

Penerapan Algoritma Nazief & Adriani Dalam Stemming Bahasa Batak Angkola

¹Muhammad Rosikhu, ¹Ema Utami, ¹Ainul Yaqin

¹ Teknik Informatika Program Pascasarjana Universitas Amikom
m.rosikhu@students.amikom.ac.id, ema.u@amikom.ac.id, ainulyaqin@amikom.ac.id

ABSTRAK

Penelitian dengan judul “Penerapan Algoritma Nazief & Adriani Dalam Stemming Bahasa Batak Angkola” dimulai dengan mengumpulkan data berupa kata-kata dan jenis imbuhan yang digunakan dalam bahasa Batak Angkola. Selanjutnya dianalisis untuk menghasilkan daftar kata dasar serta aturan dari penggunaan imbuhan. Kemudian berdasarkan daftar kata dasar dan aturan imbuhan tersebut, disusun sebuah algoritma dengan melakukan modifikasi dari algoritma Nazief & Adriani dan juga algoritma Arifin Setiono yang dapat digunakan dalam melakukan stemming kata-kata dalam bahasa Batak Angkola, pada penerapan algoritma stemming Nazief dan Adriani dengan sampel data yang diujikan berjumlah 3280 kata dibuat menjadi satu kamus dengan type data .JSON. Algoritma ini mampu memberikan tingkat akurasi dan presisi yang sangat tinggi, yaitu 99,33%, dimana kata yang mengalami Overstemming berjumlah 7 kata dan yang mengalami Understemming berjumlah 15 dari keseluruhan kata dalam kamus yang berjumlah 3280 kata.

Keyword: Algoritma, Nazief & Adriani, Stemming

1. PENDAHULUAN

Bahasa sebagai alat komunikasi yang mempunyai kekhasan tertentu yang menjadi jati diri dari suatu bangsa, negara maupun daerahnya (Taufik dkk., 2017). Banyak keistimewaan yang dimiliki oleh bangsa Indonesia. Selain bahasa Indonesia yang menjadi bahasa resmi, Indonesia juga memiliki beragam bahkan ratusan bahasa daerah yang tersebar di setiap suku yang terdapat diseluruh penjuru Indonesia.

Bahasa daerah adalah penanaman identitas karakter budaya dari suku bangsa. Bahasa daerah dapat juga di katakan sebagai induk bahasa karena awal anak mengenal bahasa yang di mengerti adalah bahasa daerah.

Begitu banyaknya suku dan bahasa daerah yang ada di Indonesia, dengan jumlah bahasa daerah 718 menurut Kementerian Pendidikan dan Kebudayaan pada tahun 2022, dengan suku bangsa lebih dari 1340 yang terdapat di Indonesia menurut Badan Pusat Statistik pada tahun 2010.

Suku Batak adalah suku dengan jumlah penduduk terbanyak ketiga di Indonesia dengan jumlah lebih dari 14 juta jiwa sesuai dengan data yang di dapat oleh Badan Pusat Statistik pada Tahun 2020. Dalam suku batak sendiri terdapat beberapa rumpun bahasa daerah batak, antara lain :

1. Bahasa Batak Toba: Digunakan di Kabupaten Toba, Kabupaten Samosir, Kabupaten Tapanuli Utara dan Kabupaten Humbang Hasundutan.
2. Bahasa Batak Angkola: Digunakan di Kabupaten Tapanuli Selatan, Kabupaten Padang Sidempuan, Kabupaten Padang Lawas Utara, Kabupaten Padang Lawas, dan sebagian Kabupaten Mandailing Natal.
3. Bahasa Batak Mandailing: Digunakan di Kabupaten Mandailing Natal dan sebagian juga digunakan di Riau Bagian Utara dan sedikit di Sumatera Barat.
4. Bahasa Batak Simalungun: Digunakan di Kabupaten

Simalungun, Kabupaten Serdang Begadai, Kota Pematang Siantar, Kabupaten Deli Serdang, Kabupaten Karo, dan Kabupaten Dairi.

5. Bahasa Batak Karo: Digunakan di Kabupaten Karo, Langkat, Dairi, Deli Serdang, Medan.
6. Bahasa Batak Pakpak: Digunakan di Kabupaten Dairi, Pakpak Barat, sebagian di wilayah Humbang Hasundutan dan Tapanuli Tengah.

Bahasa Batak Angkola adalah salah satu dari beberapa jenis bahasa batak yang ada di daerah Sumatera Utara, yaitu yang terdapat di daerah Tapanuli Bagian Selatan. Di karenakan wilayah pengguna bahasa Batak Angkola berdekatan dengan Toba dan Mandailing, bahasa mereka pun mempunyai kemiripan, namun di bedakan dari intonasi bahasa Batak Angkola yang lebih halus dibandingkan bahasa Batak Toba dan lebih tegas dibandingkan bahasa Batak Mandailing.

Sejumlah kata bahasa Indonesia memiliki padanan yang tepat di dalam Bahasa Batak Angkola. Namun, tidak sedikit kata bahasa Indonesia yang tidak memiliki padanan kata dalam bahasa batak Angkola tersebut.

Stemming adalah teknik untuk mendapatkan kata dasar, dari suatu kata yang memiliki atau mengandung prefiks (awalan), suffiks (akhiran), infiks (imbuhan), dan konfiks (awalan-akhiran) (Adriani dkk., 2007). Stemming sebagai alat yang efektif dan efisien untuk mengolah kata dasar dalam pengambilan teks. Stemming merupakan salah satu tahapan dalam text preprocessing.

Algoritma stemming beredar sangat banyak dalam teknik pencarian kata dasar suatu kata atau kalimat. Stemming digunakan pada pengambilan informasi untuk mengurangi varian bentuk kata menjadi akar kata atau kata dasar untuk meningkatkan efektifitas pengambilan suatu kata atau informasi.

Beberapa jenis Algoritma Stemming yang beredar:

1. Algoritma Nazief & Adriani: Algoritma yang saat ini populer dan banyak dipakai dalam pemrosesan *stemming* bahasa Indonesia. Waktu yang digunakan biasanya lebih lama, namun akurasi lebih baik dibandingkan dengan algoritma *stemming* lainnya.
2. Algoritma Arifin Setiono: Algoritma ini mempunyai cara dengan mengurangi awalan terlebih dahulu baru akhiran. Namun apabila tidak menemukan kata dasarnya, ia akan mengembalikan awalan tersebut dan kembali melakukan *stemming*.
3. Algoritma Porter: Merupakan algoritma yang banyak digunakan dalam proses *stemming* bahasa Inggris.

Dan masih banyak lagi algoritma *stemming* yang bisa digunakan dengan berbagai kelebihan dan kekurangannya masing-masing. Dengan memilih yang tepat atau cocok untuk *stemming* bahasa atau kata tertentu.

Beberapa penelitian sebelumnya mengatakan bahwa Algoritma Porter dapat melakukan proses *stemming* dengan waktu yang lebih cepat, dan Algoritma Nazief & Adriani dapat menghasilkan akurasi *stemming* yang lebih baik (Wahyudi dkk., 2017). Penggunaan kamus dalam proses menggunakan algoritma Nazief & Adriani sangat mempengaruhi hasil *stemming*, semakin lengkap kamus yang dimasukkan maka hasil proses *stemming* akan semakin akurat (Agusta, 2009). Begitu pula dengan kesimpulan bahwa algoritma Nazief & Adriani lebih unggul 0,1% dalam tingkat akurasinya dengan algoritma Arifin Setiono, dan juga lebih unggul 0,9% dengan algoritma Porter (Rezalina, 2016). Dalam pengujian yang menggunakan algoritma Vega dengan komparasi algoritma Nazief & Adriani, terdapat beberapa kesimpulan bahwa untuk rata-rata waktu proses *stemming*, algoritma Nazief & Adriani lebih lama dibandingkan dengan algoritma Vega. Sedangkan untuk tingkat akurasi, algoritma Nazief & Adriani mendapatkan hasil yang lebih baik dari pada algoritma Vega (Sahat & Simarangkir, 2017).

2. METODOLOGI

2.1 Jenis, Sifat dan Pendekatan Penelitian

Sifat penelitian yang dilakukan menggunakan metode deskriptif kuantitatif. Dengan menggunakan imbuhan dan kelompok kata dalam bahasa Batak Angkola yang kemudian dengan memberi batasan parameter dan juga sejumlah dataset yang telah di dapat ditetapkan dan dilakukan pengujian untuk mengetahui tingkat akurasi. Dan juga kumpulan kata dasar dan imbuhan tersebut hendak digunakan sebagai dasar dalam proses *stemming* untuk bahasa Batak Angkola.

2.2 Metode Pengumpulan Data

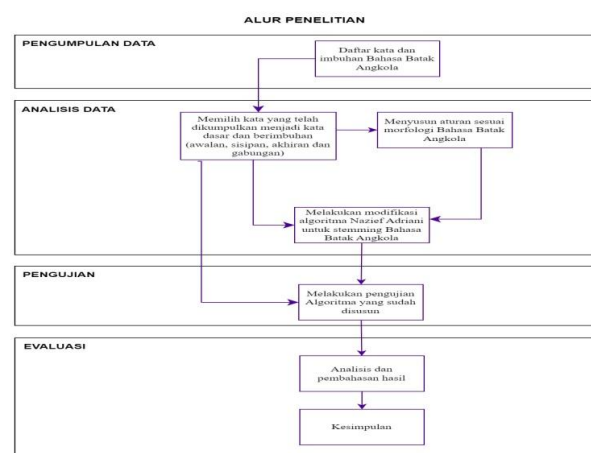
Penelitian ini dilakukan melalui studi pustaka dengan menggunakan bahan referensi yang terkait dengan penelitian, baik melalui buku, artikel, jurnal, dan lainnya yang berkaitan dengan algoritma Nazief & Adriani dan algoritma Arifin Setiono dan tata bahasa dalam bahasa Batak Angkola. Menggunakan data kumpulan morfem, imbuhan dan kata dalam bahasa Batak Angkola

2.3 Metode Analisis Data

Pada penelitian ini akan dikumpulkan data berupa kata-kata dan jenis imbuhan yang digunakan dalam bahasa Batak Angkola. Selanjutnya dianalisis untuk menghasilkan daftar kata dasar serta aturan dari penggunaan imbuhan. Kemudian berdasarkan daftar kata dasar dan aturan imbuhan tersebut, disusun sebuah algoritma dengan melakukan modifikasi dari algoritma Nazief & Adriani dan juga algoritma Arifin Setiono yang dapat digunakan dalam melakukan *stemming* kata-kata dalam bahasa Batak Angkola

2.4 Alur Penelitian

Bagian ini berisi diagram alur langkah penelitian secara lengkap dan terinci termasuk di dalamnya tercermin algoritma, rute, pemodelan-pemodelan, desain, yang terkait dengan aspek perancangan sistem

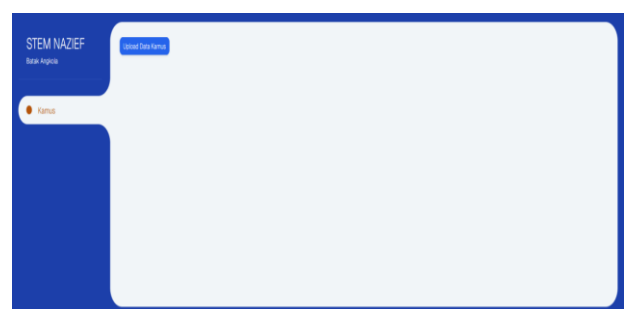


Gambar 1. Alur Penelitian

3. HASIL DAN PEMBAHASAN

3.1 Implementasi Aplikasi Stemming Bahasa Batak Angkola

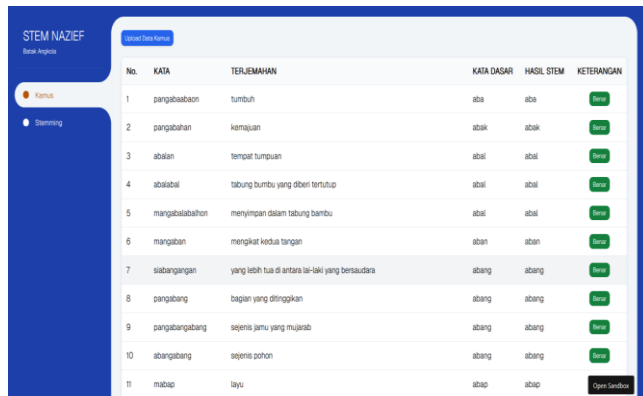
Graphical User Interface (GUI) merupakan tampilan yang berfungsi sebagai *window* atau jendela yang menghubungkan pengguna dengan aplikasi *stemming* bahasa Batak Angkola. Pada penerapan penggabungan algoritma watermarking dan parameternya menggunakan pemrograman *Java Script*. Tampilan GUI dapat dilihat pada gambar 4.1.



Gambar 2. Interface Proses Upload Kamus

Gambar di atas adalah tampilan awal saat aplikasi *Stemming* dengan algoritma Nazief & Adriani dijalankan.

Pada halaman ini ada satu tombol atau *button* yang berfungsi untuk mengupload kamus Bahasa Batak Angkola sehingga aplikasi bisa digunakan dalam melakukan *Stemming* Bahasa Batak Angkola. Setelah kamus di upload maka akan terlihat seperti pada gambar 2 di bawah.



No.	KATA	TERJEMAHAN	KATA DASAR	HASIL STEM	KETERANGAN
1	pangabababon	tumbuh	aba	aba	Benar
2	pangababan	kemajuan	abak	abak	Benar
3	ababan	tempat tumpuan	abai	abai	Benar
4	abalabal	tabung bambu yang diberi tutup	abai	abai	Benar
5	mangabababon	menyimpan dalam tabung bambu	abai	abai	Benar
6	mangan	mengikat kedua tangan	aban	aban	Benar
7	salabangan	yang lebih tua di antara laki-laki yang bersaudara	abang	abang	Benar
8	pangabang	bagian yang ditinggikan	abang	abang	Benar
9	pangabangabang	sejenis jamu yang mujarab	abang	abang	Benar
10	abangabang	sejenis pohon	abang	abang	Benar
11	masap	layu	asap	asap	Benar

Gambar 3. Interface Kamus setah di Upload

Pada gambar di atas dapat dilihat bahwa kamus yang sudah peneliti upload langsung didetailkan mulai dari kata yang diberi imbuhan, terjemahan, kata dasar dan hasil *stemming*-nya. Setelah itu baru peneliti bisa melakukan *stemming* kata. Kamus Bahasa Batak Angkola ini dibuat dalam bentuk file bertipe data .JSON, karena type data ini merupakan type data yang general dan mudah dibaca oleh semua bahasa pemrograman pada saat melakukan *stemming*. Adapun jumlah kata dalam kamus Bahasa Batak Angkola ini adalah 3280 kata.

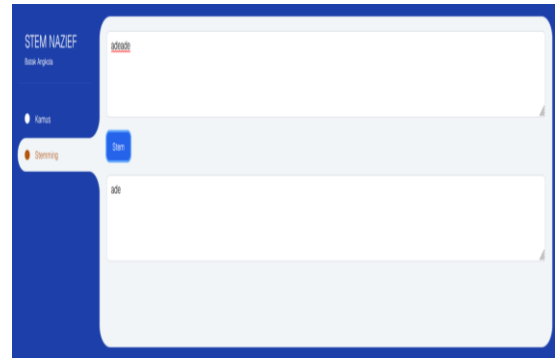
3.2 Parameter Pengujian Algoritma Stemming Nazief dan Andriani

Pada penelitian ini peneliti menggunakan dua parameter pengujian, yaitu mengukur tingkat akurasi dan presisi dalam penerapan algoritma *stemming* Nazief dan Andriani pada Bahasa Batak Angkola.

1. Akurasi
 - Menghitung presentase kata yang berhasil di-*stemming* dengan benar dibandingkan total kata yang di uji.
 - Rumus : $(\text{Jumlah kata yang benar di-stemming} / \text{total kata yang diuji}) \times 100\%$.
2. Presisi
 - Mengukur ketepatan algoritma dalam menghasilkan kata dasar yang benar
 - Rumus : $\text{True Positive} / (\text{True Positive} + \text{False Positive})$

3.3 Data Penelitian dan Pembahasan

Pengujian dilakukan dengan mencoba melakukan *stemming* pada beberapa kata pada bahasa Batak Angkola. Pengujiannya dapat dilihat pada gambar di bawah.



Gambar 4. Uji Coba Perkata

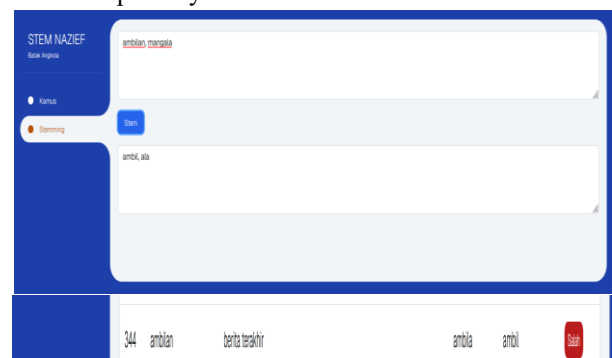
Interface di atas adalah tempat untuk melakukan *stemming* kata Bahasa Batak Angkola. Pada uji coba ini peneliti mencoba kata “adeade” yang artinya “cacing pita” dengan kata dasar dari hasil *stemming* “ade” ketika peneliti mengecek kamus menunjukkan hasil *stemming* benar artinya tidak ada *overstemming* ataupun *Understemming*.

Uji coba berikutnya peneliti lakukan dengan melakukan *stemming* kalimat atau lebih dari satu kata dapat dilihat pada gambar berikut.



Gambar 5. Uji Coba Kalimat

Pengujian *stemming* kata pada kalimat atau lebih dari satu kata termasuk berhasil atau aplikasi mampu menampilkan kata dasar dari tiap kata pada bahasa batak Angkola ini dengan benar selama kata yang di *stemming* berada dalam kamus data bahasa batak Angkola yang di *upload*. Uji coba di atas menggunakan kata “adatu soada adeade” yang artinya “Sukurlah tidak ada cacing pita” dan aplikasi *stemming* kata dengan algoritma Nazief & Adriani mampu menampilkan kata dasar dari tiap kata yaitu “ada ada ade”.



Gambar 6. Overstemming

Overstemming adalah kondisi di mana proses *stemming* menghapus bagian kata yang seharusnya tidak dihapus,

sehingga menghasilkan kata dasar yang salah dalam algoritma Nazief dan Andriani.

Tabel 1. *Overstemming*

No	Urutan	Kata	Terjemahan	Kata Dasar	Stemming
1	344	ambilan	berita terakhir	ambila	ambil
2	2250	mangala	mengalah	kala	ala
3	2549	mangagohon	menghilangkan	mago	ago
4	2602	mangaruni	membuat demam	marun	arun
5	2608	mekel-ekel	tertawa-tawa	mekel	ekel
6	2623	marimbar	berpindah	mimbar	Imbar
7	2988	tarpalan	agak pelan	palan	lan

Ada beberapa penyebab utama terjadinya *overstemming* dalam algoritma Nazief dan Andriani:

1. Penghapusan Imbuhan yang Sebenarnya Bagian dari Kata Dasar

- Kata "ambila" dianggap memiliki ahiran "an" sehingga menjadi "ambil"
- Kata "palan" dianggap memiliki awalan "pa" sehingga menjadi "lan"

Padahal kata "ambila" dan "palan" itu adalah kata dasarnya.

2. Kesalahan dalam Pengenalan Pola Imbuhan

Algoritma dapat salah dalam mengenali pola huruf yang mirip dengan imbuhan:

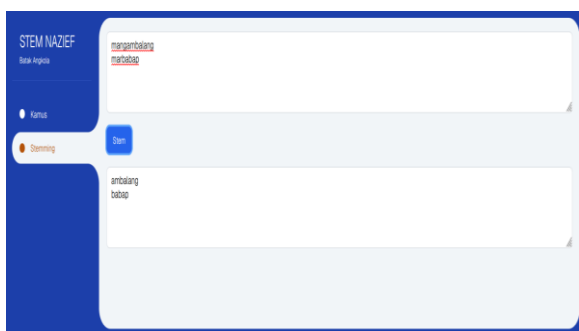
- Mengira bagian awal kata sebagai awalan
- Mengira bagian akhir kata sebagai akhiran
- Kesalahan identifikasi kombinasi imbuhan

3. Kompleksitas Morfologi Bahasa Batak Angkola

Bahasa Batak Angkola memiliki struktur morfologi yang kompleks dengan berbagai kombinasi imbuhan. Hal ini membuat algoritma kesulitan membedakan: Mana yang benar-benar imbuhan, Mana yang merupakan bagian dari kata dasar dan Bagaimana urutan penghapusan imbuhan yang tepat.

4. Aturan Penghapusan yang Terlalu Agresif

Algoritma mungkin menerapkan aturan penghapusan imbuhan secara berlebihan tanpa mempertimbangkan konteks kata secara keseluruhan.



Gambar 7. *Understemming*

Understemming adalah kondisi dimana proses *stemming* tidak berhasil menghapus semua imbuhan yang seharusnya dihilangkan, sehingga kata yang dihasilkan masih mengandung imbuhan dan bukan merupakan kata dasar yang sebenarnya, seperti uji coba di atas menggunakan kata "mangambalang dan marbabab" yang artinya "melemparkan

batu memakai balang dan berdebab" dengan kata dasar "baling dan bab" menjadi "ambalang dan babab" dimana hasil *stemming*nya melebihi kata dasar atau disebut dengan *understemming*. Ada beberapa pengujian yang mengalami *understemming* dapat dilihat pada tabel berikut.

Tabel 2. *Understemming*

No	Urutan	Kata	Terjemahan	Kata Dasar	Stemming
1	872	mangambalang	melemparkan batu memakai balang	balang	ambalang
2	945	marbabab	berdebab	bap	babap
3	2037	gumutur	gemetar	hutor	gumutur
4	2543	madung	sudah	ma	dung
5	2587	madung	sudah	mang	Dung
6	2694	marnambur	berembun	nambur	marnambur
7	2701	manapu	bertambah subur	napu	manapu
8	2735	manguas	haus	nguas	manguas
9	2880	monjap	bersembunyi	onjap	monjap
10	3010	panganon	makanan	pangan	mangan
11	3036	marparnamam	bermulut	parnamam	nanam
12	3087	mamodomi	meniduri	podom	modom
13	3089	parpodom	cara tidur	podom	modom
14	3090	podoman	tempat tidur	podom	Modom
15	3091	tarpodom	tertidor	podom	Modom

Berikut adalah penyebab-penyebab utama terjadinya *understemming* pada algoritma *stemming* Nazief dan Andriani:

1. Ketidakmampuan Mengenali Variasi Morfologis

Algoritma gagal mengenali beberapa variasi bentuk imbuhan dalam Bahasa Batak Angkola, seperti:

- Perubahan fonem pada proses pengimbuhan
- Variasi penulisan yang tidak standar
- Bentuk-bentuk tidak beraturan

2. Keterbatasan Aturan Transformasi

Algoritma memiliki keterbatasan dalam menangani:

- Kombinasi imbuhan yang kompleks
- Urutan penghapusan imbuhan yang tidak tepat
- Kasus-kasus khusus dalam pembentukan kata

3. Kesalahan dalam Urutan Penghapusan Imbuhan

Algoritma mungkin tidak menghapus imbuhan dalam urutan yang benar, menyebabkan:

- Imbuhan tertentu tetap tertinggal
- Kegagalan mengenali imbuhan berikutnya
- Proses *stemming* berhenti sebelum selesai

4. Kompleksitas Kata Majemuk

Kesulitan dalam menangani:

- Kata majemuk yang mengandung imbuhan
- Gabungan kata dengan struktur kompleks
- Kata-kata dengan makna khusus

5. Penanganan Konfiks yang Tidak Sempurna

Masalah dalam menangani:

- Konfiks yang terpisah (me-kan, ber-an, dll)
- Kombinasi konfiks dengan imbuhan lain
- Perubahan bentuk konfiks

4. Conclusion

Dari Penelitian Dengan Judul “Penerapan Algoritma Nazief & Adriani Dalam Stemming Bahasa Batak Angkola” Dapat Disimpulkan Bahwa Pada Penerapan Algoritma Stemming Nazief Dan Adriani Dengan Sampel Data Yang Diujikan Berjumlah 3280 Kata Dibuat Menjadi Satu Kamus Dengan Type Data .Json. Algoritma Ini Mampu Memberikan Tingkat Akurasi Dan Presisi Yang Sangat Tinggi, Yaitu 99,33%, Dimana Kata Yang Mengalami *Overstemming* Berjumlah 7 Kata Dan Yang Mengalami *Understemming* Berjumlah 15 Dari Keseluruhan Kata Dalam Kamus Yang Berjumlah 3280 Kata.

References

- Adriani, M., Asian, J., Nazief, B., Williams, H. E., & Tahaghoghi, S. M. M. (2007). Stemming Indonesian: A Confix-Stripping Approach. *Conferences In Research And Practice In Information Technology Series*, 38, 307–314. <https://doi.org/10.1145/1316457.1316459>
- Agusta, L. (2009). Perbandingan Algoritma Stemming Porter Dengan Algoritma Nazief & Adriani Untuk Stemming Dokumen Teks Bahasa Indonesia. Dalam *Konferensi Nasional Sistem Dan Informatika*.
- Amalia, D., Ainun, N., & Nasution, K. (2022). Analisis Komparatif Afiks Dalam Bahasa Batak Toba Dan Batak Angkola. *Talenta Conference Series: Local Wisdom, Social, And Arts*, 5(2), 50–58. <https://doi.org/10.32734/Lwsa.V5i1.1323>
- Dongoran, T. H., Naibaho, J., Sihombing, P., Sinaga, M., & Tampubolon, R. (1997). *Fonologi Bahasa Angkola*. Pusat Pembinaan Dan Pengembangan Bahasa, Departemen Pendidikan Dan Kebudayaan.
- Jaya, I., & Oktavia, Y. (2021). Analisis Kesalahan Berbahasa Indonesia Pada Media Berita Berbasis Digital. *Escience Humanity Journal*, 2(1), 59–64. <https://doi.org/10.37296/Escl.V2i1.19>
- Jumadi, J., Maylawati, D. S., Pratiwi, L. D., & Ramdhani, M. A. (2021). Comparison Of Nazief-Adriani And Paice-Husk Algorithm For Indonesian Text Stemming Process. *Iop Conference Series: Materials Science And Engineering*, 1098(3), 032044. <https://doi.org/10.1088/1757-899x/1098/3/032044>
- Kastowo, D., Saputra, A., Suryono, W. D., & Setyowati, E. (2022). Analisis Perbandingan Algoritma Nazief Adriani Dan Levenshtein Distance Untuk Mengukur Tingkat Similaritas Berita Berbahasa Jawa Menggunakan Rabin Krap. *Jnanaloka*, 3(1), 1–10. <https://doi.org/10.36802/Jnanaloka.V3-No1-1-10>
- Lakonawa, K. N., Mola, S. A. S., & Fanggidae, A. (2021). Nazief-Adriani Stemmer Dengan Imbuhan Tak Baku Pada Normalisasi Bahasa Percakapan Di Media Sosial. *Jurnal Komputer Dan Informatika*, 9(1), 65–73. <https://doi.org/10.35508/Jicon.V9i1.3749>
- Mandala, R., Koryanti, E., Munir, R., & Harlili. (2004). Sistem Stemming Otomatis Untuk Kata Dalam Bahasa Indonesia. *Seminar Nasional Aplikasi Teknologi Informasi*, 29–36.
- Meitaningsih, A., Aribowo, A. S., & Cahyana, N. H. (2020). Text Mining Untuk Mendeteksi Plagiasi Dokumen Dengan Penerapan Stemming Nazief-Adriani Dan Algoritma Smith-Waterman (2020). *Telematika (Jurnal Informatika Dan Teknologi Informasi)*, 17(2), 99–110. <https://doi.org/10.31315/Telematika.V1i1.3377>
- Nq, M. A., Manik, L. P., & Widiyatmoko, D. (2020). Stemming Javanese: Another Adaptation Of The Nazief-Adriani Algorithm. 2020 3rd International Seminar On Research Of Information Technology And Intelligent Systems, Isriti 2020, 627–631. <https://doi.org/10.1109/Isriti51436.2020.9315420>
- Ramza, H., Kadar, M. S., Abdurrahman, F., & Ab-Rahman, M. S. (2013). Analisis Hubungan Imbuhan Me Dan Ber Pada Setiap Perenggan Dalam Tulisan Cerita Pendek “Kembali Kasih” Abstrak. *Jurnal Matematika Terapan Indonesia*, 1(1).
- Rezalina, O. (2016). Perbandingan Algoritma Stemming Nazief & Adriani, Porter Dan Arifin Setiono Untuk Dokumen Teks Bahasa Indonesia. <http://repository.unmuhjember.ac.id/Id/Eprint/550>, 1–5.
- Ridwang, Afif, N., & Syahyadi, A. I. (2020). Sistem Pakar Pendeteksi Penyakit Pada Balita Menggunakan Metode Cosine Similitary Dan Algoritma Nazief Adriani. *Jurnal Instek (Informatika Sains Dan Teknologi)*, 5(1), 57–66.
- Russell-Rose, T., & Stevenson, M. (2009). The Role Of Natural Language Processing In Information Retrieval: Searching For Meaning And Structure. Dalam *Information Retrieval: Searching In The 21st Century* (Hlm. 215–231). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9780470033647.Ch10>
- Sahat, M., & Simarangkir, H. (2017). Studi Perbandingan Algoritma-Algoritma Stemming Untuk Dokumen Teks Bahasa Indonesia. Dalam *Jurnal Inkofar ** (Vol. 1, Nomor 1). Online.
- Subali, M. A. P., & Fatichah, C. (2019). Kombinasi Metode Rule-Based Dan N-Gram Stemming Untuk Mengenali Stemmer Bahasa Bali. *Jurnal Teknologi Informasi Dan Ilmu Komputer (Jtiik)*, 6(2), 219–228. <https://doi.org/10.25126/Jtiik.201961105>
- Suci, F. W., Hayatin, N., & Munarko, Y. (2022). In-Idris: Modification Of Idris Stemming Algorithm

- For Indonesian Text. *Iium Engineering Journal*, 23(1), 82–94.
<https://doi.org/10.31436/iiumej.V23i1.1783>
- Suciadi, J. (2001). Studi Analisis Metode-Metode Parsing Dan Interpretasi Semantik Pada Natural Language Processing. *Jurnal Informatika*, 2(1), 13–22.
<https://doi.org/10.9744/informatika.2.1.pp.13-22>
- Sutedi, A., Elsen, R., & Nasrulloh, M. R. (2021). Sundanese Stemming Using Syllable Pattern. *Jurnal Online Informatika*, 6(2), 218.
<https://doi.org/10.15575/join.V6i2.812>
- Taufik, M., Yuliana, Ri., Asih, I. Y. V., Kuswati, M., Rizhyalita, A., & Anggara, S. (2017). Bahasa Daerah Sebagai Mother Language Dalam Upaya Penguatan Kearifan Lokal Identitas Banten Di Kota Serang. *Jurnal Membaca Bahasa Dan Sastra Indonesia*, 2(1), 59.
- Wahyudi, D., Susyanto, T., & Nugroho, D. (2017). Implementasi Dan Analisis Algoritma Stemming Nazief & Adriani Dan Porter Pada Dokumen Berbahasa Indonesia. *Jurnal Ilmiah Sinus*, 15(2), 49–56.
<https://doi.org/10.30646/sinus.V15i2.305>