

Penerapan Metode Rule Base Approach Untuk Stemming Teks Ensiklopedia Bahasa Sasak

Ramli Ahmad^{1*}, L M Samsu², Muhammad Saiful³, Imam Fathurrahman⁴

¹Program Studi Teknik Komputer, Universitas Hamzanwadi

^{2,3}Program Studi Sistem Informasi, Universitas Hamzanwadi

⁴Program Studi Informatika, Universitas Hamzanwadi

*ramliahmad@hamzanwadi.ac.id

Abstrak

Suku Sasak adalah salah satu suku Bangsa Indonesia yang mayoritasnya berada di Pulau Lombok, bahasa yang digunakan adalah Bahasa Sasak dengan pedoman tiga tingkatan sor-singgih (tigang soroh) yaitu Basa Kasar, Basa Mada dan Basa Alus. Bahasa Sasak juga memiliki imbuhan pangater, seselan dan pangiring. Untuk memudahkan pencarian kata dasar dalam Bahasa Sasak perlu dilakukan proses *stemming*. *Stemming* adalah proses pemetaan dan penguraian bentuk dari suatu kata menjadi bentuk dasarnya. Proses *stemming* sangat penting didalam proses *information retrieval system*. Pada penelitian ini, dalam melakukan proses *stemming* Bahasa Sasak menggunakan metode *Rule Base Approach*. Data yang digunakan dalam penelitian ini adalah kata dasar dalam Bahasa Sasak sebanyak 376 kata dasar. Penelitian ini bertujuan untuk merancang aplikasi *stemming* yang tepat untuk melakukan *stemming* Bahasa Sasak. Tahap awal dalam proses *stemming* Bahasa Sasak adalah melakukan proses *input, preprocessing, filtering, case folding* dan tokenisasi. Masing-masing kata dilakukan proses *stemming* untuk menghilangkan imbuhan pangater, seselan, dan pangiring. Hasil dari penelitian menyatakan bahwa metode *Rule Base Approach* dapat digunakan untuk melakukan *stemming* teks Bahasa Sasak, hal ini dapat dilihat dari hasil akurasi mencapai angka 77.82%. Tentunya dalam pengujian masih terdapat kegagalan yang disebabkan oleh kesalahan *overstemming* akibat dari proses *stemming*.

Kata kunci: Bahasa Sasak, Ensiklopedia, *Rule Base Approach*, *Stemming*

Abstract

The Lomboknese are one of the ethnic groups of Indonesia, the majority of which are on the island of Lombok, the language used is Lomboknese with three levels of sor-singgih (tigang soroh) guidelines, namely Basa Kasar, Basa Madya and Basa Alus. Lomboknese language also has the additions of pangater, seselan and pangiring. To facilitate the search for basic words in Lomboknese, a stemming process is needed. Stemming is the process of mapping and decomposing the form of a word into its basic form. The stemming process is very important in the information retrieval system process. In this study, the Lomboknese stemming process used the Rule Base Approach method. The data used in this study are 376 basic words in Lomboknese. This study aims to design an appropriate stemming application for Lomboknese stemming. The initial stage in the Lomboknese stemming process is to carry out the input process, preprocessing, filtering, case folding and tokenization. Each word is subjected to a stemming process to remove the additions of pangater, seselan, and pangiring. The results of the study indicate that the Rule Base Approach method can be used to stem Lomboknese texts, this can be seen from the results of the accuracy reaching 77.82%. Of course, in testing there are still failures caused by overstemming errors resulting from the stemming process.

Keyword: Lomboknese, Ensiklopedia, *Rule Base Approach*, *Stemming*

1. Pendahuluan

Suku Sasak adalah salah satu suku Bangsa Indonesia yang mayoritasnya berada di Pulau Lombok, bahasa yang digunakan adalah Bahasa Sasak dengan pedoman tiga tingkatan sor- singgih (tigang soroh) yaitu Basa Kasar, Basa Madia dan Basa Alus. Morfologi Bahasa Sasak memiliki keunikan tersendiri jika dibandingkan dengan Bahasa Indonesia atau Bahasa Negara lain dalam hal penyusunan pengucapan, struktur kalimat, dan tata bahasa.

Tata bahasa itu sendiri meliputi kata dasar, pangater (awalan), seselan (sisipan) dan pangiring (akhiran) dalam sebuah kalimat. Jika sebuah kata dasar dalam Bahasa Sasak sudah mendapatkan awalan, sisipan, akhiran atau kombinasinya maka kata tersebut disebut krana tiron (kata berimbuan). Krana tiron [1] merupakan Bahasa Sasak yang memiliki arti kata berimbuan yang terbentuk dari kata dasar yang mendapatkan awalan, sisipan dan akhiran. Kata dasar dalam Bahasa Sasak yang sudah ditambahkan pangater, sisipan dan pangiring pasti memiliki permasalahannya sendiri jika diterapkan pada *information retrieval*.

Information retrieval merupakan metode mencari material (dokumen), meliputi proses mencari dan mengambil informasi berbasis pengetahuan dari kumpulan dokumen [2].

Information retrieval digunakan untuk mencari dokumen teks yang dibutuhkan oleh pengguna, dokumen teks tersebut berupa dokumen yang berbentuk digital [3]. Dalam mesin pencari (*search engine*) atau *Information retrieval* pada umum akan menghasilkan informasi dokumen teks dalam jumlah yang besar, hal ini juga akan memberikan dampak pada boros waktu dalam melakukan pencarian dokumen teks. Hasil pencarian informasi menggunakan *search engine* akan menghasilkan dokumen teks pencarian yang efektif dan efisien jika didukung proses pembuatan akar kata atau kata dasar yang tepat.

Dari permasalahan tersebut maka diperlukan aplikasi yang dapat mencari kata dasar dalam Bahasa Sasak dengan memetakan pangater (awalan), seselan (sisipan) dan pangiring (akhiran). Semakin berkembangnya teknologi dan ilmu pengetahuan, untuk menentukan kata dasar dalam Bahasa Sasak dapat direkomendasikan menggunakan *Text Mining* dengan tahap *Stemming*. Menurut Feldman [4] *text mining* adalah sebuah proses pengetahuan intensif dimana pengguna berinteraksi dan bekerja dengan sekumpulan dokumen dengan menggunakan beberapa alat analisis sedangkan *stemming* adalah proses pemetaan dan menghilangkan bentuk dari suatu kata, sehingga kata tersebut menjadi bentuk



dasarnya. Salah satu metode dasar yang banyak digunakan adalah metoda *Rule Base Approach*. Metode *Rule Base Approach* adalah metode yang membentuk akar kata dengan proses pengambilan awalan, sisipan dan akhiran.

Beberapa peneliti telah melakukan penelitian menggunakan metode *Rule Based Approach* antara lain oleh *Guterres* [5] dengan judul “*Stemming Bahasa Tetun Menggunakan Pendekatan Rule Based*”, dengan hasil proses *stemming* Bahasa Tetun dalam penelitian ini memiliki tahapan yaitu memeriksa konfiks, memeriksa gabungan awalan & akhiran yang dilarang, dan memeriksa sisipan secara urut. Hal utama yang menjadi penyebab kegagalan algoritma *stemming* Bahasa Tetun dalam proses *stemming* disebabkan karena kesalahan *overstemming*. Dalam penelitian ini algoritma *stemming* Bahasa Tetun mendapatkan tingkat akurasi sebesar 90.52%. Melihat hasil akurasi yang sangat baik dari metode *Rule Base Approach*, penulis kembali menggunakan metode tersebut untuk melakukan *stemming* pada Bahasa Sasak. Peneliti lainnya dilakukan oleh *Amin* [6], dengan judul “*Implementasi Stemmer Bahasa Jawa Dengan Metode Rule*

Base Approach Pada Sistem Temu Kembali Informasi Dokumen Teks Berbahasa Jawa”. Pada penelitian ini proses *stemming* pada dokumen teks Bahasa Jawa menggunakan metode *Rule Base Approach* dilakukan pada semua dokumen teks Bahasa Jawa. Pada penelitian ini hasil proses *stemming* Bahasa Jawa menggunakan metode *Rule Base Approach* mendapatkan tingkat akurasi yaitu sebesar 77%. Melihat hasil akurasi yang sangat baik dari metode *Rule Base Approach*, penulis kembali menggunakan metode tersebut untuk melakukan *stemming* pada Bahasa Sasak. Dari paparan penelitian sebelumnya, kontribusi ilmiah yang diterapkan yaitu penerapan metode *Rule Base Approach* untuk melakukan *stemming* pada Bahasa Sasak.

Berdasarkan uraian latar belakang penelitian yang sudah peneliti paparkan serta melihat penting peranan *stemming* pada *information retrieval* dan mengacu pada hasil penelitian sebelumnya, peneliti tertarik untuk melakukan penelitian terhadap *stemming* dokumen dengan objek teks Bahasa Sasak menggunakan metode *Rule Base Approach*. Peneliti berharap, dengan penelitian yang dilakukan, terciptanya aplikasi yang mampu mendapatkan kata dasar dari sebuah

masuk dokumen teks Bahasa Sasak.

2. Tinjauan Pustaka

Dalam penelitian ini kami melakukan berbagai macam tinjauan pustaka, diantaranya penelitian terkait dan landasan teori dimana kami jelaskan secara terperinci dibawah ini.

2.1 Penelitian Terkait

1. Penelitian Saiful M, Samsu LM, Fathurrahman I. [7] Perancangan Kerangka Crowdsourcing Berbasis Wisdom Of Crowds Untuk Kamus Naskah Lontar (Takepan) Sasak Online. Hasil penelitian bahwa dalam rancangan kerangka sistem Crowdsourcing kamus naskah Lontar (Takepan) Sasak online diperlukan adanya pengembangan dengan tujuan untuk memperbaharui dan mempermudah pengguna dalam mempelajari dan mengkaji serta menterjemahkan aksara sangsekerta melalui media online.
2. Penelitian Amin dan Alfa [6] Implementasi Stemmer Bahasa Jawa dengan Metode Rule Base Approach pada Sistem Temu Kembali Informasi Dokumen Teks Berbahasa Jawa. Hasil penelitian Proses stemmer menghasilkan kumpulan Teks kata dasar bahasa jawa. Hasil proses stemmer bahasa jawa menggunakan metode rule base approach

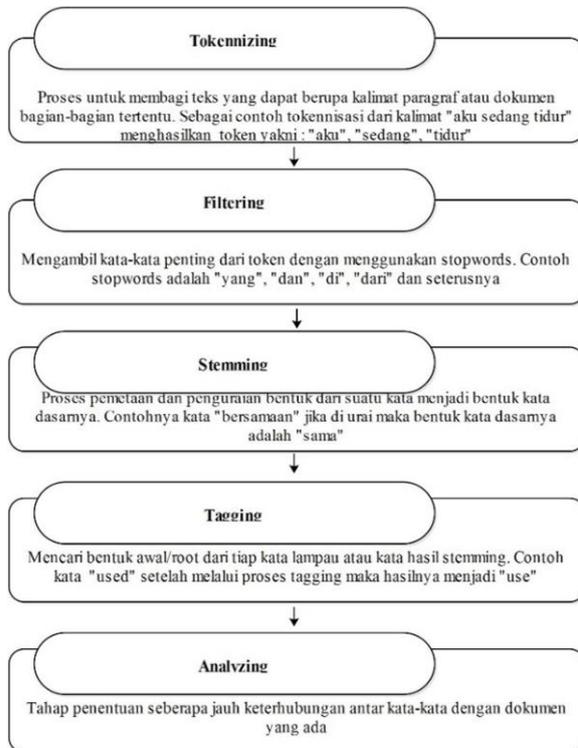
pada dokumen teks bahasa jawa yang telah di uji adalah memiliki akurasi 77%.

3. Penelitian Guterres [5] Stemming Bahasa Tetun Menggunakan Pendekatan Rule Based. Hasil penelitian bahwa analisis dilakukan berdasarkan kasus error stemming seperti overstemming, understemming, unchanged, dan spelling exception. Hasil uji coba yang didapatkan adalah algoritma stemming bahasa Tetun menghasilkan akurasi sebesar 90.52%.

2.2 . Landasan Teori

1. Text Mining

Menurut Feldman [4] *text mining* adalah sebuah proses pengetahuan intensif dimana pengguna berinteraksi dan bekerja dengan sekumpulan dokumen dengan menggunakan beberapa alat analisis. Tahapan proses *text mining* dibagi menjadi 4 tahap utama yaitu *Tokenizing, Filtering, Stemming, Tagging* dan *Analyzing*. Tahapan-tahapan dalam *text mining* bertujuan untuk menemukan kata-kata yang mewakili isi dokumen dan digambarkan pada gambar 1.



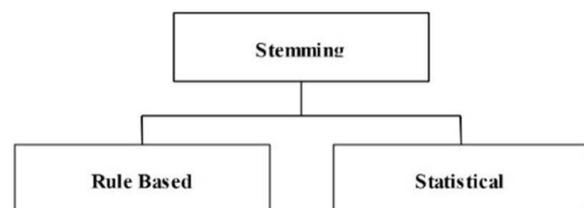
Gambar 1. Tahapan dalam *Text Mining*

2. Rule Base Stemmer

Stemming adalah proses untuk menemukan akar kata (*root*) atau kata dasar dengan memisahkan semua *affix* atau imbuhan yang melekat pada kata tersebut. *Affix* (imbuhan) bisa terdiri dari awalan (*prefix*), akhiran (*suffix*), sisipan (*infix*), dan gabungan awalan-akhiran (*confix*). Pada banyak bahasa, kata-kata biasanya dihasilkan dengan menambahkan imbuhan pada kata dasarnya (*root*). Hasil dari *stemming* adalah *stem* (akar kata) yang merupakan bagian kata yang tersisa setelah dihilangkan imbuhanannya [6].

Metode *Rule Base* adalah metode yang

membentuk akar kata dengan proses pengambilan awalan, sisipan dan akhiran. Metode *Rule Base* digunakan karena sifatnya yang fleksibel untuk digunakan sebagai *stemmer* berbagai macam bahasa dengan karakteristiknya yang lebih menekankan pada struktur morfologi suatu bahasa. Metode ini akan membuang (menghilangkan) *prefix* (awalan), *suffix* (akhiran) dan *infix* (sisipan) dari term bentukan menjadi suatu term kata dasar [6]. Algoritma *Stemming* bisa diklasifikasikan menjadi 2 (dua) yaitu *Rule Based* dan *Statistical*. Menurut Sharma (2012) *Stemmer* berbasis aturan mengkodekan aturan khusus bahasa dimana *stemmer* statistik menggunakan informasi statistik dari yang besar korpus bahasa tertentu untuk mempelajari morfologi. yang ditunjukkan pada gambar 2.



Gambar 2. *Types of Stemming Approach*

3. Bahasa Sasak

Bahasa Sasak *adalah* sebuah bahasa *Austronesia* dari cabang Sundik dan lebih spesifik dari anak cabang Bali-Sasak. Bahasa

ini terutama dipertuturkan di pulau Lombok, pulau Lombok bagian barat, dan sedikit di ujung timur pulau Bali. Di Lombok sendiri Bahasa Sasak memiliki tingkatan penggunaannya, misalnya ada yang disebut Lombok Alus, Lombok Madya dan Lombok Kasar.

Ada beberapa imbuhan dalam Bahasa Sasak, yaitu:

1. *Prefiks* yaitu imbuhan yang diletakkan di awal kata dasar, dalam Bahasa Sasak disebut dengan pangater terdiri dari: (*a-*, *ma-*, *su-*, *ka-*, *pa-*, *pati-*, *pari-*, *maka-*, *saka-*, *kuma-*, *sa-*, *pa-*, *pi-*, *dur-*, *swa-*). Adapun contoh katanya yaitu : aukud, majalan, sudharma, kaejuk, pajalan, patigrepe, makadadua, sakabesik, kumalipan.
2. *Infiks* yaitu penyisipan imbuhan di dalam kata dasar, dalam Bahasa Sasak disebut dengan Seselan terdiri dari: (*-in-*, *-um-*, *-el-*, *-er-*). Adapun contoh katanya : sinurat, sumaur, telapak, gerigi.
3. *Sufiks* yaitu imbuhan yang diletakkan di belakang kata dasar, dalam Bahasa Sasak disebut dengan pangiring terdiri dari: (*-a-*, *-e-*, *-ne-*, *-ang-*, *-in-*, *-an-*, *-n-*, *-wan-*, *-nyane-*). Adapun contoh katanya: alapa, bastise, memene, ngadepang, tulisin, menekan, bapan tiange, dharmawan

retrieval

Information retrieval merupakan

metode mencari material (dokumen), meliputi proses mencari dan mengambil informasi berbasis pengetahuan dari kumpulan dokumen [2]. *Information retrieval* digunakan untuk menemukan kembali dokumen teks yang dibutuhkan oleh pengguna, dokumen teks tersebut berupa dokumen yang berbentuk digital [3]. Konsep pengambilan informasi (IR) telah dikembangkan dalam kaitannya dengan sistem database bagi banyak orang tahun. Pengambilan informasi adalah asosiasi dan pengambilan informasi dari sejumlah besar dokumen berbasis teks. Pengambilan informasi dan sistem database, masing-masing menangani berbagai jenis data; beberapa masalah sistem database biasanya tidak hadir dalam sistem pencarian informasi, seperti kontrol konkurensi, pemulihan, transaksi manajemen, dan pembaruan. Juga, beberapa yang umum masalah pencarian informasi biasanya tidak ditemui dalam sistem database konvensional, seperti sebagai dokumen tidak terstruktur, perkiraan pencarian berdasarkan kata kunci, dan konsep relevansi. Karena sejumlah besar informasi teks, informasi retrieval telah menemukan banyak aplikasi. Terdapat banyak sistem temu kembali informasi, seperti on-line sistem katalog perpustakaan, dokumen on-line sistem manajemen, dan baru-baru ini mengembangkan mesin pencari

Web [8].

3. Metode Penelitian

Dalam penelitian ini kami membagi tahapan menjadi dua yaitu lokasi penelitian dan tahapan penelitian.

3.1. Lokasi Penelitian

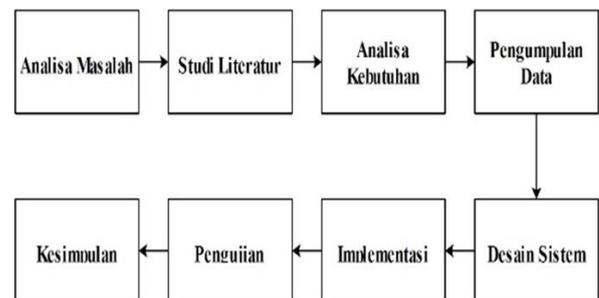
Penelitian ini berlokasi di Kabupaten Lombok Timur, Provinsi Nusa Tenggara Barat.

3.2. Tahapan Penelitian

Secara umum, penelitian ini menghasilkan aplikasi yang dapat digunakan untuk pembentukan akar kata dalam dokumen teks Bahasa Sasak. Analisa masalah penelitian ini dimulai dengan melakukan pengumpulan data dokumen teks Bahasa Sasak serta bagaimana pengembangan atau memodifikasi tabel aturan pemenggalan dari algoritma *Rule Base Approach*.

Seperti terlihat pada gambar 3, pada proses studi literatur akan dilakukan pencarian referensi pendukung yang berupa teori maupun jurnal yang memuat penelitian terkait permasalahan yang telah didefinisikan sebelumnya. Pada proses analisa kebutuhan dilakukan pengumpulan kebutuhan yang diperlukan untuk memperkuat data penelitian yang selanjutnya dilakukan pemenuhan

semua kebutuhan semua data tersebut pada proses pengumpulan data. Setelah semua data terpenuhi barulah dilakukan desain dan implementasi aplikasi tersebut. Untuk mengetahui kemampuan aplikasi dalam mengatasi permasalahan yang telah dikembangkan, maka dilakukanlah pengujian sistem. Proses paling akhir dari tahap penelitian ini adalah melakukan penarikan kesimpulan untuk penerapan solusi yang ditawarkan berdasarkan masalah yang telah didefinisikan sebelumnya.



Gambar 3. Rancangan Penelitian *Stemming* Bahasa Sasak

3.2.1 Data yang Digunakan

Data yang digunakan pada penelitian stemming Bahasa Sasak [9], berupa kumpulan teks kata dasar yang menggunakan Bahasa Sasak yang dapat diakses di <https://kansas94.blogspot.com/2017/02/takepan-sasak.html>. Arsitektur umum pada sistem pembentukan akar kata dalam dokumen teks Bahasa Sasak memiliki beberapa tahapan. Seperti terlihat pada

gambar 4, adapun tahapan yang ada dalam sistem sebagai berikut:

a. *Input*

Pada tahap pertama sistem adalah input dokumen teks Bahasa Sasak.

b. *Preprocessing*

Pada tahap *preprocessing* dilakukan proses perubahan bentuk teks yang belum terstruktur menjadi teks yang terstruktur, yang mana teks tersebut disesuaikan dengan kebutuhan aplikasi. Tahap yang dilakukan, yaitu:

- *Filtering*

Pada proses *filtering* dilakukan pengambilan karakter alfabet dari dokumen teks Bahasa Sasak. selanjutnya pada proses ini, dilakukan penghapusan karakter, tanda baca dan angka.

- *Case Folding*

Pada proses *case folding* dilakukan perubahan untuk semua huruf dalam dokumen teks Bahasa Sasak menjadi huruf kecil. Huruf yang dapat diterima yaitu huruf 'a' sampai dengan huruf 'z'.

- *Tokenisasi*

Proses ini bertujuan untuk memisahkan setiap kata yang menyusun dokumen menjadi kata per kata dengan menandai karakter spasi. Tiap kata hasil *tokenisasi* akan

disimpan ke dalam katagori.

c. *Proses*

Pada tahap ini, dilakukan proses *stemming* dengan metoda *Rule Base Approach*. Satu per satu kata akan dicek, kata yang terdapat imbuhan akan dilakukan pengambilan imbuhan seperti awalan, sisipan atau akhiran sehingga terbentuk akar kata. Pembentukan kata dalam Bahasa Sasak di paparkan pada struktur berikut:

[awalan]+[sisipan]+[dasar]+[akhiran]

Masing-masing imbuhan digabungkan dengan kata dasar sehingga membentuk kata berimbuhan baik yang berisi awalan, berisi sisipan, berisi akhiran maupun berisi semua model imbuhan. Pada *stemming* Bahasa Sasak terdapat beberapa proses, proses tersebut adalah sebagai berikut:

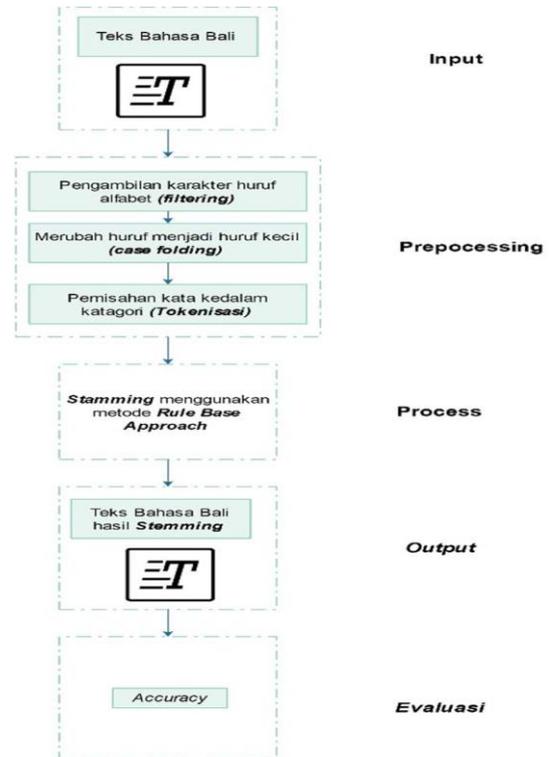
1. Menghilangkan awalan (awalan-4, awalan-3, awalan-2 dan awalan-1) Jika suatu aturan terpenuhi jalankan sebagai berikut: Hilangkan akhiran (akhiran-3, akhiran-2, akhiran-1), jika suatu aturan terpenuhi, hilangkan awalan, jika tidak proses *stemming* selesai.
2. Jika tidak ada aturan yang terpenuhi jalankan sebagai berikut: hilangkan awalan, hilangkan akhiran dan proses *stemming*.

d. *Output*

Setelah semua kata dalam dokumen

selesai di proses, maka hasil dari proses *stemming* akan ditampilkan dalam bentuk teks. Kata yang di *stemming* awalan, sisipan dan akhiran akan diberikan tanda dengan warna biru dan warna merah. Pada pengujian proses *stemming* menggunakan metode *Rule Base Approach*, dalam menghitung tingkat akurasi *stemming* dapat menggunakan persamaan (1). Pada persamaan (1), terdapat simbol x yang merupakan tingkat akurasi *stemming*. Simbol T merupakan jumlah kata yang di *stemming* dengan hasil benar. Simbol N merupakan jumlah kata yang di *stemming*.

$$x = \frac{T}{N} \times 100\% \quad (1)$$



Gambar 4. Gambaran Umum Aplikasi

4. Implementasi dan Pembahasan

1. Sistem Pencarian Kata Dasar

Data pada penelitian ini berupa kumpulan teks kata dasar yang menggunakan Bahasa Sasak. Pada gambar 5 menampilkan tampilan aplikasi *stemming*, yang mana pada aplikasi ini terdapat 2 kolom. Masing-masing kolom memiliki fungsi yaitu untuk menampilkan teks awal dan menampilkan teks yang telah dilakukan proses *stemming*. Masukan dalam aplikasi ini yaitu *load* kamus, *load* leburan, *load* text dan *load* sisipan. Pada aplikasi *stemming*, untuk tahap awalan aplikasi memberikan tanda warna

merah, pada tahap akhiran aplikasi memberikan tanda warna biru dan pada tahap sisipan aplikasi akan langsung memunculkan kata dasarnya.



Gambar 5. Tampilan Sistem Pencarian Kata Dasar

2. Pengujian Prefix (Pengater)

Pada tahap ini dilakukan pengujian penghapusan pengater dari teks Bahasa Sasak yang telah di masukan. Sebagai contoh “mabaju” bentuk dasarnya “baju”. Pengujian dilakukan dengan menggunakan teks kata dasar pengater sebanyak 240 kata dasar.

Tabel 1. Pengujian Pengater

Pengater	Teks	Hasil Stemming
be-	be-silo	besilo
pe-	pe-takak	penakak
te-	te-jagur	jagur
ke-	ke-solah	solah

3. Pengujian Sufiks (Pangiring)

Pada tahap ini dilakukan pengujian penghapusan pangiring dari teks Bahasa Sasak yang telah di inputkan. Sebagai contoh “jemakin” bentuk dasarnya “jemak”.

Pengujian dilakukan dengan menggunakan teks kata dasar pangiring sebanyak 119 kata dasar.

Tabel 2. Pengujian Pangiring

Pangiring	Teks	Hasil Stemming
- be - an	jemaka	jemak
- ing	carike	carik
- ne	bapanne	bapan
- ang	dengehang	dengah
- in	negakin	negak
- an	cenikan	cenik
- n	mejan	meja

4. Pengujian Infiks (Seselan)

Pada tahap ini dilakukan pengujian penghapusan seselan dari teks Bahasa Sasak yang telah di inputkan. Sebagai contoh “sinurat” bentuk dasarnya “surat”. Pengujian dilakukan dengan menggunakan teks kata dasar seselan sebanyak 17 kata dasar.

Tabel 3. Pengujian Seselan

Sisipan	Teks	Hasil Stemming
-in-	sinurat	surat
-um-	rumaksa	maksa
-el-	telapak	tapak
-er-	keresek	kesek

5. Hasil Pengujian

Setelah dilakukan pengujian *stemming* teks kata dasar pada pengater, seselan dan pangiring menggunakan aplikasi *stemming*



kata dasar Bahasa Sasak, dilihat hasil pengujian menghasilkan uji *stemming* benar pada pengater sebanyak 159 kata dan uji *stemming* salah pada pengater sebanyak 81 kata. Uji *stemming* benar pada seselan sebanyak 17 kata dan uji *stemming* salah pada seselan sebanyak 0 kata. Uji *stemming* benar pada pangiring sebanyak 80 kata dan uji *stemming* salah pada pangiring sebanyak 39 kata.

Tabel 4. Hasil Pengujian

Imbuhan	Jumlah Kata Dasar	Uji Stemming (Benar)	Uji Stemming (Salah)
Pengater	240	159	81
Seselan	17	17	0
Pangiring	119	80	39

6. Tingkat Akurasi

Tingkat akurasi aplikasi *stemming* kata dasar Bahasa Sasak dengan metode *Rule Base Approach*, menghasilkan tingkat akurasi pengater 66,25%, tingkat akurasi seselan 100% dan tingkat akurasi pangiring 67.22%. Hasil rata-rata tingkat akurasi yang diperoleh aplikasi *stemming* Bahasa Sasak sebesar 77.82%. Dari hasil tingkat akurasi yang diperoleh dapat disampaikan bahwa aplikasi *stemming* kata dasar Bahasa Sasak masih tergolong baik dalam melakukan proses *stemming* kata dasar Bahasa Sasak.

Jika mengacu pada hasil pengujian tepatnya dengan hasil uji *stemming* salah, masih terdapat banyak kesalahan pada proses *stemming* hal ini di karenakan Metode *Rule Base Approach* memiliki kelemahan yaitu jika diterapkan pada kata yang memiliki kompleksitas imbuhan yang tinggi dan aplikasi *stemming* tidak dapat mengenali *rule* maka akan terjadi kesalahan pada proses *stemming*. Selain itu penyebab kesalahan juga terjadi karena *overstemming*. *Overstemming* adalah sebuah permasalahan pada proses *stemming* dimana kata dasar dari hasil *stemming* mengalami kekurangan suku kata atau huruf yang dianggap sebagai imbuhan akibat dari proses *stemming*.

Dalam penelitian ini salah satu contoh permasalahan *overstemming* seperti proses pengiring “-ang” pada kata “sambatang” menjadi “mbat” padahal yang benar adalah “sambat”. Dari hasil tersebut dapat kita lihat ada huruf atau suku kata yang menjadi bagian kata dasar yang terhapus atau hilang. Yang menjadi penyebab kegagalan pada proses *stemming* adalah karena ada bagian huruf yang dianggap sebagai pengater oleh aplikasi yaitu “sa”, sehingga huruf atau suku kata tersebut dihapus.

Pada penelitian terdahulu metode *Rule*

Base Approach sudah digunakan untuk melakukan proses *stemming* pada Bahasa Tetun yang mana penelitian dilakukan oleh Guterres [5]. Jika dilakukan perbandingan hasil dari penelitian yang sudah dilakukan, dapat dilihat bahwa tingkat akurasi yang peneliti peroleh sebesar 77.82% lebih kecil jika dibandingkan dengan penelitian Guterres sebesar 90.52%, hal ini dikarenakan morfologi dari Bahasa Tetun sebagian besar masih menggunakan Bahasa Lain seperti Bahasa Indonesia dan Bahasa Portugis.

Pada penelitian terdahulu lainnya yaitu proses *stemming* pada Bahasa Jawa, yang mana penelitian dilakukan oleh Amin [6]. Jika dilakukan perbandingan hasil penelitian yang sudah dilakukan, dapat dilihat bahwa tingkat akurasi yang peneliti peroleh sebesar 77.82% sama jika dibandingkan dengan penelitian amin sebesar 77%, hal ini dikarenakan morfologi Bahasa Jawa dan Bahasa Sasak memiliki tingkat kesulitan yang hampir sama.

5. Kesimpulan

Berdasarkan proses penelitian yang sudah dilakukan maka peneliti dapat mengambil kesimpulan bahwa metode *Rule Base Approach* dapat digunakan untuk melakukan *stemming* teks Bahasa Sasak, hal ini dapat dilihat dari hasil tingkat akurasi mencapai angka 77.82%. Tentunya dalam

pengujian masih terdapat kegagalan yang disebabkan oleh kesalahan *overstemming* akibat dari proses *stemming*.

Daftar Pustaka

- [1] I. B. Purwa, I. N. Suandi, and I. N. Martha, "Representasi Nilai Kearifan Lokal Masyarakat Bali pada Lagu Pop Bali (Kajian Bentuk Lingual Bahasa)," *J. Pendidik. Bhs. dan Sastra Indones.*, vol. 14, pp. 230–245, 2024, doi: <https://doi.org/10.23887/jpbsi.v14i2.86033>.
- [2] A. Roshdi and A. Roohparvar, "Review: Information Retrieval Techniques and Applications," in *International Journal of Computer Networks and Communications Security*, vol. 3, no. 9, 2015, pp. 373–377. [Online]. Available: www.ijcnscs.org
- [3] L. Afuan, "Stemming Dokumen Teks Bahasa Indonesia Menggunakan Algoritma Porter," *J. Telemat.*, vol. 6, no. 2, pp. 34–40, 2013, doi: <http://dx.doi.org/10.35671/telematika.v6i2.237>.
- [4] R. Feldman, "The text mining handbook: Advanced approaches in analyzing unstructured data," in *Educational and Psychological Measurement*, vol. 28, no. 3, 2007, pp. 123–127. doi: 10.1177/001316446802800332.
- [5] A. Guterres, Gunawan, and J. Santoso, "Stemming Bahasa Tetun Menggunakan Pendekatan Rule Based," *Teknika*, vol. 8, no. 2, pp. 142–147, 2019, doi: 10.34148/teknika.v8i2.224.
- [6] F. Amin and J. Alfa Razaq, "Implementasi Stemmer Bahasa Jawa dengan Metode Rule Base Approach pada Sistem Temu Kembali Informasi Dokumen Teks Berbahasa Jawa,"



- Pros. SENDI_U*, pp. 199–206, 2018, [Online]. Available: <https://www.unisbank.ac.id/ojs/index.php/sendu/article/view/5982>
- [7] M. Saiful, L. M. Samsu, and I. Fathurrahman, “Perancangan Kerangka Crowdsourcing Berbasis Wisdom Of Crowds Untuk Kamus Naskah Lontar (Takepan) Sasak Online,” *Infotek J. Inform. dan Teknol.*, vol. 3, no. 2, pp. 165–173, 2020, doi: 10.29408/jit.v3i2.2300.
- [8] V. Gupta and G. S. Lehal, “A survey of text mining techniques and applications,” *J. Emerg. Technol. Web Intell.*, vol. 1, no. 1, pp. 60–76, 2009, doi: 10.4304/jetwi.1.1.60-76.
- [9] M. S. Qodri, Murahim, M. Efendi, and N. Abdullah, “Belajar Takepan Sasak Bersama Siswa Smk Shofwatul Khaer Desa Bilog Petung, Sembalun, Lombok Timur,” *J. Pendidik. dan Pengabd. Masy.*, vol. 2, no. 2, pp. 1–23, 2019, [Online]. Available: <https://www.google.com/url?sa=t&source=web&rct=j&opi=89978449&url=https://jurnalfkip.unram.ac.id/index.php/JPPM/article/download>