

Stemmer Bahasa Jawa Ngoko dengan Metode Affix Removal Stemmers (Rule Based Approach)

Fatkul Amin dan Purwatiningsih

Fakultas Teknologi Informasi, Universitas Stikubank Semarang
email: fatkhulamin@gmail.com, diba_ian@yahoo.com

Abstrak

Proses pengembalian kata dasar bahasa *Jawa ngoko* dari kata jadian (*tembung andhahan*) menjadi kata dasar yang salah akan mengakibatkan kata dasar menjadi rusak, tidak memiliki arti yang benar dan tidak bisa digunakan untuk proses setelah *Stemming*. Bahasa *Jawa Ngoko* memiliki morfologi dalam penyusunan kata yang didalamnya terdapat awalan (*ater-ater*), sisipan (*seselan*) dan akhiran (*penambang*). Penguasaan morfologi Bahasa Jawa ngoko membantu menciptakan cara untuk pengambilan kata dasar dalam bahasa *jawa ngoko* secara benar. *Stemmer Bahasa Jawa ngoko* metode *Affix Removal* digunakan untuk mendapatkan kata dasar dari hasil proses pengurangan awalan, sisipan dan akhiran secara benar. Hasil pengembalian kata dasar menggunakan *Stemmer Bahasa Jawa ngoko* metode *Affix Removal* mampu membuat kata dasar dalam bahasa Jawa ngoko dengan hasil benar mencapai 62%. Kemampuan *stemmer* ini masih harus ditingkatkan sampai mencapai tingkat kebenaran dalam pengembalian kata dasar mencapai 100%.

PENDAHULUAN

Proses pencarian pada mesin pencari terbukti efektif dan efisien dalam melaksanakan kerjanya sebagai penyedia informasi bagi penggunaannya. Proses pencarian informasi menggunakan mesin pencari bisa menghasilkan hasil pencarian yang banyak namun juga bisa menghasilkan hasil pencarian yang sedikit. Hal ini karena algoritma yang digunakan berbeda. Salah satu komponen dalam proses pencarian adalah proses *stemmer*. 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.

Bahasa *Jawa ngoko* memiliki morfologi yang berbeda dengan bahasa Indonesia atau bahasa negara lain dalam hal penyusunan kata. Proses penyusunan kata dalam bahasa *jawa ngoko* pada proses penyusunannya akan menggunakan *ater-ater* (awalan), *Seselan* (sisipan) dan *penambang* (akhiran). Jika sebuah kata dasar dalam bahasa *jawa ngoko* sudah mendapatkan awalan, sisipan, akhiran atau kombinasinya maka kata tersebut disebut *tembung andhahan* (kata jadian). *Tembung andhahan* adalah kata yang telah mendapatkan awalan, sisipan, akhiran atau kombinasi dari awalan, sisipan, akhiran. Berikut ini adalah *Ater-ater* (awalan) terdiri dari; *ater-ater Hanuswara* (m, n, ng, ny), *ater-ater tripurasa* (dak, ko, di) dan *ater-ater liyane* (a, ma, ka, ke, sa, pa, pi, pra, kuma, kami, kapi, tar). *Seselan* (sisipan) terdiri dari: *Seselan* (-um, -in, -el, lan -er). *Penambang* (akhiran) terdiri dari: *Penambang* (-i, -ake, -e, -ane, -ke, -a, -ana, -na, -ku, -mu, -en). Morfologi bahasa *jawa ngoko* yang berbeda dengan bahasa Indonesia memiliki keunikan dan kesulitan tersendiri dalam proses *stemmingnya*. Begitu banyaknya awalan, sisipan dan akhiran dalam bahasa *jawa ngoko* membuat tingkat kesulitan yang semakin kompleks dalam proses pembuatan *stemmer* bahasa *jawa ngoko*.

Solusi untuk mengatasi masalah *stemmer* bahasa *jawa ngoko* adalah dengan membuat

Stemmer Bahasa Jawa Ngoko dengan Metode *Affix Removal Stemmers (Rule Based Approach)*. Metode *Affix Removal Stemmers (Rule Based Approach)* dipilih karena menghasilkan *stemmer* bahasa jawa ngoko yang tepat.

Rumusan Masalah

Bagaimana membuat *Stemmer* Bahasa Jawa Ngoko dengan Metode *Affix Removal Stemmers (Rule Based Approach)*?

Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian ini adalah;

- a. Membuat *Stemmer* Bahasa Jawa Ngoko dengan Metode *Affix Removal Stemmers*
- b. Membuat panduan *stemmer* bahasa jawa ngoko yang benar sehingga bisa digunakan untuk penelitian-penelitian terkait dengan Sistem Temu Kembali Informasi.

METODE

1. *Stemmer* Bahasa Jawa Ngoko

a. *Stemming*

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.

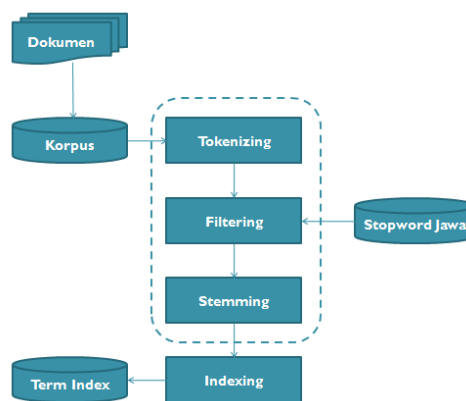
Metode *Affix Removal* digunakan karena sifatnya yang fleksibel untuk digunakan sebagai *stemmer* berbagai macam bahasa dengan

karakteristiknya yang lebih menekankan pada struktur morfologi suatu bahasa. Metode ini akan membuang Awalan (*prefix*), *suffix* (akhiran) dan *infix* (sisipan) dari *term* menjadi suatu stem.

Stemmer diharapkan akan mampu mengurangi dimensi data sehingga nantinya akan meningkatkan performansi 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 Information Retrieval System*

Arsitektur Informasi pada *Stemmer* Bahasa Jawa Ngoko 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 1 menunjukkan *arsitektur informasi stemmer* bahasa Jawa ngoko



Gambar 1. *Arsitektur Informasi Stemmer* Bahasa Jawa Ngoko

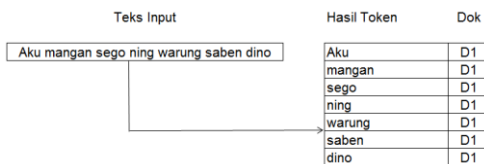
c. *Corpus*

Corpus digunakan sebagai media untuk menempatkan data di database untuk dilakukan proses berikutnya dari sistem *stemmer* bahasa jawa ngoko. Penelitian dengan menggunakan

database pada aplikasinya biasanya memakai korpus untuk proses pembuatan tabel pendukungnya. Penelitian empiris dapat dilakukan dengan menggunakan teks tertulis atau lisan, seperti teks-teks dasar dari berbagai jenis sastra dan analisis linguistik. Tapi gagasan tentang korpus sebagai dasar untuk sebuah bentuk linguistic empiris berbeda dalam beberapa cara mendasar dari teks-teks tertentu. Pada prinsipnya, setiap koleksi lebih dari satu teks dapat disebut dengan *korpus* (McEnery dan Wilson, 2001): istilah *korpus* dalam bahasa latin berarti *body*, maka *korpus* dapat didefinisikan sebagai isi setiap teks.

d. Proses Tokenisasi

Sistem temu kembali Informasi dimulai dengan proses memisahkan kata yang ada pada dokumen berdasarkan spasi kemudian memproses kata yang telah dipisahkan tersebut kedalam sebuah tabel untuk dilakukan proses berikutnya. *Tokenizing* merupakan proses pemisahan suatu rangkaian karakter berdasarkan karakter spasi, dan mungkin pada waktu yang bersamaan dilakukan juga proses penghapusan karakter tertentu, seperti tanda baca. Sebagai contoh, kata-kata “computer”, “computing”, dan “compute” semua berasal dari term yang sama yaitu “comput”, tanpa pengetahuan sebelumnya dari morfologi bahasa Inggris. Token seringkali disebut sebagai istilah (*term*) atau kata, sebagai contoh sebuah token merupakan suatu urutan karakter dari dokumen tertentu yang dikelompokkan sebagai unit semantik yang berguna untuk diproses (Salton, 1989). contoh tokenisasi bisa dilihat pada gambar 2. Input : Aku mangan sega ning warung saben dino (amin, 2015)

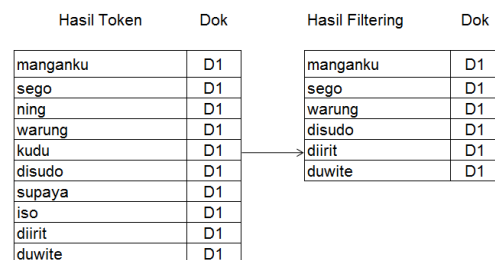


Gambar 2. Contoh hasil proses tokenisasi

e. Proses *Filtering*

Pada riset *stemmer jawa ngoko*, proses *filtering* tidak dilakukan karena yang menjadi konsentrasi adalah proses *stemming*. *Filtering*

akan memproses kata hasil proses tokenisasi menjadi lebih sedikit dengan cara mengurangi kata tersebut dengan kata yang termasuk dalam *stopwords*. Eliminasi *stopwords* memiliki banyak keuntungan, yaitu akan mengurangi *space* pada tabel *term index* hingga 40% atau lebih (Baeza, 1999,h.167). Proses *stopword removal* merupakan proses penghapusan *term* yang tidak memiliki arti atau tidak relevan. Proses ini dilakukan pada saat proses tokenisasi. Proses tokenisasi menghasilkan sebuah *term*, dan *term* tersebut selanjutnya di periksa dalam daftar *stopword*. Apabila *term* tersebut terdapat dalam daftar *stopword* maka *term* tersebut tidak akan dimasukkan dalam tabel *term*. Sebaliknya *term* hasil tokenisasi apabila diperiksa ke dalam daftar *stopword* dan hasilnya nihil maka *term* tersebut akan dimasukkan ke dalam tabel *term*(Baeza, 1999,h.167). beberapa cara dalam proses *stopword removal*, antara lain meletakkan proses *stopword* sebelum *term* hasil tokenisasi dimasukkan ke dalam tabel *term*, cara yang kedua menempatkan proses *stopword* setelah *term* hasil tokenisasi masuk kedalam tabel (amin, 2015). Gambar 3 menunjukkan proses *filtering*.



Gambar 3. Contoh hasil proses *Filtering*

f. 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 ngoko yang benar. Contoh *Stemming* bisa dilihat pada gambar 4.

Stemmer bahasa *Jawa ngoko* dibuat menggunakan pendekatan morfologi bahasa

jawa ngoko. 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 ngoko 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 ngoko lebih kompleks jika dibandingkan dengan imbuhan pada Bahasa Inggris. Imbuhan pada Bahasa Jawa ngoko 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. *Steming* teks berbahasa Jawa ngoko 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 *Stemming* bisa dilihat pada gambar 4. (amin, 2015)

Hasil Filtering	Dok	Hasil Stemming	Dok
manganku	D1	mangan	D1
sego	D1	sego	D1
warung	D1	warung	D1
disudo	D1	sudo	D1
diirit	D1	irit	D1
duwite	D1	duwit	D1

Gambar 4. Contoh hasil proses Stemming

g. *Stemmer* Bahasa Jawa ngoko

Pada penelitian *Stemmer* Bahasa Jawa Ngoko 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 Ngoko adalah sebagai berikut:

$$[\text{awalan}] + [\text{sisipan}] + [\text{dasar}] + [\text{akhiran}]$$

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

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

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

2. Metodologi Penelitian

a. Obyek Penelitian

Obyek penelitian dari penelitian ini adalah *Stemmer* Bahasa Jawa Ngoko.

b. Teknik Pengumpulan Data

Pengumpulan data dimaksudkan agar mendapatkan bahan-bahan yang relevan, akurat dan reliable. Maka teknik pengumpulan data yang dilakukan dalam penelitian ini adalah sebagai berikut:

1) Observasi

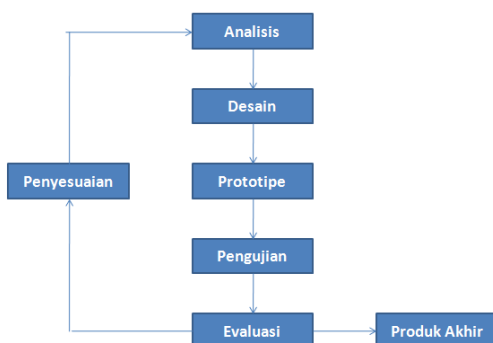
Melakukan pengamatan dan pencatatan secara sistematis tentang hal-hal yang berhubungan dengan basis data dokumen teks dan kemampuan pencarian kemiripan dokumen.

2) Studi Pustaka

Pengumpulan data dari bahan-bahan referensi, arsip, dan dokumen yang berhubungan dengan permasalahan dalam penelitian ini.

c. Metode Pengembangan

Penelitian ini menggunakan model *prototype*. Di dalam model ini sistem dirancang dan dibangun secara bertahap dan untuk setiap tahap pengembangan dilakukan percobaan-percobaan untuk melihat apakah sistem sudah bekerja sesuai dengan yang diinginkan. Sifatnya model *prototype* terdapat pada Gambar 5 memperlihatkan tahapan pada *prototype*.



Gambar 5. Tahapan *Prototyping* (Pressman, 2001)

Berikut adalah tahapan yang dilakukan pada penelitian ini dengan metode pengembangan *prototype*

a. Analisis

Pada tahap ini dilakukan analisa tentang masalah morfologi bahasa *jawa ngoko* dan menentukan pemecahan masalah yang tepat untuk menyelesaikannya.

b. Desain

Pada tahap ini dibangun rancangan sistem dengan beberapa diagram bantu seperti

Arsitektur Informasi dari program dan *Flowchart*. Pembuatan *Flowchart* dilakukan untuk membuat proses menjadi runtut dan membantu proses desain.

c. *Prototype*

Pada tahap ini dibangun aplikasi *Stemmer Bahasa Jawa Ngoko dengan Metode Rule Based Analysis* yang sesuai dengan desain dan kebutuhan sistem.

d. Pengujian

Pada tahap ini dilakukan pengujian *software Stemmer* dengan uji Kamus Bahasa *Jawa Ngoko*.

e. Evaluasi

Pada tahap ini dilakukan evaluasi apakah performa aplikasi sudah sesuai dengan yang diharapkan, apabila belum maka dilakukan penyesuaian-penyesuaian secukupnya.

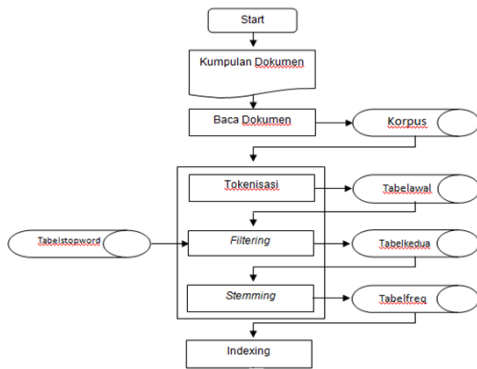
f. Penyesuaian

Tahap ini dilakukan apabila pada evaluasi performa aplikasi kurang memadai dan dibutuhkan perbaikan, tahap ini melakukan penyesuaian dan perbaikan pada aplikasi sesuai dengan kebutuhan

HASIL

1. Rancangan *Stemmer*

Stemmer dirancang bisa menghasilkan kata dasar bahasa *jawa ngoko* dengan benar. *Stemmer Bahasa Jawa Ngoko* dirancang melalui beberapa tahap, yaitu: Tokenisasi, *filtering* dan *Stemming*. Gambar 6 menunjukkan proses *Stemmer Bahasa Jawa Ngoko* dimulai dengan proses pengumpulan dokumen yang diinputkan 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 Ngoko*.



Gambar 6. Rancangan Stemmer Bahasa Jawa Ngoko

2. Flowchart Stemming

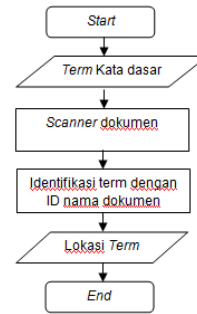
Proses *stemming* dirancang agar *term* hasil *filtering* diubah menjadi *term* kata dasar. Proses *stemming* dimulai dengan menghilangkan awalan dan akhiran. Proses ini juga dirancang dapat melakukan *replace* ketika awalan dihilangkan dan menggantinya dengan huruf yang sesuai. Proses menghilangkan awalan, akhiran, dan *replace* sisipan dilakukan dalam satu tahap proses. Gambar 7 menunjukkan aplikasi *flowchart stemming*.



Gambar 7. Flowchart Proses Stemming

3. Flowchar Indexing

Term kata dasar hasil proses *stemming* selanjutnya dimasukkan dalam tabel untuk diproses pada perhitungan *Vector Space Model*. Proses *indexing* menggunakan metode *inverted indexing*, yaitu dengan membedakan letak tiap *term* dalam dokumen. Gambar 8. menunjukkan *flowchart indexing*.



Gambar 8. Flowchart Proses Indexing

4. Interface

Rancangan menu Hasil Stemmer Bahasa Jawa Ngoko akan ditampilkan per kata. Setiap kata jadian (tebung andhahan) akan dilakukan proses pemisahan kata dengan awalan dan akhiran. Aplikasi hasil stemmer bisa dilihat pada gambar 9.

Judul	Ater-ater Tripuroso
Term	dibathin
Frekuensi	1
Awalan	di
Kata Dasar	bathin
Akhiran	-

Gambar 9. Rancangan Interface Hasil Stemmer Bahasa Jawa Ngoko

5. Aplikasi Stemmer Bahasa Jawa Ngoko

Proses *stemming* yang digunakan adalah proses *stemmer* menggunakan *stemmer* untuk bahasa Jawa ngoko berdasarkan *stemmer* bahasa Indonesia yang dibuat Tala. Proses *stemming* dengan menggunakan *stemmer* jawa ngoko melalui beberapa tahapan seperti terlihat pada gambar 7 dan untuk mendukung proses ini juga digunakan *stopword list jawa ngoko*. Hasil akhir dari proses *stemming* adalah kumpulan *term* yang sudah menjadi kata dasar yang diinput dalam tabel *freq*

Stemmer Bahasa Jawa Ngoko ini dirancang agar didapatkan bentuk kata dasar bahasa jawa ngoko yang benar. Tampilan *interface* dirancang dalam bentuk terpisah mulai dari *term* kata jadian, awalan, akhiran dan kata dasar hasil proses *stemming*. (gambar 10). Prosedur menggunakan *Stemmer* bahasa Jawa Ngoko ini

sangat mudah, yaitu dengan menampilkan softwarena

10
Palintangan : Ater-ater Tripurasa
Term : **dakdhudhuk**
Frekuensi = 1
Awalan : **dak**
Kata Dasar : **dhudhuk**
Akhiran : -

11
Palintangan : Ater-ater Tripurasa
Term : **kojupuk**
Frekuensi = 1
Awalan : **ko**
Kata Dasar : **jupuk**
Akhiran : -

12
Palintangan : Ater-ater Tripurasa
Term : **kogoreng**
Frekuensi = 1
Awalan : **ko**
Kata Dasar : **goreng**
Akhiran : -

13
Palintangan : Ater-ater Tripurasa
Term : **diambung**
Frekuensi = 1
Awalan : **di**
Kata Dasar : **ambung**

9
Palintangan : Ater-ater Tripurasa
Term : **dakpangan**
Frekuensi = 1
Awalan : **dak**
Kata Dasar : **pang**
Akhiran : **an**

14
Palintangan : Ater-ater Tripurasa
Term : **dibatin**
Frekuensi = 1
Awalan : **di**
Kata Dasar : **batin**
Akhiran : -

Hasil Proses Tokenizing dan Filtering

Ater-ater Tripurasa
isi: Dakpangan dakdhudhuk kojupuk kogoreng diambung dibatin
A2
masuk Dakpangan
masuk dakdhudhuk
masuk kojupuk
masuk kogoreng
masuk diambung
masuk dibatin
--

Hasil Proses Stemming

Gambar 10. Interface stemmer

PEMBAHASAN

1. Implementasi Stemmer Bahasa Jawa Ngoko

Studi kasus pada aplikasi Stemmer bahasa Jawa Ngoko 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*” Gambar 10 menunjukkan implementasi proses tersebut. 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*

2. Pengujian hasil Stemmer Bahasa Jawa Ngoko

Pengujian program dilakukan dengan cara membandingkan hasil proses stemming dengan kamus bahasa jawa ngoko. 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 ngoko. Adapun yang termasuk dalam ater-ater hanuswara antara lain; m, n, ng, ny (tabel 1)

Tabel 1. . ater-ater hanuswara

m	bathik	mbathik
m	pupus	mupus
n	dudut	ndudut
n	tulis	nulis
ng	gawa	nggawa
ng	kethok	ngethok
ny	cuwil	nyuwil
ny	sikut	nyikut

b. Uji Hasil ater-ater Tripurasa

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

Tabel 2. . ater-ater Tripurasa

dak	pangan	dakpangan
dak	dhudhuk	dakdhudhuk
ko	jupuk	kojupuk
ko	goreng	kogoreng
di	ambung	diambung
di	batin	dibatin

c. Uji Hasil ater-ater Liyane

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

Tabel 4. . *Seselan*

-um	guyu	gumuyu
-um	ayu	kemayu
-in	serat	sinerat
-in	gawat	ginawa
-el	dewer	delewer
-el	kumpruk	kelumpruk
-er	cuwil	ceruwil
-er	kelip	kerelip

e. Uji Hasil *ater-ater Penambang*

Penambang adalah akhiran dalam bahasa Jawa ngoko. 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	turaa
weneh	-ana	wenehana
gawa	-na	gawakna
buku	-ku	bukuku
omah	-mu	omahmu
lepeh	-en	lepehen

3. Hasil Uji *Stemmer* Bahasa Jawa Ngoko

a. Hasil *ater-ater Hanuswara*

Hasil dari proses pengujian hasil *stemmer* Bahasa Jawa ngoko adalah *ater-ater Hanuswara*: m-, n-, ng-, ny-, belum bisa distemmer dengan hasil yang benar.

b. Hasil *ater-ater Tripurasa*

Hasil dari proses pengujian hasil *stemmer* Bahasa Jawa ngoko adalah *ater-ater Tripurasa*: Dak-, belum bisa distemmer dengan hasil yang benar. Sedangkan *ater-ater Ko-* dan *di-* berhasil mendapatkan term kata dasar benar

c. Hasil *ater-ater Liyane*

Hasil dari proses pengujian hasil *stemmer* Bahasa Jawa ngoko adalah *ater-ater liyane*: a-, belum bisa distemmer dengan hasil yang benar. Sedangkan *ater-ater ma, ka, ke, sa, pa, pi, pra, kuma, kami, kapi,* dan *tar* berhasil mendapatkan term kata dasar benar

d. Hasil *ater-ater Seselan*

Hasil dari proses pengujian hasil *stemmer* Bahasa Jawa ngoko adalah *seselan*: -um, -in, -el lan -er belum bisa distemmer dengan hasil yang benar

e. Hasil *ater-ater Penambang*

Hasil dari proses pengujian hasil *stemmer* Bahasa Jawa ngoko adalah *penambang*: -e, -ke, dan -a belum bisa distemmer dengan hasil yang benar. Sedangkan *penambang -I, -ake, -ane, -, -ana, -na, -ku, -mu, -en* berhasil mendapatkan term kata dasar benar

KESIMPULAN DAN SARAN

1. Kesimpulan

- a. *Stemmer* Bahasa Jawa ngoko mampu membuat kata dasar *jawa ngoko* dengan benar 62 % atau 21 dari 34 (*ater-ater/awalan, seselan/sisipan dan penambang/akhiran*)
- b. *Stemmer* Bahasa Jawa ngoko mampu melakukan proses *stemming* dengan hasil yang mudah di pahami karena dibuat terpisah antara awalan, kata dasar dan akhiran.

2. Saran

- a. Perlunya Perbaikan kode untuk: *ater-ater Hanuswara* (m-, n-, ng-, ny-), *ater-ater tripurasa* (dak-), *ater-ater liyane* (a-), *Seselan* (-um, -in, -el, -er) dan *Penambang* (-e, -ke, -a) sehingga akan menghasilkan *stemmer* Bahasa Jawa Ngoko dengan 100% benar
- b. Penggunaan kamus akan membantu proses *Stemming*

DAFTAR PUSTAKA

- Budi, I., Aji, R.F., 2006. Efektifitas Seleksi Fitur dalam Sistem Temu Kembali Informasi. Seminar Nasional Aplikasi Teknologi Informasi (SNATI), ISSN : 1907-5022.
- Bum, K.Y., 2010. *An autonomous assessment system based on combined latent semantic kernels. Expert Systems with Applications: An International Journal* , Volume 37 Issue 4.
- Deepika Sharma (2012) Stemming Algorithms: A Comparative Study and their Analysis, International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868 Foundation of Computer Science FCS, Volume 4– No.3, September 2012, New York, USA
- Kadir, A., 2001. Dasar Pemrograman Web Dinamis menggunakan PHP. Penerbit Andi. Yogyakarta.
- May Y. Al-Nashashibi, D.(2010), Stemming Techniques for Arabic Words: A Comparative Study, 2010 2nd International Conference on Computer Technology and Development (CCTD 2010)
- 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.
- Porter, M. F. (1980). An algorithm for suffix stripping'. Program, 14, 1304137.
- Salton, G., 1989, *Automatic Text Processing, The Transformation, Analysis, and Retrieval of information by computer*. Addison – Wesley Publishing Company, Inc. USA.
- Sandeep R. Sirsat (2013), Strength and Accuracy Analysis of Affix Removal Stemming Algorithms, International Journal of Computer Science and Information Technology, Vol 4 (2), IJCSIT
- Yates, R.B, 1999. *Modern Information Retrieval*, Addison Wesley-Pearson international edition, Boston. USA