding* dan tidak juga ditemukan dalam kamus, maka algoritma ini akan manganggap kata semula sebagai kata dasar.

2. METODE PENELITIAN

Metode pengembangan sistem yang digunakan adalah model *watfall*. Berikut merupakan jbaran mengenai metode pengembangan sistem model *waterfall*. [3]

1. Analisis Kebutuhan

Dalam tahapan ini penulis menentukan kebutuhan-kebutuhan pada sistem Analisa

Perbandingan Algoritma Nazief & Adriani dengan Algoritma Idris untuk Pencarian Kata Dasar pada Teks Bahasa Indonesia Berbasis *Website* baik itu kebutuhan fungsional maupun kebutuhan non fungsional.

2. Desain Sistem

Dalam tahapan desain sistem ini, penulis membuat perancangan dari model atau desain sistem dengan menggunakan beberapa alat bantu untuk menggambarkan sistem berjalan ataupun sistem baru yang akan dikembangkan secara logika. Untuk rincian prosedur menggunakan *flowchart* sedangkan untuk menggambarkan susunan logis antar data dan hubungannya dengan sistem penulis menggunakan *Entity Relationship Diagram* (ERD).

3. Implementasi dan Pengujian Unit

Pada tahap ini, penulis melakukan penerjemahan desain yang telah dibuat ke dalam bentuk *software* yang dirancang dengan bahasa pemrograman PHP dan basis data menggunakan MySQL. Selanjutnya melakukan pengujian terhadap program yang dibangun per unit atau per modul kerja. Dimana semua fungsi-fungsi *software* tersebut diuji cobakan, agar *software* bebas dari *error* dan hasilnya harus benar-benar sesuai dengan kebutuhan yang sudah didefinisikan sebelumnya.

4. Pengujian Sistem

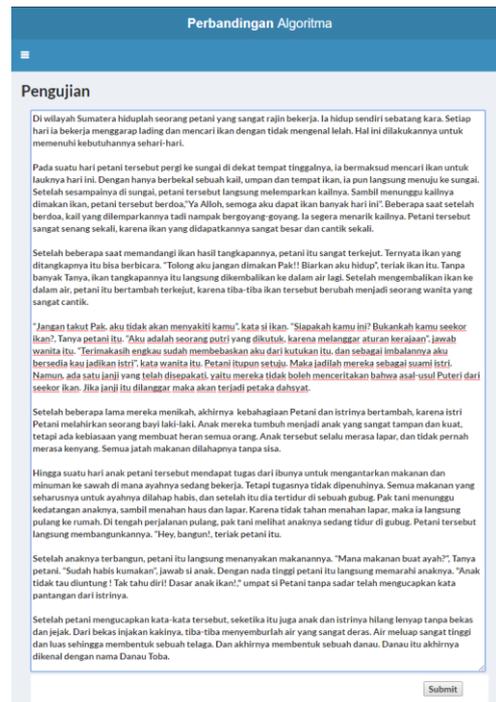
Pada tahap ini program yang telah dibuat dan diuji tiap unitnya kemudian disatukan menjadi suatu sistem dan diuji secara keseluruhan berguna untuk menguji tingkat integrasi antar unit yang dibuat sebelumnya.

5. Maintenance atau Perawatan

Pada tahap ini program akan diawasi kinerja dan hasil keluarannya. Jika terjadi sebuah kesalahan pada saat program beroperasi atau hasil keluaran tidak sesuai dengan yang diharapkan, maka akan dilakukan perawatan yang berguna untuk memperbaiki program.

3. HASIL DAN PEMBAHASAN

Objek pengujian untuk mengetahui kecepatan proses dan akurasi dari algoritma Nazief & Adriani dan Idris adalah berupa wacana berupa teks. Berikut merupakan tampilan dari laman masukkan pengujian dengan wacana berjudul "Legenda Danau Toba" dengan jumlah kata 520.



Gambar 1. Laman masukkan pengujian

Pada gambar 1 menampilkan bagaimana teks wacana dimasukkan untuk memulai pengujian. Setelah teks wacana selesai di masukkan. Pengujian dapat memulai proses pengujian dengan menekan tombol *submit*.

Hasil Stemming Nazief & Adriani

Durasi : 7.800290393829 detik
Jumlah Kata Hasil 255
wilayah sumatera hidup petan rajin hidup batang kara garap lading cari ikan kenal lelah laku penuh butuh hari petan pergi sungai tinggal cari ikan lauk bekal kail umpan ikan langsung sungai sampai sungai petan langsung lempar kail tunggu kail makan ikan petan doa ya alloh moga ikan doa kail lempar nampak goyang teriak kail petan senang ikan dapat cantik pandang ikan hasil tangkap petan kejut ikan tangkap bicara tolong makan biar hidup teriak ikan ikan tangkap langsung kembali air kembali ikan air petan tambah kejut ikan ubah wanita cantik takut sakti si ikan ekor ikan petan putri kutuk langgar atur raja wanita terimakasih engkau bebas kutuk imbal sedia kau jadi istri wanita petan itu tuju suami istri janji sepakat cerita usul puter ekor ikan janji langgar petaka dahsyat meni bahgia petan istri tambah istri petan lahir bayi laki laki anak tumbuh anak tapan kuat biasa heran orang anak lapar kenyang jatah makan lapah sisa anak petan tugas ibu antar makan minum sawah ayah tugas penuh makan ayah lapah habis tidur gubug tani tunggu datang anak nah haus lapar tahan nah lapar langsung pulang rumah jalan pulang tani anak tidur gubug petan langsung bangun hey bangun teriak petan anak bangun petan langsung makan makan anak petan habis kumakan si anak nada petan langsung marah anak anak tau untung dasar anak ikan umpat si petan sadar pantang istri petan anak istri hilang lempay bekas jejak bekas injak kail sembur air deras air luap luas bentuk telaga bentuk danau danau kenal nama danau toba

Jumlah Kata Salah : 19

Gambar 2. Hasil stemming Nazief & Adriani

Pada gambar 2 menampilkan hasil dari pengolahan algoritma Nazief & Adriani. Atribut yang ditampilkan pada hasil *stemming* Nazief & Adriani berupa durasi pemrosesan, jumlah kata yang dihasilkan, data kata dari hasil *stemming*, dan jumlah kata salah. Jumlah kata salah merupakan jumlah dari kata yang dihasilkan dari *stemming* namun tidak sesuai dengan kata dasarnya.

Hasil Stemming Idris

Durasi : 7.9879069328308 detik
Jumlah Kata Hasil 255
wilayah sumatera hidup tani rajin hidup batang kara garap lading cari ikan mengenal lelah dilakukannya penuh kebutuhannya hari tani pergi sungai tinggal cari ikan lauk bekal kail umpan ikan langsung sungai sampai sungai tani langsung empur kail tunggu kail makan ikan tani doa ya alloh moga ikan doa kail dilemparkannya nampak goyang tarik kail tani senang ikan dilakukannya cantik pandang ikan hasil tangkapannya tani kejut ikan tangkap bicara tolong makan biar hidup teriak ikan ikan tangkapannya langsung kembali air mengembalikan ikan air tani tambah kejut ikan ubah wanita cantik takut menyakiti si ikan ekor ikan tani putri kutuk langgar atur raja wanita terimakasih engkau membebaskan kutuk imbalannya sedia kau jadi istri wanita tani itu tuju suami istri janji sepakat cerita usul puter ekor ikan janji langgar petaka dahsyat nikah bahgia tani istri tambah istri tani melahirkan bayi laki laki anak tumbuh anak tapan kuat biasa heran orang anak lapar kenyang jatah makan lapah sisa anak tani tugas ibu gantar makan minum sawah ayah tugas dipemuhnya makan ayah lapah habis tidur gubug tani tunggu datang anak tahan haus lapar tahan tahan lapar langsung pulang rumah jalan pulang tani anak tidur gubug tani langsung membangunkannya hey bangun teriak tani anak bangun tani langsung makanannya makan ayah tani habis kumakan si anak nada tani langsung parah anak anak tau untung dasar anak ikan umpat si tani sadar pantang istri tani anak istri hilang lempay bekas jejak bekas injak kail menyembur air deras air luap luas bentuk telaga bentuk danau danau kenal nama danau toba

Jumlah Kata Salah : 19

Gambar 3. Hasil stemming Idris

Pada gambar 3 menampilkan hasil dari pengolahan algoritma Idris. Atribut yang ditampilkan pada hasil *stemming* Idris berupa durasi pemrosesan, jumlah kata yang dihasilkan, data kata dari hasil *stemming*, dan jumlah kata salah. Jumlah kata salah merupakan jumlah dari kata yang dihasilkan dari *stemming* namun tidak sesuai dengan kata dasarnya.

Rekap	
Jumlah Data yang telah dimasukkan : 5	
Jumlah Kata yang telah dimasukkan : 3174	
Nazief & Adriani	
Kecepatan : 0.035065172133544 detik/kata	
Akurasi : 97.50271759033203%	
Idris	
Kecepatan : 0.027079922330063 detik/kata	
Akurasi : 91.36622314453125%	

Gambar 4. Laman Rekap hasil Pengujian

Pada gambar 4 menampilkan rekap dari pengujian yang telah dilakukan, seperti jumlah data yang telah dimasukkan, total jumlah kata, hingga kecepatan dan akurasi dari masing-masing algoritma. Kecepatan didapatkan dari membagi jumlah waktu pemrosesan dengan jumlah kata. Sementara akurasi didapatkan dari rata-rata hasil jumlah kata yang benar dibagi dengan jumlah kata yang dimasukkan di awal.

Berikut merupakan hasil pengujian yang telah dilakukan terhadap 5 teks wacana berbeda.

Tabel 1. Hasil Pengujian Algoritma

No	Nama wacana	Jumlah kata	Algoritma	Waktu proses (detik)	Akurasi (%)
1	Legenda Danau Toba	520	Nazief & Adriani	7.80	92.55
2	Legenda Danau Toba	520	Idris	7.99	92.55
3	Legenda Sangkuriang	645	Nazief & Adriani	15.8	97.07
4	Legenda Sangkuriang	645	Idris	11.09	88.56
5	Cerita Rakyat Timun Mas	637	Nazief & Adriani	10.98	99.1
6	Cerita Rakyat Timun Mas	637	Idris	8.15	92.84
7	Legenda	639	Nazief	10.27	99.68

8	Harimau Makan Durian Legenda Harimau Makan Durian	639	& Adriani Idris	8.24	88.22
9	Cinderelas	733	Nazief & Adriani	14.58	99.11
10	Cinderelas	733	Idris	10.43	94.67

4. KESIMPULAN

Setelah dilakukan pengujian, maka diperoleh kesimpulan sebagai berikut:

1. Algoritma *stemming* Idris memiliki waktu pemrosesan yang lebih cepat dibandingkan dengan algoritma *stemming* Nazief & Adriani.
2. Algoritma *stemming* Nazief & Adriani memiliki akurasi hasil yang lebih tinggi dibandingkan dengan algoritma *stemming* Idris.

5. REFERENSI

- [1] Hasanah, Nur. 2017. Sistem Pencarian Skripsi Berbasis *Information Retrieval* di FASTIKOM UNSIQ. *Jurnal PPKM*. 1 (1):105-113.
- [2] Nugroho, Hargyo Tri. 2017. Pengaruh Algoritma *Stemming* Nazief-Adriani Terhadap Kinerja Algoritma *Winnowing* Untuk Mendeteksi Plagiarisme Bahasa Indonesia. *ULTIMA Computing*. 9 (1):36-40.
- [3] Oktarino, Ade. 2015. Perancangan Sistem Informasi Rekam Medis Pasien pada Klinik Bersalin Kasih Ibu Menggunakan Metode *Waterfall*. *Scientia Journal*. 4 (3):239-247.
- [4] Permatasari, Nindy. 2016. Analisa Perbandingan Algoritma Idris dan Algoritma *Enhanced Confix* (ECS) *Stemmer* Pada Dokumen Teks Bahasa Indonesia. *Skripsi*. Universitas Komputer Indonesia, Bandung.
- [5] Rahmawati, L. 2016. *Analisa Clustering Menggunakan Metode K-Means dan Hierarchical Clustering (Studi Kasus: Dokumen Skripsi Jurusan Kimia, Fmipa, Universitas Sebelas Maret)*. Surakarta: ITSmart: Jurnal Teknologi dan Informasi. Vol. 3, No. 2: 66-73.