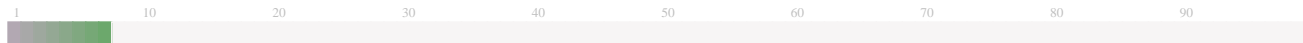


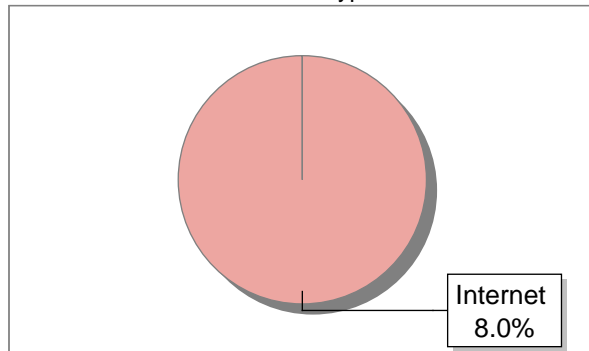
Submission Information

| | |
|--------------------------|--|
| Author Name | Jeffri Alfa Razaq |
| Title | Implementasi Stemmer Bahasa Jawa Dengan Metode Rule Base Approach Pada Sistem Temu Kembali Informasi Dokumen Teks Berbahasa Jawa |
| Paper/Submission ID | 2247637 |
| Submitted by | atikrakhma@edu.unisbank.ac.id |
| Submission Date | 2024-08-20 23:21:26 |
| Total Pages, Total Words | 8, 3338 |
| Document type | Article |

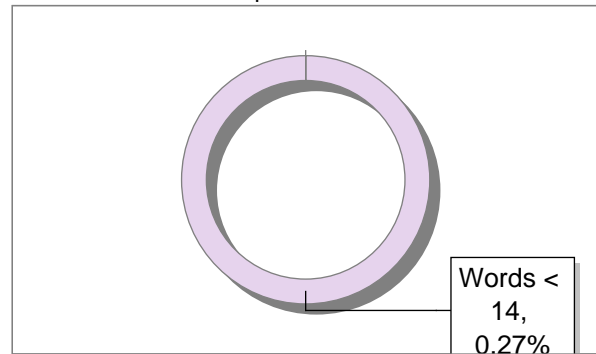
Result Information

Similarity **8 %**

Sources Type



Report Content



Exclude Information

| | |
|-----------------------------|--------------|
| Quotes | Not Excluded |
| References/Bibliography | Not Excluded |
| Source: Excluded < 14 Words | Not Excluded |
| Excluded Source | 98 % |
| Excluded Phrases | Not Excluded |

Database Selection

| | |
|------------------------|-------------|
| Language | Non-English |
| Student Papers | No |
| Journals & publishers | No |
| Internet or Web | Yes |
| Institution Repository | No |

A Unique QR Code use to View/Download/Share Pdf File





DrillBit Similarity Report

8

SIMILARITY %

7

MATCHED SOURCES

A

GRADE

A-Satisfactory (0-10%)
B-Upgrade (11-40%)
C-Poor (41-60%)
D-Unacceptable (61-100%)

| LOCATION | MATCHED DOMAIN | % | SOURCE TYPE |
|----------|----------------|----|---------------|
| 2 | adoc.pub | <1 | Internet Data |
| 3 | docplayer.info | 6 | Internet Data |
| 4 | docplayer.info | 1 | Internet Data |
| 5 | docplayer.info | <1 | Internet Data |
| 6 | adoc.pub | <1 | Internet Data |
| 7 | docplayer.info | <1 | Internet Data |
| 8 | 123dok.com | <1 | Internet Data |

EXCLUDED SOURCES

| | | | |
|---|--------------------|----|-------------|
| 1 | www.unisbank.ac.id | 98 | Publication |
|---|--------------------|----|-------------|

IMPLEMENTASI STEMMER BAHASA JAWA DENGAN METODE RULE BASE APPROACH PADA SISTEM TEMU KEMBALI INFORMASI DOKUMEN TEKS BERBAHASA JAWA

Fatkhul Amin¹, Jeffri Alfa Razaq²

^{1,2} Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Stikubank
e-mail: ¹fatkhulamin@edu.unisbank.ac.id, ²mrjf@edu.unisbank.ac.id

ABSTRAK

Tingkat akurasi hasil pencarian dalam sistem temu kembali dokumen teks ditentukan dalam proses Pre processing yang didalamnya terdapat proses Tokenisasi (tokenizing), Penyaringan (filtering), dan Pembuatan akar kata (stemming). Proses pembuatan akar kata (Stemming) pada dokumen teks berbahasa Jawa menggunakan metode Rule Base Approach dilakukan pada semua dokumen teks bahasa Jawa yang didalamnya terdapat Bahasa Jawa, Bahasa Jawa Madya dan Bahasa Jawa Krama. Proses pembentukan kata pada bahasa Jawa berpengaruh dalam proses pembuatan akar kata yang dilakukan pada proses stemming kata bahasa Jawa. Proses penyusunan kata jadian (tembung andhahan) yang sebelumnya telah diberikan ater-ater (awalan), seselan (sisipan), dan penambang (akhiran) akan dikembalikan menjadi kata dasar pada proses stemmer bahasa Jawa. Proses stemmer bahasa Jawa dilakukan diawali dengan pengambilan ater-ater, sisipan dan penambang. 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%.

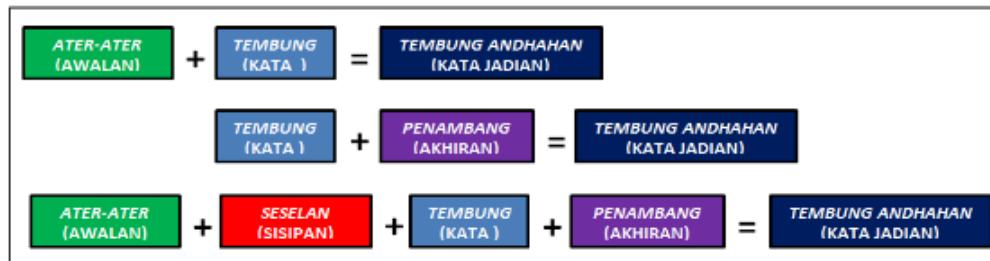
Kata Kunci: Bahasa Jawa, Rule Base Stemmer, Stemmer Jawa

1. PENDAHULUAN

Hasil pencarian informasi dokumen teks menggunakan mesin pencari (*search engine*) atau sistem temu kembali informasi (*information retrieval system*) pada umumnya akan menghasilkan dokumen yang terambil dalam jumlah yang besar (dokumen teks yang ditampilkan pada hasil pencarian dalam jumlah yang besar). Dokumen teks yang dihasilkan dalam jumlah yang besar dalam sebuah pencarian pada mesin pencari akan mengakibatkan pemborosan waktu pencarian yang berakibat lamanya waktu yang dibutuhkan user untuk memilih dokumen yang diinginkan dan akan menghasilkan tingkat akurasi dokumen yang kurang akurat. Hasil Pencarian informasi menggunakan mesin pencari akan menghasilkan dokumen Hasil pencarian yang efektif (akurat) dan efisien (dokumen yang dihasilkan proporsional) jika didukung oleh proses pembuatan akar kata/kata dasar (*stemming processing*) yang tepat.

Langkah pertama dalam proses pengambilan Informasi adalah proses *stemming* (Salton, 1971). Sebuah algoritma stemmer, bertujuan memperoleh batang dari sebuah kata, yaitu akar morfologi dengan menghilangkan afiks yang membawa informasi tata bahasa atau leksikal tentang kata. Dalam kedua kasus, ini imbuhan tidak mengubah konsep kata tersebut terkait sebagai informasi semantik telah terbukti dalam literatur, terutama dalam bahasa yang sangat inflective (Popovic dan Willett, 1992). Proses Stemmer dalam sistem temu kembali informasi (*information retrieval system*) akan berdampak kepada hasil pencarian kata yang akan sangat menentukan pada database yang dihasilkan. Proses stemmer yang tidak benar akan menghasilkan sejumlah kata yang terambil tidak benar dan tidak bisa dilakukan proses selanjutnya. Proses stemmer yang tidak benar bisa terjadi karena terlalu sedikit awalan, sisipan atau akhiran yang diambil dalam sebuah kata. Proses stemmer juga bisa salah karena terlalu banyak awalan, sisipan atau akhiran yang diambil terlalu banyak. Proses stemmer mengharuskan dibuat dengan cara mempelajari morfologi dari suatu bahasa dengan benar sehingga akan didapatkan proses pengambilan awalan, sisipan, akhiran atau kombinasinya dengan benar.

Morfologi bahasa Jawa memiliki keunikan tersendiri jika dibandingkan dengan bahasa Indonesia atau bahasa negara lain dalam hal penyusunan katanya. Proses penyusunan kata dalam bahasa Jawa pada proses penyusunannya akan menggunakan *ater-ater* (awalan), *Seselan* (sisipan) dan *penambang* (akhiran). Jika sebuah kata dasar dalam bahasa Jawa sudah mendapatkan awalan, sisipan, akhiran atau kombinasinya maka kata tersebut disebut *tembung andhahan* (kata jadian). Proses penyusunan kata dalam bahasa Jawa bisa dilihat pada gambar 1.



Gambar 1. Proses Penyusunan *Tembung Andhahan* Kata Jadian dalam Bahasa Jawa

Tembung andhahan merupakan bahasa Jawa yang memiliki arti kata jadian yang terbentuk dari kata dasar yang mendapatkan awalan, sisipan dan akhiran. Dalam bahasa Jawa awalan disebut dengan *ater-ater* yang dibagi kedalam 3 jenis *ater-ater* yaitu; *ater-ater Hanuswara* (n, m, ny, ng), *ater-ater tripurasa* (di, ko, dak), dan *ater-ater liyane* (a, ma, ka, ke, sa, pa, pi, pra, kuma, kami, kapi, tar). Sisipan atau dalam bahasa Jawa disebut *seselan* terdiri dari; *Seselan* (-in, -el, -er, -um).

Penambang (akhiran) terdiri dari: *Penambang* (-i, -ake, -e, -ane, -ke, -a, -ana, -na, -ku, -mu, -en). Morfologi bahasa Jawa yang berbeda dengan bahasa Indonesia memiliki keunikan dan kesulitan tersendiri dalam proses pembuatan akar katanya (*stemming*). Begitu banyaknya awalan, sisipan dan akhiran dalam bahasa Jawa membuat tingkat kesulitan yang semakin kompleks dalam proses pembuatan *stemmer* bahasa Jawa.

Solusi untuk mendapatkan hasil akar kata bahasa Jawa yang tepat dilakukan dengan membuat *Stemmer* Bahasa Jawa dengan Metode *Rule Based Approach* yang akan menghasilkan pembuatan kata dasar bahasa Jawa yang sesuai.

2. TINJAUAN PUSTAKA

Penelitian terdahulu dengan judul "*A survey of Stemming Algorithms for Information Retrieval*" oleh Rajput, B.S., Dkk (2015). Telah menyimpulkan bahwa Sekarang dokumen teks sehari terkumpul melalui internet, e-mail dan halaman web. Seperti penggunaan internet adalah tumbuh secara eksponensial, kebutuhan penyimpanan data yang besar meningkat. Biasanya banyak dokumen berisi variabel morfologi, sehingga berasal yaitu teknik preprocessing memberikan pemetaan yang berbeda varian morfologi dari kata-kata menjadi kata dasar mereka disebut batang. Proses *Stemming* digunakan dalam informasi Pengambilan sebagai cara untuk meningkatkan kinerja pengambilan didasarkan pada asumsi bahwa istilah dengan batang yang sama biasanya memiliki arti yang sama. Untuk melakukan operasi pada data yang besar, kita memerlukan biasanya lebih perhitungan waktu dan tenaga, untuk mengatasi dengan kebutuhan untuk mencari kata tertentu dalam data. Dalam tulisan ini, berbagai algoritma berasal dianalisis dengan manfaat dan keterbatasan dari teknik yang ada saat ini. Mempelajari berbagai metode dan harus tahu bahwa berasal lumayan meningkatkan Hasil pencarian untuk kedua aturan pendekatan tergantung dan statistik. Hal ini juga berguna dalam mengurangi ukuran indeks file dan set fitur atau atribut sebagai jumlah kata yang akan diindeks dikurangi untuk bentuk umum disebut batang. Kinerja *stemmers* statistik jauh lebih unggul beberapa *stemmers* berbasis aturan yang terkenal tapi kali mengkonsumsi. Aturan tergantung *stemmer* seperti porter *stemmer* adalah pilihan yang baik untuk pengolahan dokumen bahasa Inggris.

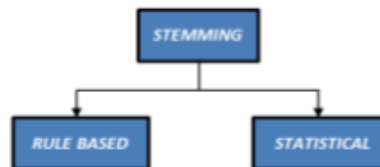
Penelitian terkait *stemmer* dengan judul "*A Literature Review: Stemming Algorithms for Indian Languages*", oleh Thangarasu, M., DKK(2013). *Stemming* adalah proses penggalian akar kata dari kata infleksi yang diberikan. Hal ini juga memainkan peran penting dalam berbagai aplikasi *Natural Language Processing* (NLP). Berbagai Masalah dibahas dalam banyak konteks dan oleh para peneliti di banyak disiplin pengetahuan. Makalah ini menyajikan survei dari beberapa dari perkembangan terbaru di algoritma dalam data mining dan juga menyajikan dengan beberapa solusi untuk berbagai bahasa yang berasal algoritma dari India. *Stemming* memainkan peran penting dalam Sistem pencarian informasi dan efeknya sangat besar, dibandingkan dengan yang ditemukan di berbagai algoritma berasal. makalah ini memiliki fokus tentang berbagai algoritma berasal dan yang efektifitas pada berbagai bahasa India. Hal ini tidak cukup untuk sistem pencarian informasi. Sehingga di masa depan, peneliti akan mencoba untuk jumlah lebih dari implementasi untuk metode algoritma berasal dan mereka menggunakan berbagai bahasa India pada sistem pencarian informasi.

Penelitian terdahulu juga pernah dilakukan dengan topik "*Enhancing Malay Stemming Algorithm with Background Knowledge*" oleh Leong, L.C., DKK (2012). *Stemming* merupakan proses mengurangi kata-kata infleksi untuk akar mereka bentuk. Berasal algoritma untuk bahasa Melayu sangat penting terutama dalam membangun sistem pencarian informasi yang efektif. Meskipun ada banyak *stemmers* Melayu seperti Othman dan algoritma Fatimah, *stemmer* yang mereka buat tidak lengkap karena algoritma mereka gagal untuk membendung semua bahasa Melayu, beberapa kata-kata masih ada ruang untuk perbaikan. Sulit untuk menerapkan *stemmer* sempurna untuk bahasa Melayu karena kompleksitas dari morfologi dalam bahasa

Melayu. Makalah ini menyajikan pendekatan baru untuk membuat kata dasar Melayu dengan persentase yang lebih tinggi dari benar berasal kata-kata. Dalam pendekatan yang diusulkan, latar belakang pengetahuan tambahan disediakan untuk meningkatkan akurasi membendung kata-kata dalam bahasa Melayu. Pendekatan baru ini disebut Melayu stemmer dengan latar belakang pengetahuan. Selain memiliki referensi ke kamus yang berisi semua kata-kata akar, referensi kedua untuk kamus ditambahkan bahwa mengandung semua kata ditempel. Kedua file dianggap sebagai latar belakang pengetahuan yang akan berfungsi sebagai referensi untuk proses stemming. Sebuah *Rule frequency Order (RFO)* diterapkan sebagai dasar berasal algoritma karena akurasi yang tinggi berasal kata Melayu. Berdasarkan hasil yang diperoleh, terbukti bahwa stemmer yang diusulkan dengan latar belakang pengetahuan menghasilkan kurang error dibandingkan dengan stemmers diterbitkan sebelumnya yang tidak menerapkan latar belakang pengetahuan dalam membendung kata-kata Melayu.

Proses pembuatan akar kata atau kata dasar (*Stemming*) digunakan untuk mengubah *term* yang masih melekat dalam *term* (kata) tersebut awalan, sisipan, dan akhiran. Selanjutnya *term* tersebut diproses untuk dihilangkan awalannya (jika ada), sisipan (jika ada) dan akhiran (jika ada) sehingga menjadi *term* kata dasar. Proses membuat *term* dasar ini mengacu kepada bahasa Jawa (ngoko, madya dan krama) yang benar. Proses stemming dilakukan dengan cara menghilangkan semua imbuhan (*affixes*) baik yang terdiri dari awalan (*prefixes*), sisipan (*infixes*), akhiran (*suffixes*) dan *confixes* (kombinasi dari awalan dan akhiran) pada kata turunan atau kata jadian. *Stemming* digunakan untuk mengganti bentuk dari suatu kata menjadi kata dasar dari kata tersebut yang sesuai dengan struktur morfologi bahasa Jawa yang benar. Imbuhan (*affixes*) pada Bahasa Jawa lebih kompleks jika dibandingkan dengan imbuhan pada Bahasa Inggris. Imbuhan pada Bahasa Jawa terdiri dari *ater-ater* atau awalan (*prefixes*), *seselan* atau sisipan (*infixes*), *penambang* atau akhiran (*suffixes*), bentuk perulangan (*repeated forms*) dan kombinasi awalan akhiran (*confixes*). Imbuhan-imbuhan yang melekat pada suatu kata harus dihilangkan untuk mengubah bentuk kata tersebut menjadi bentuk kata dasarnya. Penelitian terhadap *stemming* untuk *retrieval*, *machine translation*, *document summarization* dan *text classification* sudah pernah dilakukan sebelumnya. *Stemming* yang dilakukan pada *text retrieval*, *stemming* ini meningkatkan kesensitifan *retrieval* dengan meningkatkan kemampuan untuk menemukan dokumen yang relevan, tetapi hal itu terkait dengan pengurangan pada pemilihan dimana pengelompokan menjadi kata dasar menyebabkan penghilangan makna kata.

Algoritma *Stemming* bisa diklasifikasikan menjadi 2 (dua) yaitu; *Rule Based* dan *Statistical* (gambar 2).



Gambar 2. *Types of Stemming Approach* (Sharma, 2012)

3. METODE PENELITIAN

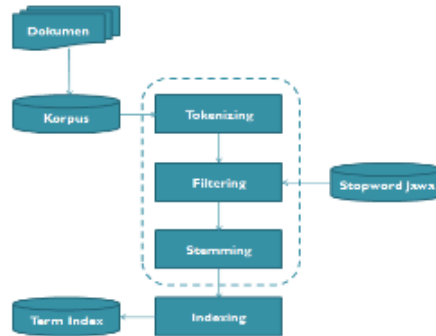
a. *Rule Base Stemmer*

Information Retrieval System menemukan informasi yang biasanya dalam bentuk dokumen dari sebuah data yang tidak terstruktur dalam bentuk teks untuk memenuhi kebutuhan informasi dari koleksi data yang sangat besar umumnya tersimpan dalam *database computer* (Manning, 2008). *Stemming* adalah proses untuk menemukan akar kata (*root*) atau kata dasar dengan memisahkan semua affix atau imbuhan yang melekat pada kata tersebut (Indrajaya, 2003). *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 imbuhan.

Metode *Rule Base Stemmer* 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) Awalan (*prefix*), *suffix* (akhiran) dan *infix* (sisipan) dari *term* bentukannya menjadi suatu *term* kata dasar. *Stemmer* diharapkan akan mampu mengurangi dimensi data (jumlah data terambil) sehingga nantinya akan meningkatkan performansi (akurasi) dari proses kategorisasi. Semakin sedikit dimensi data, maka akan semakin sedikit pula *rule-rule* untuk menentukan suatu dokumen terkategori ke suatu kelas atau kategori tertentu sehingga bisa meningkatkan hasil kategorisasi. Pengurangan dimensi data dengan menggunakan *stemming* terjadi karena kata-kata yang memiliki kata dasar yang sama dikelompokkan menjadi satu atribut sehingga dimensi data bisa direduksi.

b. Arsitektur Stemmer Bahasa Jawa

Arsitektur Informasi pada Stemmer Bahasa Jawa menggunakan model Sistem temu kembali Informasi yang hanya diambil pada proses *Pre Processing*. Penekanan pada kajian stemmer sehingga yang dilakukan sistem hanya dimulai pada saat proses input dokumen kedalam korpus, proses tokenisasi, proses filtering dan Stemming. Gambar 3 menunjukkan arsitektur informasi stemmer bahasa Jawa dalam sebuah sistem temu kembali informasi (STKI).



Gambar 3. Arsitektur Informasi Stemmer Bahasa Jawa dalam sebuah STKI

c. Flowchart Stemming Bahasa Jawa

Proses *stemming* bahasa Jawa dirancang agar *term* hasil *filtering* diubah menjadi *term* kata dasar melalui proses pembuatan akar kata melalui beberapa tahapan. Proses *stemming* dimulai dengan menghilangkan awalan dan akhiran. Proses ini juga dirancang dapat melakukan penggantian (*replace*) ketika awalan dihilangkan dan menggantinya dengan huruf yang sesuai. Proses menghilangkan awalan, akhiran, dan *replace* sisipan dilakukan dalam satu tahap proses. Gambar 4 menunjukkan aplikasi *flowchart stemming*.



Gambar 4. Flowchart Proses Stemming

c. Proses Stemming

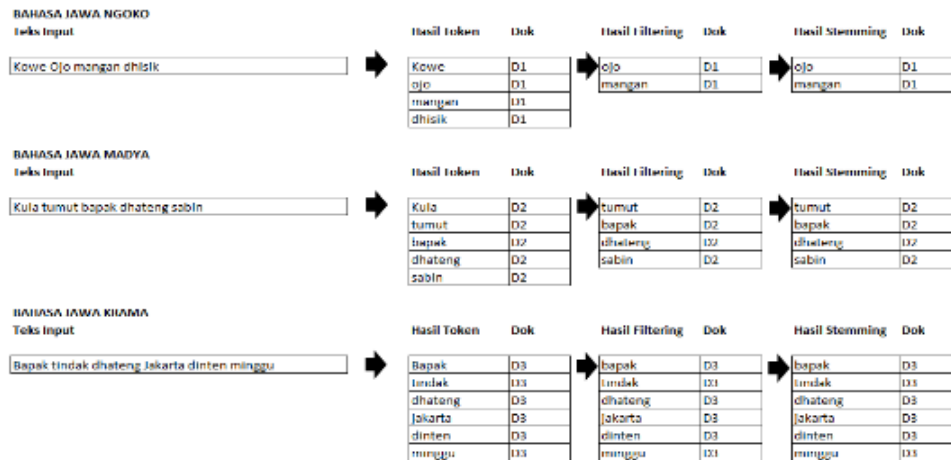
Proses *Stemming* digunakan untuk mengubah *term* yang masih melekat dalam *term* tersebut awalan, sisipan, dan akhiran. Selanjutnya *term* tersebut diproses untuk dihilangkan awalan, sisipan dan akhiran sehingga menjadi *term* kata dasar. Proses membuat *term* dasar ini mengacu kepada bahasa Jawa yang benar.

Stemmer bahasa Jawa dibuat menggunakan pendekatan morfologi bahasa Jawa. Proses *Stemming* digunakan untuk mengubah *term* yang masih melekat dalam *term* tersebut awalan, sisipan, dan akhiran. Selanjutnya *term* tersebut diproses untuk dihilangkan awalan, sisipan dan akhiran sehingga menjadi *term* kata dasar. Proses membuat *term* dasar ini mengacu kepada bahasa Jawa yang benar (amin, 2015).

Proses *stemming* dilakukan dengan cara menghilangkan semua imbuhan (*affixes*) baik yang terdiri dari awalan (*prefixes*), sisipan (*infixes*), akhiran (*suffixes*) dan *confixes* (kombinasi dari awalan dan akhiran) pada kata turunan. *Stemming* digunakan untuk mengganti bentuk dari suatu kata menjadi kata dasar dari kata tersebut yang sesuai dengan struktur morfologi bahasa Jawa yang benar

Imbuhan (*affixes*) pada Bahasa Jawa lebih kompleks jika dibandingkan dengan imbuhan pada Bahasa Inggris. Imbuhan pada Bahasa Jawa terdiri dari awalan (*prefixes*), sisipan (*infixes*), akhiran (*suffixes*), bentuk perulangan (*repeated forms*) dan kombinasi awalan akhiran (*confixes*). Imbuhan-imbuhan yang melekat pada suatu kata harus dihilangkan untuk mengubah bentuk kata tersebut menjadi bentuk kata dasarnya. *Stemming* teks berbahasa Jawa memiliki beberapa masalah yang sangat khusus terhadap bahasa. Salah satu masalah tersebut

adalah perbedaan tipe dari imbuhan-imbuhan (*affixes*), bahwa awalan (*prefixes*) dapat berubah tergantung dari huruf pertama pada kata dasar. Sebagai contoh “ng-” dapat berubah “k-” ketika huruf pertama dari kata dasar tersebut adalah “ng”, misalkan “ngetok”(kata dasar kethok), tetapi dapat berubah menjadi “ng-” ketika huruf pertama dari kata dasar melekat adalah “k”, misalkan “ngethok” (kata dasar kethok). Contoh proses Token, Filtering dan *Stemming* pada aplikasi bahasa Jawa ngoko, Jawa Madya dan Jawa Krama bisa dilihat pada gambar 5.



Gambar 5. Contoh hasil proses stemming pada bahasa Jawa ngoko, Jawa Madya dan Jawa Krama

d. *Stemmer* Bahasa Jawa

Pada penelitian *Stemmer* Bahasa Jawa di adopsi dari algoritma *porter stemmer*. Metode *Stemmer* menggunakan *rule base* analisis untuk mencari *root* sebuah kata. *Stemmer* ini sama sekali tidak menggunakan kamus sebagai acuan. Struktur pembentukan kata dalam bahasa Jawa adalah sebagai berikut:

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

Masing-masing bagian digabungkan dengan kata dasar membentuk kata berimbuhan. Algoritma Bahasa Jawa menggunakan algoritma *rule based stemming* seperti halnya dengan algoritma *porter* pada *stemming* bahasa Inggris.

Pada *stemmer* Bahasa Jawa Terdapat 5 langkah utama dengan 2 langkah awal dan 2 langkah pilihan, langkah-langkah tersebut adalah sebagai berikut:

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

4. HASIL DAN PEMBAHASAN

Stemmer dirancang bisa menghasilkan kata dasar bahasa Jawa dengan benar. *Stemmer* Bahasa Jawa dirancang melalui beberapa tahap, yaitu: Tokenisasi, *filtering* dan *Stemming*. Gambar 6 menunjukkan proses *Stemmer* Bahasa Jawa dimulai dengan proses pengumpulan dokumen yang diinput kedalam Korpus (tabel). Data berupa dokumen kemudian di proses tokenisasi yang menghasilkan *term-term* yang terpisah didasarkan pada spasi saat pemrosesannya. *Term* hasil proses tokenisasi selanjutnya dilakukan proses *filtering*. *Term* hasil proses *filtering* kemudian dilakukan *stemming* dan menghasilkan kata dasar Bahasa Jawa

4.1. Interface

Menu Hasil *Stemmer* Bahasa Jawa akan ditampilkan per kata. Setiap kata jadian (tembung andhahan) akan dilakukan proses pemisahan kata dengan awalan dan akhiran. Aplikasi hasil *stemmer* bisa dilihat pada gambar 6.

| | |
|------------|-----------|
| JUDUL | ATER-ATER |
| KATA KUNCI | DIBATHIN |
| FREKUENSI | 1 |
| AWALAN | DI |
| KATA DASAR | BATHIN |
| AKHIRAN | - |

Gambar 6. Interface Hasil Stemmer Bahasa Jawa

4.2. Aplikasi Stemmer Bahasa Jawa

Proses *stemming* yang digunakan adalah proses *stemmer* menggunakan *stemmer* untuk bahasa Jawa. Proses *stemming* dengan menggunakan *stemmer* Jawa melalui beberapa tahapan dan untuk mendukung proses ini digunakan *stopword list* Jawa. Hasil akhir dari proses *stemming* adalah kumpulan *term* yang sudah menjadi kata dasar yang diinput dalam database kumpulan term hasil *stemming*. *Stemmer* Bahasa Jawa ini dirancang agar didapatkan bentuk kata dasar bahasa Jawa yang benar. Tampilan *interface* dirancang dalam bentuk terpisah mulai dari term kata jadian, awalan, akhiran dan kata dasar hasil proses *stemming*. (gambar 7). Prosedur menggunakan *Stemmer* bahasa Jawa ini sangat mudah, yaitu dengan menampilkan softwarenya

The screenshot shows the 'Hasil Proses Tokenizing dan Filtering' window. It displays the analysis of the term 'Ater-ater Tripurasa'. The analysis includes:

- Term: **dakdhudhuk**
- Frekuensi = 1
- Awalan: **dak**
- Kata Dasar: **dhudhuk**
- Akhiran: -

 A callout box lists the components:

- masuk Dolepangan
- masuk dakdhudhuk
- masuk kojupuk
- masuk kogoreng
- masuk diambung
- masuk dibatin

 Below this, other terms are listed:

- 11. Term: **kojupuk**, Frekuensi = 1, Awalan: **ko**, Kata Dasar: **jupuk**, Akhiran: -
- 12. Term: **kogoreng**, Frekuensi = 1, Awalan: **ko**, Kata Dasar: **goreng**, Akhiran: -
- 13. Term: **diambung**, Frekuensi = 1, Awalan: **di**, Kata Dasar: **ambung**
- 9. Term: **dakpangan**, Frekuensi = 1, Awalan: **dak**, Kata Dasar: **pang**, Akhiran: **an**
- 14. Term: **dibatin**, Frekuensi = 1, Awalan: **di**, Kata Dasar: **batin**, Akhiran: -

Gambar 7. Interface stemmer

4.3. Implementasi Stemmer Bahasa Jawa

Studi kasus pada aplikasi *Stemmer* bahasa Jawa ini menggunakan kumpulan tembung andhahan (kata jadian). Sebagai contoh digunakan tembung andhahan yang telah diberi ater-ater tripurasa. Term yang akan diproses adalah: "*dakpangan dakdhudhuk kojupuk kogoreng diambung dibatin*". Proses *stemmer* menghasilkan program bisa memisahkan awalan dengan kata dengan benar, meskipun ada satu kata yang terambil tidak benar yaitu kata "*dakpangan*". Beberapa kata yang terambil dan diproses dengan benar diantaranya:

- "*dakdhudhuk*" = *dak* + *duduk*
- "*kojupuk*" = *ko* + *jupuk*
- "*kogoreng*" = *ko* + *goreng*
- "*diambung*" = *di* + *ambung*
- "*dibatin*" = *di* + *batin*

4.4. Pengujian hasil Stemmer Bahasa Jawa

Pengujian program dilakukan dengan cara membandingkan hasil proses *stemming* dengan kamus bahasa Jawa. Pelaksanaan pengujian hasil dilakukan pada setiap awalan (*ater-ater*), sisipan (*seselan*) dan akhiran (*penambang*).

a. Uji Hasil *ater-ater* Hanuswara

Ater-ater Hanuswara adalah awalan dalam bahasa Jawa. Adapun yang termasuk dalam ater-ater hanuswara antara lain; m, n, ng, ny (tabel 1)

Tabel 1. *ater-ater hanuswara*

| | | |
|----|--------|---------|
| m | bathik | mbathik |
| n | tulis | nullis |
| ng | gawa | nggawa |
| ny | sikut | nyikut |

b. Uji Hasil *ater-ater Tripurasa*

Ater-ater Tripurasa adalah awalan dalam bahasa Jawa. Adapun yang termasuk dalam *ater-ater Tripurasa* antara lain; dak, ko, di (tabel 2)

Tabel 2. *ater-ater Tripurasa*

| | | |
|-----|--------|-----------|
| dak | pangan | dakpangan |
| ko | jupuk | kojupuk |
| di | batin | dibatin |

c. Uji Hasil *ater-ater Liyane*

Ater-ater Liyane adalah awalan dalam bahasa Jawa. Adapun yang termasuk dalam *ater-ater Liyane* antara lain; a, ma, ka, ke, sa, pa, pi, pra, kuma, kami, kapi, tar (tabel 3)

Tabel 3. *ater-ater Liyane*

| | | |
|------|----------|------------|
| a | lungguh | alungguh |
| ma | lumpat | malumpat |
| ka | limpe | kalimpe |
| ke | sandhung | kesandhung |
| sa | gegem | sagegem |
| pa | lilah | palilah |
| pi | tutur | pitutur |
| pra | wira | prawira |
| kuma | wani | kumawani |
| kami | tuwo | kamituwo |
| kapi | lare | kapilare |
| tar | waca | tarwaca |

d. Uji Hasil *ater-ater Seselan*

Seselan adalah sisipan dalam bahasa Jawa. Adapun yang termasuk dalam *Seselan* antara lain; -um, -in, -el lan -er (tabel 4)

Tabel 4. *Seselan*

| | | |
|-----|-------|---------|
| -um | ayu | kemayu |
| -in | serat | sinerat |
| -el | dewer | delewer |
| -er | cuwil | ceruwil |

e. Uji Hasil *ater-ater Penambang*

Penambang adalah akhiran dalam bahasa Jawa. Adapun yang termasuk dalam *penambang* antara lain; -I, -ake, -e, -ane, -ke, -a, -ana, -na, -ku, -mu, -en (tabel 5)

Tabel 5. *Penambang*

| | | |
|--------|------|----------|
| turut | -i | turuti |
| jupuk | -ake | jupukake |
| teka | -ne | tekane |
| bapak | -e | bapake |
| jaluk | -ane | jalukane |
| kethok | -ke | kethoke |
| turu | -a | turua |
| weneh | -ana | wenehana |
| gawa | -na | gawakna |
| buku | -ku | bukuku |
| omah | -mu | omahmu |
| lepeh | -en | lepehen |

4.5. Hasil Uji Stemmer Bahasa Jawa

Hasil pengujian stemmer bahasa jawa bisa dilihat pada tabel 6.

Tabel 6. Hasil Uji stemmer jawa

| No | Ater-ater | Hasil |
|----|--|--|
| 1 | <i>Hanuswara:</i> m-, n-, ng-, ny- | Uji stemmer jawa menghasilkan kata dasar dengan Tepat |
| 2 | <i>Tripurasa:</i> Dak- | Uji stemmer jawa menghasilkan kata dasar dengan Belum tepat |
| 3 | <i>Tripurasa:</i> Ko- dan di- | Uji stemmer jawa menghasilkan kata dasar dengan Tepat |
| 4 | <i>Liyane:</i> ma, ka, ke, sa, pa, pi, pra, kuma, kami, kapi, dan tar | Uji stemmer jawa menghasilkan kata dasar dengan Tepat |
| 5 | <i>Seselan:</i> -um, -in, -el lan -er | Uji stemmer jawa menghasilkan kata dasar dengan Belum tepat |
| 6 | <i>Penambang:</i> -e, -ke, dan -a | Uji stemmer jawa menghasilkan kata dasar dengan Belum tepat |
| 7 | <i>Penambang:</i> -I, -ake, -ane, -, -ana, -na, -ku, -mu, -, -en | Uji stemmer jawa menghasilkan kata dasar dengan Tepat |

5. KESIMPULAN

- Stemmer Bahasa Jawa yang dibuat* mampu menghasilkan kata dasar jawa dengan benar 77 % (*ater-ater/awalan, seselan/sisipan dan penambang/akhiran*)
- Stemmer Bahasa Jawa* mampu melakukan proses *stemming* dengan hasil yang mudah di pahami karena dibuat terpisah antara awalan, kata dasar dan akhiran.

UCAPAN TERIMA KASIH (APABILA DIPERLUKAN)

Penulis mengucapkan terima kasih kepada;

- Direktorat Riset Pengabdian Kepada Masyarakat Direktorat Jenderal Penguatan Riset dan Pengembangan Kementerian Riset, Teknologi, dan Pendidikan Tinggi.
- Bapak **Dr. H. Hasan Abdul Rozak, S.H., C.N.,M.M.** selaku Rektor Universitas Stikubank (Unisbank) Semarang.
- Bapak **Dr. Agus Budi Santosa, M.Si.**,selaku Kepala Lembaga Penelitian dan pengabdian Masyarakat (LPPM) Universitas Stikubank (Unisbank) Semarang.

DAFTAR PUSTAKA

- Leong L.C., Dkk., 2012. *Enhancing Malay Stemming Algorithm with Background Knowledge*. Pacific Rim International Conference on Artificial Intelligence (PRICAI). Trends in Artificial Intelligencepp 753-758
- Manning, C., Raghavan, P., 2007. *An Introduction to Information Retrieval*, Stanford. USA.
- Meadow, C.T., 1997. *Text Information Retrieval Systems*. Academic Press.New York.
- Tala, F.Z., 2003, *A Study of Stemming Effects on Information Retrieval in bahasa Indonesia*. Institut for logic, Language and Computation Universiteit van Amsterdam The Netherlands.
- M.Thangarasu, Dkk. 2013., A Literature Review: Stemming Algorithms for Indian Languages, *International Journal of Computer Trends and Technology (IJCTT) – volume 4 Issue 8–August 2013*
- Popovic, M. & Willett, P. 1992. The effectiveness of stemming for natural-language access to Slovene textual data. *Journal of the American Society for Information Science*, 43(5), 384-390.
- Porter, M. F. 1980. *An algorithm for suffix stripping*. *Program*, 14, 130-137.
- Rajput B.S., Dkk., 2015. A survey of Stemming Algorithms for Information Retrieval. *IOSR Journal of Computer Engineering (IOSR-JCE)* e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 17, Issue 3, Ver. VI (May – Jun. 2015), PP 76-80
- Salton, G., 1971., *The SMART retrieval system - experiments in automatic document processing*. Upper Saddle River, NJ: Prentice-Hall, Inc.
- Salton, G., 1989, *Automatic Text Processing, The Transformation, Analysis, and Retrieval of information by computer*. Addison – Wesley Publishing Company, Inc. USA.
- Yates, R.B, 1999. *Modern Information Retrieval, Addison Wesley-Pearson* international edition, Boston. USA.