



ALGORITMA STEMMING TEKS BAHASA BATAK ANGKOLA BERBASIS ATURAN TATA BAHASA

TUGAS AKHIR

Disusun Sebagai Salah Satu Syarat
Untuk Memperoleh Gelar Sarjana Teknik
Pada Jurusan Teknik Informatika

Oleh

NUR HASANAH HRP

NIM. 11950121723



UIN SUSKA RIAU

FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU
PEKANBARU

2023

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LEMBAR PERSETUJUAN

**ALGORITMA *STEMMING* TEKS BAHASA BATAK ANGKOLA
BERBASIS ATURAN TATA BAHASA**

TUGAS AKHIR

Oleh

NUR HASANAH HRP

NIM. 11950121723

Telah diperiksa dan disetujui sebagai Laporan Tugas Akhir
di Pekanbaru, pada tanggal 14 Juli 2023

Pembimbing I,

Muhammad Fikry, ST, M.Sc
NIP. 19801018 200710 1 002

Pembimbing II,

Yusra, ST, M.T
NIP. 19840123 201503 2 001

**Hak Cipta Diindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LEMBAR PENGESAHAN**ALGORITMA STEMMING TEKS BAHASA BATAK ANGKOLA
BERBASIS ATURAN TATA BAHASA**

Oleh

NUR HASANAH HRP**NIM. 11950121723**

Telah dipertahankan di depan sidang dewan penguji
sebagai salah satu syarat untuk memperoleh gelar Sarjana Teknik
pada Universitas Islam Negeri Sultan Syarif Kasim Riau

Pekanbaru, 14 Juli 2023

Mengesahkan,
Ketua Jurusan,

Dr. Hartono, M.Pd
NIP. 19640301 199203 1 003

Iwan Iskandar, M.T
NIP. 19821216 201503 1 003

DEWAN PENGUJI

Ketua	: Jasril, S.Si, M.Sc
Pembimbing I	: Muhammad Fikry, ST, M.Sc
Pembimbing II	: Yusra, ST, MT
Penguji I	: Novriyanto, ST, M.Sc
Penguji II	: Muhammad Affandes, ST, MT



LEMBAR HAK ATAS KEKAYAAN INTELEKTUAL

Tugas Akhir yang tidak diterbitkan ini terdaftar dan tersedia di Perpustakaan Universitas Islam Negeri Sultan Syarif Kasim Riau adalah terbuka untuk umum dengan ketentuan bahwa hak cipta pada penulis. Referensi kepustakaan diperkenankan dicatat, tetapi pengutipan atau ringkasan hanya dapat dilakukan seizin penulis dan harus disertai dengan kebiasaan ilmiah untuk menyebutkan sumbernya.

Penggandaan atau penerbitan sebagian atau seluruh Tugas Akhir ini harus memperoleh izin dari Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau. Perpustakaan yang meminjamkan Tugas Akhir ini untuk anggotanya diharapkan untuk mengisi nama, tanda peminjaman dan tanggal pinjam.

Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



LEMBAR PERNYATAAN

Dengan ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain kecuali yang secara tertulis diacukan dalam naskah ini dan disebutkan didalam daftar pustaka.

Pekanbaru, 18 Juli 2023

Yang membuat pernyataan,

NUR HASANAH HRP

NIM. 11950121723

UIN SUSKA RIAU

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



LEMBAR PERSEMBAHAN

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Alhamdulillah Rabbil'alamin

Dengan Rahmat Allah Yang Maha Pengasih Lagi Maha Penyayang. Atas kasih sayang Allah yang dilimpahkan kepada saya untuk dapat menyelesaikan Tugas Akhir ini.

Shalawat beserta salam tak lupa kita hadiahkan buat junjungan alam yakni Nabi besar Muhammad SAW yang mana membebaskan manusia dari zaman kegelapan jahiliyah menuju zaman yang terang benderang.

Saya ucapkan terimakasih dan Saya persembahkan Tugas Akhir ini kepada kedua Orang Tua, Adik-adikku, Dosen, dan teman-teman yang telah mendukung dan membantu dalam pengerjaan Tugas Akhir.

Semoga Tugas Akhir ini bermanfaat bagi pembacanya dan untuk adik-adikku, semoga karya ini bisa menjadi pedoman dan semangat dalam menyelesaikan pendidikan.

Aamiin Allahuma Aamiin

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

SURAT PERNYATAAN

Saya yang bertandatangan di bawah ini:

Nama : Nur Hasanah Hrp
 NIM : 11950121723
 Tempat/Tgl. Lahir : Sipirok Baru, 26 Mei 2001
 Fakultas : Sains dan Teknologi
 Prodi : Teknik Informatika
 Judul Skripsi : Algoritma *Stemming* Teks Bahasa Batak Angkola
 Berbasis Aturan Tata Bahasa

Menyatakan dengan sebenar-benarnya bahwa:

1. Penulisan Skripsi dengan judul sebagaimana tersebut diatas adalah hasil pemikiran dan penelitian saya sendiri.
2. Semua kutipan pada karya tulis saya ini sudah disebutkan sumbernya.
3. Oleh karena itu, Skripsi saya ini, saya nyatakan bebas dari plagiat
4. Apabila dikemudian hari terbukti terdapat plagiat dalam penulisan Skripsi saya tersebut, maka saya bersedia menerima sanksi sesuai perundang-undangan.

Demikianlah Surat Pernyataan ini saya buat dengan penuh kesadaran dan tanpa paksaan dari pihak manapun juga.

Pekanbaru, 17 Juli 2023

Yang membuat pernyataan



NUR HASANAH HRP

NIM. 11950121723



Algoritma Stemming Teks Bahasa Batak Angkola Berbasis Aturan Tata Bahasa

Nur Hasanah Hrp, Muhammad Fikry*, Yusra

Fakultas Sains dan Teknologi, Teknik Informatika, Universitas Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

Email: ¹11950121723@students.uin-suska.ac.id, ^{2,*}muhammad.fikry@uin-suska.ac.id, ³yusra@uin-suska.ac.id

Email Penulis Korespondensi: muhammad.fikry@uin-suska.ac.id

Bahasa Batak Angkola merupakan salah satu ragam bahasa batak tepatnya di daerah Tapanuli bagian Selatan yang hingga kini masih dipakai dan dipelihara sebagai bahasa sehari-hari. Sumber daya bahasa Batak Angkola hingga saat ini belum tersedia dalam bentuk digital yang dapat dipergunakan oleh peneliti dalam tahapan analisis pemrosesan bahasa alami manusia. Pemrosesan bahasa alami (NLP Taks) untuk bahasa Batak Angkola harus mengikuti tahapan-tahapan pemrosesan teks dimulai dari tokenisasi, analisis leksikal, sintaksis, semantik, dan fragmatis. Penelitian ini melakukan pemrosesan bahasa alami pada tahapan pertama, yaitu analisis leksikal. Pada tahap analisis leksikal salah satu taks NLP yang penting adalah stemming. Stemming adalah proses menentukan kata dasar dari kata berimbuhan. Pada penelitian ini telah dilakukan analisis dan perancangan algoritma stemming Batak Angkola berdasarkan aturan tata bahasa. Tahapan pada penelitian ini yaitu dimulai dari mengumpulkan aturan-aturan tata bahasa Batak Angkola, mengumpulkan kata dasar berbahasa Batak Angkola sebagai kamus database, dan menghilangkan imbuhan dari akar kata. Keluaran penelitian ini berupa stemmer bahasa Batak Angkola dalam bentuk PHP. Berdasarkan pengujian yang dilakukan terhadap 450 kata yang berasal dari cerita rakyat Batak Angkola, diperoleh 448 kata uji yang benar (99,56%) dan 2 kata uji yang salah (0,44%). Kata uji yang salah didapatkan karena kata dasartidak ditemukan pada kamus.

Kata Kunci: Bahasa Batak Angkola; Kata Dasar; Aturan Tata Bahasa; Pemrosesan Bahasa Alami; Stemming

Abstract–The Angkola Batak language is a variety of Batak languages, to be precise in the southern Tapanuli area, which is still used and maintained as an everyday language. Until now, the resources of the Angkola Batak language are not yet available in digital form that can be used by researchers in the analytical stages of human natural language processing. Natural language processing (NLP Taks) for the Angkola Batak language must follow the stages of text processing starting from tokenization, lexical analysis, syntax, semantics, and pragmatics. This study conducted natural language processing in the first stage, namely lexical analysis. At the lexical analysis stage, one of the most important NLP tasks is stemming. Stemming is the process of determining root words from affixed words. In this research, an analysis and design of the Angkola Batak stemming algorithm have been carried out based on grammar rules. The stages in this research are starting from collecting the grammar rules of the Angkola Batak language, collecting basic words in the Angkola Batak language as a database dictionary, and removing affixes from root words. The output of this research is the stemmer of the Angkola Batak language in the form of PHP. Based on tests conducted on 450 words originating from the Batak Angkola folklore, 448 test words were correct (99.56%) and 2 test words were wrong (0.44%). The wrong test word is obtained because the root word is not found in the dictionary.

Keywords: Angkola Batak Language; Basic Word; Grammar Rules; Natural Language Processing; Stemming

1. PENDAHULUAN

Bahasa daerah dituturkan oleh penduduk setempat di satu bagian suatu negara sebagai bahasa komunikasi sehari-hari. Selain itu, bahasa daerah merupakan sebuah ciri khas suku bangsa di suatu daerah yang menjadi salah satu pembeda dengan suku bangsa di wilayah lain [1]. Indonesia memiliki lebih dari 700 bahasa daerah, dengan beragam kosakata, penutur, dialek, wilayah, dan konteks budaya. Bahasa daerah merupakan komponen penting warisan budaya Indonesia yang harus dipelihara, dilestarikan, dan dipertahankan untuk mencegah kepunahannya [2].

Suku Batak merupakan mayoritas penduduk Sumatera Utara. Terdapat enam kategori Suku Batak, yaitu Batak Toba, Batak Karo, Batak Pakpak Dairi, Batak Simalungun, Batak Mandailing, dan Batak Angkola [3]. Setiap suku memiliki ciri khas nama marga yang berfungsi sebagai simbol persaudaraan, dan setiap suku dapat dibedakan satu sama lain dengan bahasa daerah yang mereka gunakan. Bahasa Batak Angkola adalah salah satu ragam bahasa batak tepatnya di daerah Tapanuli bagian Selatan yang hingga kini masih dituturkan dan dilestarikan dalam kehidupan sehari-hari [4].

Pemrosesan bahasa alami (NLP Taks) untuk bahasa Batak Angkola harus mengikuti tahapan-tahapan pemrosesan teks dimulai dari tokenisasi, analisis leksikal, sintaksis, semantik, dan fragmatis [5]. Pada tahap analisis leksikal salah satu taks NLP yang penting adalah stemming. Stemming adalah mengubah kata-kata dalam teks menjadi kata dasar dengan menerapkan aturan tata bahasa (morfologi). Contohnya kata bersama, kebersamaan, menyamai, akan di stem ke kata dasarnya yaitu “sama”. Proses stemming untuk bahasa Indonesia dengan stemming teks bahasa Inggris memiliki perbedaan. Proses stemming teks bahasa Inggris yang dilakukan hanya penghapusan akhiran, sedangkan proses stemming bahasa Indonesia semua imbuhan dihapus yaitu akhiran, awalan, sisipan, dan perulangan [6].

Sayangnya, di perkembangan jaman dan teknologi saat ini, belum tersedia sumber daya bahasa (language resource) Batak Angkola yang dapat dimanfaatkan oleh peneliti dan pengembang perangkat lunak, khususnya di bidang Natural Language Processing (NLP), untuk mempermudah mempelajari bahasa Batak Angkola. Tahapan pemrosesan teks dapat dilihat seperti Gambar 1 dibawah ini.

2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

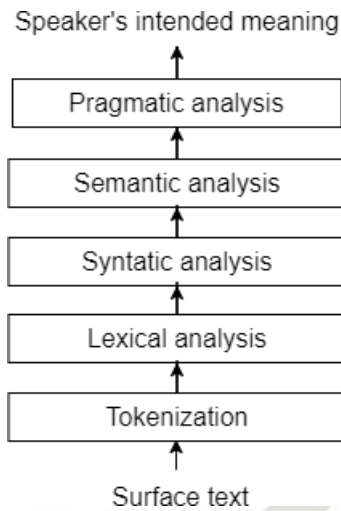
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

1. Dilarang mengutip, menyalin, atau memperbanyak sebagian atau seluruh karya tulis ini untuk tujuan komersial atau untuk tujuan lain tanpa izin UIN Suska Riau.

Hak cipta ini dilindungi oleh Undang-Undang. Untuk lebih jelasnya, silakan kunjungi laman www.uin-suska.ac.id atau hubungi kami di nomor telepon 081-7500-0000.

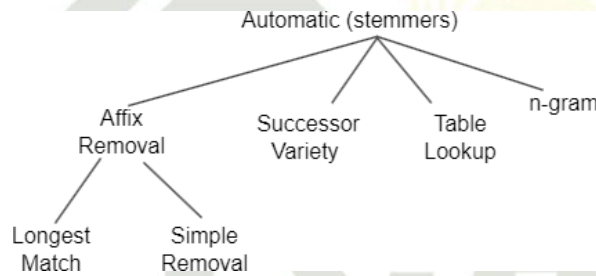
Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumbernya.
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mempublikasikan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Gambar 1. Tahapan analisis dalam pemrosesan teks [5]

Proses stemming disetiap bahasa berbeda-beda dan bergantung dari aturan morfologi yang ada pada bahasa. Maka dari itu untuk memperoleh kata dasar dari kata yang berimbuhan dibutuhkan algoritma yang berbeda sesuai aturan tata bahasanya (morfologi) dari bahasa itu sendiri [7]. Gambar 2 dibawah ini terdapat empat pendekatan yang dapat digunakan oleh stemmer yaitu Affix Removal, Successor Variety, Table Lookup, dan n-gram. Pendekatan yang sering dipakai adalah penghapusan imbuhan (Affix Removal) berdasarkan aturan tata bahasa, dan memakai kamus kata dasar (Table Lookup) sebagai acuan. Pendekatan stemmer diperlihatkan pada Gambar 2 dibawah ini.



Gambar 2. Pendekatan Stemmer [8]

Penelitian terdahulu seperti yang dilakukan Prastyo [9] dengan judul Algoritma stemming teks Bahasa Massenrempulu Berbasis Aturan Tata Bahasa. Bahasa Massenrempulu adalah bahasa daerah di Provinsi Sulawesi Selatan. Penelitian ini menggunakan algoritma stemming, berdasarkan pengujian yang dilakukan terhadap 530 kata uji diperoleh 526 kata uji yang benar (99,25%) dan 4 kata uji yang salah (0,76%). Karena kata yang diperoleh tidak sesuai dengan morfologi bahasa Massenrempulu, maka diperoleh kata uji yang salah. Penelitian lain oleh Sabali [10] dengan judul Pengembangan Metode Stemmer untuk Bahasa Bali dengan Pendekatan Rule-Based dan N-Gram Stemming. Penelitian ini bertujuan untuk mengembangkan metode stemmer yang meluluhkan seluruh variasi afiks pada bahasa Bali dengan mengombinasikan pendekatan Rule-Based dan metode N-Gram Stemming. Untuk 10 query metode yang telah diusulkan memperoleh akurasi stemming lebih baik yaitu sebesar 85,83% dari metode terdahulu 56,67%, Sedangkan untuk pengujian 5 dari 15 query menggunakan metode N-Gramstemming dapat mengidentifikasi kata-kata yang berafiks diluar rules.

Penelitian lain dengan judul Stemmer Bahasa Melayu Riau Berdasarkan Aturan Morfologi. Penelitian ini menggunakan algoritma stemming untuk bahasa daerah Melayu Riau dalam bentuk PHP Extension dan memperoleh akurasi 100% yang memperlihatkan bahwa stemmer dapat digunakan untuk pemrosesan teks berbahasa Melayu Riau [11].

Penelitian terkait lainnya mengenai stemming pada bahasa daerah lain di Indonesia diantaranya berbentuk aplikasi web (stemming teks Bahasa Karo [12], Simalungun [13], Pakpak Dairi [14], Lampung [15], Jawa [16], Sasak [17], Bali [18]), aplikasi android (Minang [19]), Pengaruh Stemmer Bahasa Indonesia terhadap Performa Analisis Semantimen Terjemahan Ulasan Film [20], Ekstraksi Fi'il dan Isim Pada Kaidah Nahwu Shorof Berbasis Android [21] dan Analisis Pengembangan Algoritma Porter Stemming dalam Bahasa Indonesia [22].

Berdasarkan pencarian belum ada penelitian tentang stemming Bahasa Batak Angkola. Akibatnya pemrosesan teks Bahasa Batak Angkola belum dapat dilakukan ke tahap selanjutnya, baik itu tahapan analisis leksikal (seperti normalisasi teks), tahapan analisis sintaksis (seperti Pos Parsing), tahapan analisis semantic (seperti translate). Jadi tujuan penelitian ini adalah menyediakan sumber daya bahasa (language resource) dalam

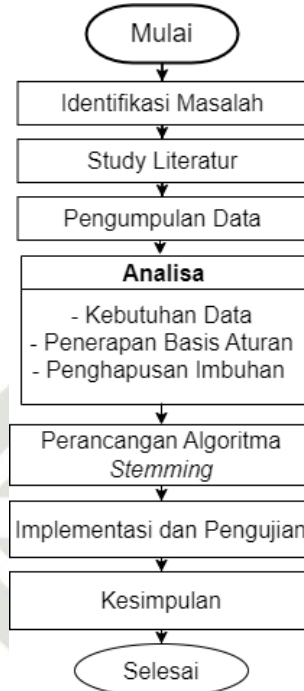


bentuk stemmer dan algoritma. Pendekatan yang dilakukan yaitu penghapusan imbuhan berdasarkan aturan morfologi yang mengacu pada kamus bahasa. Hasil dari stemming ini kemudian dapat digunakan pada tahap komposisi teks selanjutnya.

2. METODOLOGI PENELITIAN

1. Tahapan Penelitian

Tahap-tahap yang dilakukan selama proses penelitian untuk mencapai tujuan yang diharapkan dikenal dengan metode penelitian. Gambar 3 di bawah ini merupakan metode penelitian.



Gambar 3. Metode Penelitian

Tahap penelitian yang dilakukan:

Identifikasi masalah untuk memahami dan mencari solusi dari permasalahan yaitu bagaimana membangun algoritma stemming Bahasa Batak Angkola berbasis aturan tata bahasa.

Study Literatur adalah tahap dimana teori-teori tentang stemming dikaji. Teori-teori ini dapat dikumpulkan dari berbagai sumber seperti jurnal-jurnal terkait, e-book dan sumber pelengkap lainnya.

Pengumpulan data, data yang digunakan pada penelitian berupa kamus dasar bahasa Batak Angkola yang diterbitkan oleh Balai Bahasa Provinsi Sumatera Utara tahun 2021 [23]. Jumlah kata yang berhasil dikumpulkan sebanyak 3608 kata. Selanjutnya data yang dikumpulkan berupa aturan pembentukan kata (morfologi) dari buku Bahasa Bantu Batak Angkola (2021) [24]. Untuk data pengujian dipersiapkan sebanyak 450 kata berimbuhan yang diperoleh dari cerita rakyat berjudul Pabagas Boru (2017) [25]. Pemeriksaan dan koreksi dilakukan secara manual oleh ahli bahasa/validator (Tarmizi Tahir Pohan, S.Pd) untuk memperoleh masukan.

Tabel 1 berikut merupakan contoh sebagian kata berimbuhan yang digunakan sebagai data uji.

Tabel 1. Contoh kata berimbuhan dalam bahasa Batak Angkola

Bahasa Batak Angkola	Kata Dasar	Arti Dalam Bahasa Indonesia
diboto	boto	tahu
dibuat	buat	ambil
taligi	ligi	lihat
tonunan	tonun	tenun
mangupa	upa	hadiah
manjorat	jorat	jerat
manaek	naek	naik
marsuo	suo	jumpa
songoni	songon	seperti
mangajari	ajar	ajar

1. Diarahkan mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



- Hak Cipta Dilindungi Undang-Undang**
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mengacantumkan dan menyebutkan sumbernya.
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Bahasa Batak Angkola	Kata Dasar	Arti Dalam Bahasa Indonesia
harajahon	raja	raja
marlojong	lojong	lari
markehean	kehe	pergi
mardongan	dongan	teman
hatobangon	tobang	tua
manopotkon	topot	datang
dipamatang	pamatang	badan
pardomuan	domu	temu
parburangiran	burangir	sirih

Analisa, pada tahap Analisa, imbuhan dihilangkan sesuai dengan aturan tata bahasa (morfologi) untuk mendapatkan kata dasarnya. Aturan morfologi bahasa Batak Angkola diambil dari buku Bahasa Bantu Batak Angkola (2021) [24] meliputi:

- a. Penggunaan awalan (prefiks) ma- (termasuk maN-, mam-, mang-, manga-, mangka-, many-), mar- (termasuk marpa-, marpar-, mampar-, marsi-, matu), pa- (termasuk paN-, pam-, panga-, pany-, par-, parpa-), tar- (termasuk tarpa-, tarpan-, tarpam-, tarpang-), di-, ni-, sa-, saN-, um-, ha-.
- b. Penggunaan akhiran (sufiks) -i, -an, -on, -hon, -kon, -ma, -tu, dan -na.
- c. Penggunaan sisipan (infiks) -in-, -um-, dan -na-.
- d. Penggunaan gabungan (konfiks) mar-...-hon, mar-...-an, ma-...-tu, mang-...-kon, maN-...-ma, maN-...-i, maN-...-kon, paN-...-an, par-...-an, pa-...-hon, sa-...-an, sa-...-na, ha-...-an, ha-...-na, dan ha-...-na.

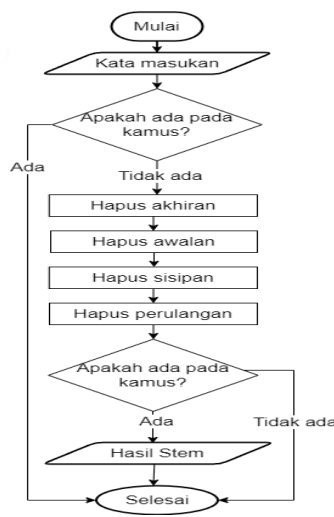
Perancangan algoritma merupakan perancangan algoritma stemming sesuai hasil analisa. Tahap perancangan menggunakan flowchart dan pseudocode. Flowchart adalah diagram alir yang merupakan diagram yang digambarkan menggunakan symbol-simbol guna merepresentasikan algoritma atau langkah-langkah yang berurutan dari sistem yang dibangun [26]. Pseudocode (kode semu) merupakan deskripsi singkat dari algoritma pemrograman komputer dengan struktur sederhana ditujukan untuk dibaca manusia bukan mesin. Implementasi dan Pengujian. Penerapan hasil algoritma dari langkah sebelumnya dikenal sebagai implementasi. Pada tahap pengujian, validator menentukan kata dasar dari kata uji. Pengujian tingkat akurasi diketahui dengan membandingkan kata dasar yang diperoleh validator terhadap kata dasar hasil stemmer.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Penelitian

Hasil buah flowchart dan algoritma stemming teks bahasa Batak Angkola dibuat dalam perancangan. Terdapat empat pendekatan yang dapat digunakan oleh stemmer yaitu Affix Removal, Successor Variety, Table Lookup, dan N-gram. Pada penelitian ini pendekatan Affix Removal digunakan untuk membuat Stemmer bahasa Batak Angkola, yang memerlukan kamus kata dasar (Table Lookup) sebagai acuan. Urutan penghapusan imbuhan yang digunakan yaitu penghapusan akhiran, penghapusan awalan, penghapusan sisipan, dan di akhiri penghapusan perulangan.

Flowchart merupakan alur atau proses kerja dari algoritma yang dibangun. Flowchart algoritma pada gambar 4. Diperlihatkan proses penghapusan imbuhan (akhiran, awalan, sisipan, perulangan), dan pengecekan pada kamus bahasa.



Gambar 4. Flowchart Algoritma



© Hak cipta milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim Riau

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mempublikasikan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Proses pertama kali yang dilakukan adalah dengan menginputkan kata masukan ke sistem, kemudian dilakukan pengecekan apakah kata inputan tersebut merupakan kata dasar dalam kamus bahasa, apabila tidak dan pengecekan imbuhan dilakukan secara berurutan, dimana jika hasil penghapusan suatu imbuhan tidak terdapat pada kamus bahasa, maka dilanjutkan ke penghapusan imbuhan berikutnya. Hasil akhirnya algoritma akan menampilkan hasil stem berupa kata dasar dan keterangan dalam bahasa Indonesia. Sebaliknya jika masukan ke sistem tidak ditemukan pada kamus bahasa, algoritma akan mengembalikan kata masukan dengan keterangan bahwa kata tersebut bukan kata dasar di kamus. Sebagian pseudocode algoritma diperlihatkan sebagai berikut.

Tabel 2. Pseudocode Stem

```

Function Stem(kata)
Input : Kata Berimbuhan
If cek kamus (kata)= true then
Result(kata)
Else if hapus akhiran (kata) then
If cek kamus (kata) = true then
Result (kata)
End if
Else if hapus awalan (kata) then
If cek kamus (kata) = true then
Result (kata)
End if
Else if hapus sisipan (kata) then
If cek kamus (kata) = true then
Result (kata)
End if
Else if hapus perulangan (kata) then
If cek kamus (kata) = true then
Result (kata)
End if
Return kata
End function Stem
    
```

Implementasi

Proses pembuatan sistem berdasarkan hasil analisa dan perancangan sistem pada tahap sebelumnya dikenal sebagai implementasi. Hal ini dilakukan untuk memastikan bahwa sistem dapat bekerja dan mencapai tujuan yang diinginkan. Implementasi stemmer menggunakan bahasa pemrograman berupa PHP. Berikut beberapa tampilan (interface) antarmuka yang diimplementasikan.

Tampilan halaman beranda Saat programnya dijalankan, halaman beranda merupakan halaman awal yang pertama kali ditampilkan. Setelah dijalankan, halaman awal akan menampilkan beberapa menu yaitu kamus, stemmer, dan pengujian. Adapun tampilannya seperti Gambar 5 berikut.



Gambar 5. Tampilan halaman beranda

2. Tampilan menu kamus Pada menu kamus menampilkan entrian kata dasar bahasa Batak Angkola yang berurutan sesuai abjad dan artinya dalam bahasa Indonesia. Jumlah kata dasar yang diperoleh dari kamus dasar bahasa Batak Angkola yang diterbitkan oleh Balai Bahasa Provinsi Sumatera Utara tahun 2021 sebanyak 3608 kata [23]. Pada halaman ini, untuk mencari kata dalam bahasa batak angkola dengan mengetikkan kata tersebut pada menu reserch, secara otomatis sistem akan menampilkan urutan kata yang dicari pada kamus beserta artinya dalam bahasa Indonesia. Tampilan menu kamus dapat dilihat Gambar 6 sebagai berikut.



Hak Cipta Dilindungi Undang-undang

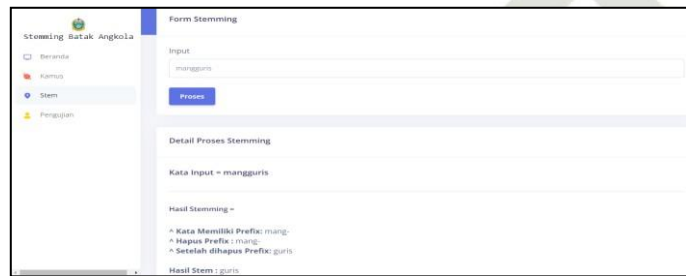
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mempublikasikan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Gambar 6. Tampilan menu kamus

Tampilan menu stemmer

Menu stemmer merupakan halaman untuk pemrosesan kata uji, kata uji yang dimasukkan akan diproses dan dicek pada kamus, sehingga menghasilkan kata dasar. Gambar 7 merupakan tampilan menu stemmer.



Gambar 7. Tampilan menu stemmer

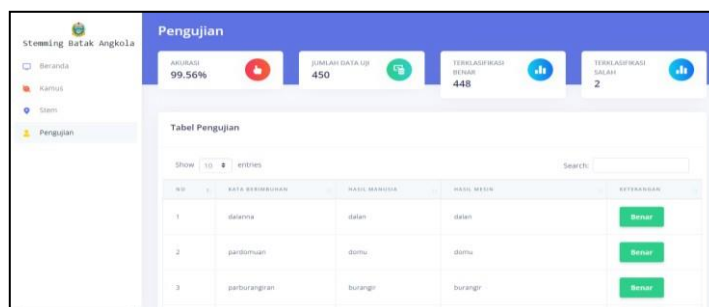
Proses kerja pada menu stemmer yaitu pertama jika kata yang diinputkan kata dasar. Kemudian kata dasar tersebut akan dilakukan pengecekan apakah kata tersebut sudah ada dalam kamus. Proses stemming tidak akan digunakan jika kata dasar yang dimasukkan adalah kata yang sudah ada dalam kamus. Contoh kata dasar bahasa Batak Angkola yang diinputkan adalah kehe yang artinya dalam bahasa Indonesia pergi. Maka kata kehe tidak akan dilakukan proses stemming karena kata kehe tidak mengandung imbuhan sama sekali dan merupakan kata dasar.

Kemudian yang kedua, jika kata yang diinputkan bukan kata dasar tetapi kata yang mengandung imbuhan (akhiran, awalan, sisipan, dan perulangan), proses pertama kali dilakukan yaitu pengecekan pada kamus apakah kata tersebut merupakan kata dasar yang ada pada kamus. Setelah dilakukan pengecekan maka dihasilkan kata tersebut dinyatakan tidak ada pada kamus dan bernilai false. Kemudian dilanjutkan ke proses pengecekan imbuhan yang dilakukan secara berurut, dimulai dari pengecekan imbuhan akhiran, setelah dilakukan pengecekan dan ditemukan mengandung imbuhan maka imbuhan yang melekat akan dihapus. Selanjutnya hasil penghapusan imbuhan akhiran akan dicek kembali pada kamus kata dasar. Jika hasil stemming ditemukan pada kamus kata dasar maka proses dihentikan. Contohnya tonunan, kata tonunan merupakan kata dasar tonun yang memiliki arti tenun, dan mengandung imbuhan akhiran -an yang mengubah artinya menjadi menenun.

Pengecekan imbuhan dilakuan secara berurutan, dimana jika hasil penghapusan suatu imbuhan tidak ditemukan pada kamus bahasa, maka dilanjutkan ke penghapusan imbuhan berikutnya. Hasil akhirnya algoritma akan menampilkan hasil stem berupa kata dasar dan keterangan dalam bahasa indonesia. Sebaliknya jika masukan hasil stem tidak ditemukan pada kamus bahasa, algoritma akan mengembalikan kata masukan dengan keterangan tidak ada kata dasar di kamus.

4. Tampilan halaman pengujian

Halaman pengujian menampilkan entrian kata yang telah diuji, jumlah klasifikasi data uji yang benar dan salah beserta total akurasinya. Total akurasi didapatkan dengan membandingkan kata dasar yang diperoleh validator terhadap kata dasar hasil stemmer. Gambar 8 berikut ini merupakan tampilan halaman pengujian.



Gambar 8. Tampilan halaman Pengujuan



Pada tahap pengujian dilakukan dengan cara membandingkan kata dasar yang diperoleh validator terhadap kata dasar hasil stemmer. Kata yang diuji sebanyak 450 kata yang berasal dari cerita rakyat Batak Angkola, diperoleh 448 kata uji yang benar (99,56%) dan 2 kata uji yang salah (0,44%). Kata uji yang salah didapatkan karena kata dasar tidak ditemukan pada kamus.

4. KESIMPULAN

Dari penelitian ini, dihasilkan sumber daya bahasa berupa algoritma stemming bahasa Batak Angkola berdasarkan aturan tata bahasa (morfologi). Algoritma stemming yang dibangun setelah dilakukan pengujian menghasilkan akurasi sebesar 99,56%. Berdasarkan hasil akurasi yang diperoleh, algoritma stemming dapat digunakan untuk memproses teks berbahasa Batak Angkola. Penelitian ini dapat dilanjutkan ke tahap pemrosesan teks berikutnya, yaitu pengoreksian kata, POS tagging, dan alih bahasa.

REFERENCES

- [1] Asrif, "Pembinaan dan Pengembangan Bahasa Daerah dalam Memantapkan Kedudukan dan Fungsi Bahasa Indonesia," *Mabasas*, vol. 4, no. 1, hal. 11–23, 2010.
- [2] N. W. Putri, "Pergeseran Bahasa Daerah Lampung Pada Masyarakat Kota Bandar Lampung," *J. Penelit. Hum.*, vol. 19, no. 14, hal. 77–86, 2018.
- [3] A. F. A. Batubara dan M. L. Anggapuspa, "Perancangan Pop-Up Book Ilustrasi Etnis Batak sebagai Media Interaktif untuk Anak Usia 9-10 Tahun," *J. Barik*, vol. 2, no. 2, hal. 108–120, 2021.
- [4] T. H. Dongoran, J. Naibaho, P. Sihombing, M. Sinaga, dan R. Tampubolon, *Fonologi Bahasa Angkola*. Jakarta : Pusat Pembinaan dan Pengembangan Bahasa, 1997.
- [5] N. Indurkha dan F. J. Damerou, *Handbook of Natural Language Processing*, 2 ed. Chapman & Hall/CRC, 2010.
- [6] H. R. Pramudita, "Penerapan Algoritma Stemming Nazief & Adriani dan Similarity pada Penerimaan Judul Tesis," *J. Ilm. DAS1*, vol. 15, no. 04, hal. 15–19, 2014.
- [7] L. Agusta, "Perbandingan Algoritma Stemming Porter dengan Algoritma Nazief & Adriani untuk Stemming Dokumen Teks Bahasa Indonesia," *Konf. Nas. Sist. dan Inform.*, hal. 196–201, 2009.
- [8] W. B. Frankes dan R. Baeza-Yates, *Information Retrieval: Data Structures & Algorithms*, 1992 ed. Prentice-Hall, 1992.
- [9] W. Prasetyo, "Algoritma Stemming Teks Bahasa Massenrempulu Berbasis Aturan Tata Bahasa," Universitas Islam Negeri Sultan Syarif Kasim Riau, 2019.
- [10] Made Agus Putra Subali, "Pengembangan Metode Stemmer untuk Bahasa Bali dengan Pendekatan Rule-Based dan N-Gram Stemming," Institut Teknologi Sepuluh Nopembe, Surabaya, 2019.
- [11] Yusra, M. Fikry, dan Hendri, "Stemmer Bahasa Melayu Riau Berdasarkan Aturan Morfologi," *Semin. Nas. Teknol. Informasi, Komun. dan Ind.* 13, no. November, hal. 118–124, 2021.
- [12] S. Megi, "Algoritma Stemming Teks Bahasa Karo Berdasarkan Aturan TataBahasa," Universitas Islam Negeri Sultan Syarif Kasim Riau, 2021.
- [13] F. Alfajri, "Algoritma Stemming Tekas Bahasa Batak Simalungun Baerbasis Aturan," Universitas Islam Negeri Sultan Syarif Kasim Riau, 2020.
- [14] W. Anisah, "Algoritma Stemming Bahasa Pakpak Dairi Menggunakan Aturan Tata Bahasa," Universitas Islam Negeri Sultan Syarif Kasim Riau, 2022.
- [15] Z. Abidin, A. Wijaya, dan D. Pasha, "Aplikasi Stemming Kata Bahasa Lampung Dialek Api Menggunakan Pendekatan Brute-Force dan Pemograman C#," *J. Media Inform. Budidarma*, vol. 5, no. 1, hal. 1, 2021, doi: 10.30865/mib.v5i1.2483.
- [16] M. Fauziyah, "Stemming Bahasa Jawa Menggunakan Algoritma Levenshtein dan Analisa Morfologi," Universitas Islam Negeri Maulana Malik Ibrahim, 2019.
- [17] Y. F. Andriani, E. Utami, dan S. Raharjo, "Modifikasi Algoritma Porter Stemmer Untuk Stemming Bahasa Sasak," *J. Inf. J. Penelit. dan Pengabd. Masy.*, vol. 5, no. 3, 2019, doi: 10.46808/informa.v5i3.147.
- [18] P. G. S. C. Nugraha dan N. W. Wardani, "Stemming Dokumen Teks Bahasa Bali Dengan Metode Rule Base Approach," *J. Tek. Inform. dan Sist. Inf.*, vol. 7, no. 3, hal. 510–521, 2020.
- [19] N. Sari dan K. Ummi, "Perancangan Aplikasi kamus Bahasa Minang Indonesia Dan Indonesia Minang Menggunakan Algoritma Levenshtein," *J. FTIK*, vol. 1, no. 1, hal. 1113–1124, 2020.
- [20] I. M. A. Agastya, "Pengaruh Stemmer Bahasa Indonesia Terhadap Performa Analisis Sentimen Terjemahan Ulasan Film," *TEKNOKOMPAK*, vol. 12, no. 1, hal. 18–23, 2018.
- [21] N. Himawan, G. W. Wicaksono, dan I. Nuryasin, "Ekstraksi Fi'il dan Isim Pada Kaidah Nahwu Shorof Berbasis Android," *J. Repos.*, vol. 2, no. 5, hal. 619–626, 2020, doi: 10.22219/repositor.v2i5.110.
- [22] A. A. Magriyanti, "Analisis Pengembangan Algoritma Porter Stemming dalam Bahasa Indonesia," 2018.
- [23] Maryanto, A. Hutasuhut, C. Nasution, Sriasrianti, Z. Hidayat, dan H. Al Banna, *Kamus Angkola Mandailing-Indonesia. Medan-Sumatera Utara: Balai Bahasa Provinsi Sumatera Utara, Badan Pengembangan dan Pembinaan Bahasa*, 2021.
- [24] M. P. Lusa Septia Dewi Br. Ginting, S.Pd., *Bahasa Bantu Batak Angkola*. Medan: Guepedia, 2021.
- [25] G. S. Bumi dan P. Ritonga, *Pelajaran Adat Tapanuli Selatan Pabagas Boru*. Marindal-Medan: Pertama Mitra Sari, 2017.
- [26] R. Rosaly dan M. K. Andy Prasetyo, ST., "Pengertian Flowchart Beserta Fungsi dan Simbol-simbol Flowchart yang Paling Umum Digunakan," <https://www.Nesabamedia.Com>, vol. 2, hal. 2, 2019, [Daring]. Tersedia pada: <https://www.nesabamedia.com/pengertian-flowchart>/<https://www.nesabamedia.com/pengertian-flowchart>

1. Dilarang mengutip, menyalin, atau menjiplak seluruh atau sebagian isi tanpa izin dari penerbit. Seluruh isi tanpa izin akan dicantumkan dan menyebabkan kerugian.

2. Dilarang mengutip dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

1. Dilarang mengutip, menyalin, atau menjiplak seluruh atau sebagian isi tanpa izin dari penerbit.

2. Dilarang mengutip dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.