

PENERAPAN TEKNIK STEMMING PADA BAHASA JAWA NGOKO BERBASIS ALGORITMA PORTER

¹Henis Cahyaning Kartika (0910651212), ²Wiwik Suharso, S.Kom, M.Kom
Jurusan Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember
Heniz.kartika@yahoo.com

ABSTRAK

Dalam Teknologi Informatika terdapat ilmu yang disebut dengan *Information Retrieval* (Temu Kembali Informasi). Didalam *Information Retrival* terdapat ilmu yang bertujuan untuk menemukan informasi baru yang sebelumnya tidak terungkap dan mencari kata-kata yang dapat mewakili apa yang ada dalam dokumen sehingga dapat dilakukan analisa keterhubungan antar dokumen, dengan memproses dan menganalisa data dalam jumlah besar berupa teks atau dokumen yang disebut dengan *Text Mining*. Dokumen yang dibutuhkan di sini adalah dokumen berbahasa Jawa, di dalam sebuah dokumen tersebut akan dicari kata dasarnya. Algoritma Stemming masih di bagi menjadi beberapa algoritma, salah satunya Algoritma Porter yang khusus menghilangkan ahiran dan awalan. Pada Tugas Akhir ini membahas tentang penerapan Stemming bahasa Jawa Ngoko berbasis algoritma porter bahasa Indonesia, dalam penerapan ini bentuk aplikasi yang dapat dijalankan di komputer sehingga mempermudah pengguna dalam pencarian bentuk dasar suatu kata.

Kata kunci: Text Mining, Algoritma Stemming, Algoritma Porter, Bahasa Jawa Ngoko.

ABSTRACT

Looking for a document based on the needs or desires relefan tekes a long time, usually users looking for specific information by reading word for word or document by document. In Information Technology are science called the Information Retrieval. Information contained in the retrival science that aims to discover new information not previously disclosed and search for words that can represent what is in the document so thet it can be analyed connectivity between document, whit processing and analyzing large amounts of data in the from of text or a document called by Text Mining. Documents required here is the Java language documents, in a document that will be searched said basically. Almost the same as Indonesia, in the Java language are also affixes prefixes, suffixes and inserts. Java language is divided into serveral specifications, which in this select is Ngoko Java language, since many documents encountered in the Java language using Ngoko. Algorithms in use is Stemming Algorithm, Stemming Algorithms is a process of converting a word and eliminate suffix and prefix in a word Indonesia language. In this final project is about the implementation of the Java language Ngoko Stemming algorithm based Indonesia porter, in this application form application that can be run on computers that allows users to search the basic form of a word.

Key words: Text Mining, Stemming Algorithm, Algorithm Porter, Ngoko Java language.

1. Latar Belakang

Bahasa sebagai alat komunikasi, sangat penting bagi semua manusia. Bahasa Jawa adalah bahasa yang berasal dari suku Jawa. Bahasa Jawa yang sering digunakan adalah bahasa Jawa Krama dan Ngoko. Bahasa Jawa Krama biasanya digunakan berbicara dengan kedua orang tua atau orang yang lebih tua, dan juga orang yang lebih tinggi kedudukannya. Bahasa Jawa Ngoko sendiri digunakan berbicara pada teman, atau orang yang lebih muda.

Dalam kemajuan teknologi saat ini bahasa jawa khususnya bahasa Jawa Ngoko kurang berkembang dan sedikit sekali dukungan yang didapat. Bahasa jawa ngoko sulit berkembang dan bersaing dengan bahasa lain dalam kemajuan teknologi, dikarenakan tingkat kata dan tata bahasa yang sulit digunakan.

Dibandingkan dengan bahasa lain Bahasa Jawa ngoko terdapat kerumitan yang sangat tinggi.

Perkembangan teknologi saat ini diharapkan dapat membantu untuk mengembangkan bahasa Jawa Ngoko, salah satu bidang ilmu yang dapat digunakan untuk pemrosesan kata adalah sistem temu kembali informasi, Sistem ini memiliki kemampuan untuk membentuk kata dasar (*Stemming*). Saat ini telah diterapkan algoritma porter, algoritma ini membentuk kata dasar dengan menghilangkan awalan dan akhiran. Oleh karena itu Tugas Akhir ini mengangkat judul tentang “**Penerapan Teknik Stemming pada Bahasa Jawa Ngoko Berbasis Algoritma Porter**”. Aplikasi ini diharapkan dapat mencari sebuah kata dasar dari masing-masing kata dalam sebuah dokumen bahasa Jawa Ngoko.

1.2 Rumusan Masalah

Perumusan masalah dalam skripsi ini yaitu

- a. Bagaimana penyesuaian algoritma stemming porter ke dalam bahasa Jawa Ngoko.
- b. Bagaimana pembentukan kata dasar dalam bahasa Jawa Ngoko.

1.3 Batasan Masalah

Batasan-batasan masalah dalam skripsi ini adalah

- a. Sumber dokumen yang digunakan berbentuk teks bahasa Jawa Ngoko dari berita, jurnal, soal ujian, dan cerita.
- b. Algoritma Stemming yang digunakan adalah penyesuaian dari Algoritma Porter Bahasa Indonesia.
- c. Aspek yang dipertimbangkan dalam Stemming adalah imbuhan berupa awalan dan akhiran.

1.4 Tujuan

- a. Aplikasi ini bertujuan untuk membentuk kata dasar atau kata baku yang sesuai dengan bahasa Jawa Ngoko dengan membuang imbuhan yang terdapat dimasing-masing kata.
- b. Membangun suatu sistem pembentukan kata dasar bahasa Jawa Ngoko dengan penyesuaian Algoritma Porter Stemming bahasa Indonesia.

1.5 Manfaat

- a. Kata dasar yang terbentuk dapat digunakan dalam operasi *text mining* bahasa Jawa dan dapat meningkatkan akurasi pencarian dalam Bahasa Jawa.
- b. Dapat membangun suatu sistem yang membentuk kata dasar Jawa Ngoko.

2. Landasan Teori

Tembung Andap(Imbuhan)

Tembung andhap merupakan ater-ater atau imbuhan pada bahasa Jawa.Tembung Andap pada bahasa Jawa Ngoko dapat di lihat di table berikut ini (Nirmala, 2012).

Tabel. 1 Awalan dan Akhiran Bahasa Jawa Ngoko

Awalan (ater-ater)	Akhiran (panambang)
a-	-i
ma-	-ku
ka-	-mu
ke-	-e
dak-	-an
tak-	-a
kok-	-na
di-	-ana
n-	-en
ny-	-ne
ng-	-ing
m-	-ake
sa-	
pa-	
pi-	
pra-	
pari-	
tar-	

kuma- kami- kapi-	
-------------------------	--

Text Mining

Text Mining adalah suatu proses yang bertujuan untuk menemukan informasi atau trem terbaru yang sebelumnya tidak terungkap, dengan memproses dan menganalisa data dalam jumlah besar. Dalam menganalisis sebagian atau keseluruhan teks yang tidak terstruktur (*Unstructured text*), *text mining* mencoba untuk mengasosiasikan satu bagian teks dengan yang lainnya berdasarkan aturan-aturan tertentu. Selain itu *text mining* juga diartika sebagai kegiatan menambang data dari data yang berupa teks atau dokumen, dengan tujuan mencari kata-kata yang dapat mewakili apa yang ada dalam dokumen sehingga dapat dilakukan analisa keterhubungan antar dokumen. Hasil yang diharapkan adalah informasi baru atau “*insight*” yang tidak terungkap jelas sebelumnya. Menurut para ahli *text mining*, *text mining* adalah proses penemuan kembali relasi dan fakta yang terkubur didalam teks, dan tidak harus baru (Sulistyo, 2008).

Case Folding

Case folding yaitu mengubah semua huruf dalam dokumen menjadi huruf kecil. Hanya huruf ‘a’ sampai dengan ‘z’ yang diterima (Choiroh, 2011).

Tokenizing

Tokenizing / parsing adalah tahap pemotongan string input berdasarkan tiap kata yang menyusunnya. Pemecahan kalimat menjadi kata-kata tunggal dilakukan dengan men-*scan* kalimat dengan pemisah *white space* (spasi, tab, *newline*) (Choiroh, 2011).

Wordlist/ Stoplist

Wordlist/stoplist adalah proses penyaringan (*filtering*) terhadap kata-kata yang tidak layak untuk dijadikan sebagai pembeda atau kata kunci sehingga kata-kata tersebut dapat dilakukan dari dokumen. *Wordlist* sendiri merupakan daftar kata-kata yang mungkin digunakan sebagai kata kunci dalam pencarian dokumen.Sedangkan *stoplist* berisi kata-kata yang tidak mungkin digunakan sebagai kata kunci dalam pencarian dokumen. Dengan demikian, maka tentu jumlah kata dalam *wordlist* akan lebih banyak daripada *stoplist* (Sulistyo, 2008).

Stemming

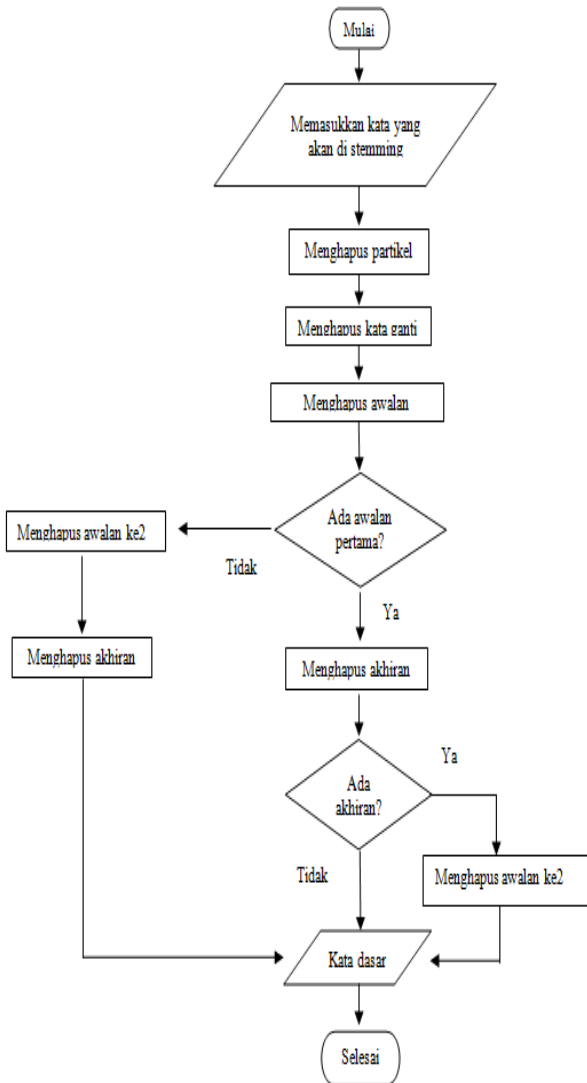
Setelah kata-kata yang terdapat dalam dokumen menjalani proses *tokenizing* dan *stoplist*, maka selanjutnya kata-kata yang tersisa akan menjalani proses *stemming*. Proses *stemming* bertujuan untuk mengubah kata dasarnya dengan menghilangkan imbuhan-imbuhan pada kata dalam

dokumen. Proses stemming dilakukan dengan mengecek kata apakah mengandung imbuhan atau tidak. Proses stemming kata dalam Bahasa Inggris memiliki karakteristik tersendiri, yang tidak lepas dari pengaruh tata bahasanya (Sulistyo, 2008).

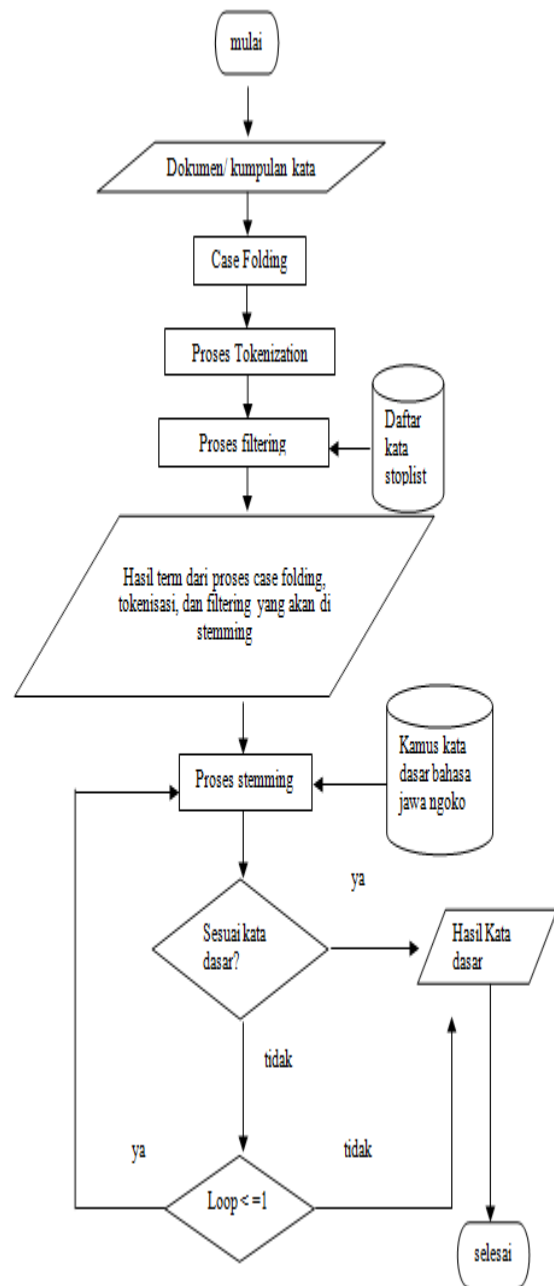
Algoritma Porter

Algoritma Porter merupakan Algoritma Stemming khusus bahasa Inggris yang ditemukan oleh Martin Porter 1980. Mekanisme algoritma dalam mencari kata dasar suatu kata berimbuhan dengan membuang imbuhan-imbuhan (atau lebih tepatnya akhirnya) pada kata-kata bahasa Inggris karena dalam bahasa Inggris tidak mengenal awalan. Karena bahasa Inggris datang dari kelas yang berbeda, beberapa modifikasi telah dilakukan untuk membuat Algoritma Porter sehingga dapat digunakan sesuai dengan bahasa Indonesia (Baskoro, 2012).

3. Metode Penelitian



Gambar 1 Proses Stemming Bahasa Jawa Ngoko



Gambar 1 Flowchat Rancangan Penelitian

4. Hasil dan Pembahasan



Gambar 2 Tampilan Stemming

Proses Stemming Cara :

1. Untuk cara yang pertama user dapat memasukkan kata, kalimat, paragraf, atau dokumen pada teks area sebelah kiri dengan cara mengkopi atau mengetik langsung, setelah itu terdapat tombol peroses yang nantinya akan memproses hasil case folding, tokenizing, dan filtering yang ditampilkan pada teks area sebelah kanan, dan juga langsung menghasilkan hasil peoses stemmingnya yang terdapat pada tab hasil stemming. Pada tampilan ini terdapat tombol proses yang berfungsi untuk menghapus isi pada teks area sebelah kiri dan sebelah kanan dan juga pada teks field jika ada. Dan juga ada tombol back yang berfungsi untuk kembali pada menu utama.
2. Untuk cara yang kedua user bisa mencari file dokumen berformat txt. yang sudah tersimpan pada komputer dengan cara klik tombol browser lalu cari dokumen yang akan di stemming, lalu klik open. Akan langsung keluar hasil case folding, tokenizing, filtering, dan hasil stemming.



Gambar 3 Hasil Stimming

Hasil proses stemming terdapat dalam tabel hasil stemming yang ditampilkan perkata seperti Gambar 3 diatas. Pada sebelah kanan terdapat imbuhan-imbuhan yang terdapat pada bahasa jawa ngoko, gunanya untuk membantu user memahami apa saja imbuhan-imbuhan yang terdapat dalam bahasa jawa ngoko. Pada hasil stemming ini terdapat tombol

clear yang berfungsi untuk menghapus hasil stemming dan tombol back untuk kembali pada menu utama.

- Rumus jumlah semua rata-rata yang di hitung dari tingkat keberhasilan dan kegagalan dari kata yang di uji.

$$\frac{\text{Jumlah tingkat keberhasilan}}{\text{jumlah kata yang berhasil di stemming}} \times \frac{\text{jumlah keseluruhan kata hasil stemming}}{\text{Jumlah tingkat kegagalan}} \times \frac{\text{jumlah kata yang gagal di stemming}}{\text{jumlah keseluruhan kata hasil stemming}}$$

- Rumus jumlah semua rata-rata yang dihitung dari semua rata-rata pada seluruh dokumen yang di uji cobakan.

$$\frac{\text{Jumlah rata-rata tingkat keberhasilan}}{\text{jumlah semua rata - rata kata yang berhasil di stemming}} \times \frac{\text{jumlah semua dokumen yang di uji}}{\text{Jumlah rata-rata tingkat kegagalan}} \times \frac{\text{jumlah semua rata - rata kata yang gagal di stemming}}{\text{jumlah semua dokumen yang di uji}}$$

Tabel 2 Contoh Kalimat yang di Stemming

No	Kalimat	hasil
1	Ing jaman sakwise, yaiku jaman Jenggala kesenian wayang purwa kuwi disampurnakake wujud karo dening Raja Jenggala Raden Panji Rawisrengga sing bergelar Sri Suryawisesa.	jaman jaman jenggala s wayang purwa mpurnak wujud raja jenggala raden panji rawisrengga bergelar sri suryawisesa
2	Kocap cinarita terus wayang – wayang sing wis kawujud mau banjur diklumpukake lan disimpen ing jeroning peti “khusus” sing “nyeni” indah.	kocap cinarit wayang wayang wujud klumpuk simp jero peti khusus s indah
3	Rikala jaman Majapahit gegambarane kesenian wayang kuwi saya luwih disampurnaake lan dadi tambah apike, amarga wis ditambah bagian kana-kene sing disik – disikane dadi kekurangane, terus digulung dadi siji.	rikala jaman majapahit gegambar s wayang mpur tambah p tambah bag k disik sik kurang gulung siji

4.1 Hasil Pengujian Kata

Rumus kata yang berhasil di stemming:

$$\frac{\text{jumlah kata yang berhasil di stemming}}{\text{jumlah keseluruhan kata hasil stemming}} = \frac{8}{10} = 0,8$$

Rumus kata yang gagal di stemming:

$$\frac{\text{jumlah kata gagal di stemming}}{\text{jumlah keseluruhan kata hasil stemming}} = \frac{2}{10} = 0,2$$

4.2 Hasil Pengujian Kalimat

Kata yang berhasil di stemming:

Kalimat 1

$$\frac{\text{jumlah kata yang berhasil di stemming}}{\text{jumlah keseluruhan kata hasil stemming}} = \frac{14}{16} = 0,875$$

Kalimat 2

$$\frac{\text{jumlah kata yang berhasil di stemming}}{\text{jumlah keseluruhan kata hasil stemming}} = \frac{9}{12} = 0,75$$

Kalimat 3

$$\frac{\text{jumlah kata yang berhasil di stemming}}{\text{jumlah keseluruhan kata hasil stemming}} = \frac{11}{17} = 0,647$$

Kalimat 4

$$\frac{\text{jumlah kata yang berhasil di stemming}}{\text{jumlah keseluruhan kata hasil stemming}} = \frac{7}{7} = 1$$

Kalimat 5

$$\frac{\text{jumlah kata yang berhasil di stemming}}{\text{jumlah keseluruhan kata hasil stemming}} = \frac{9}{11} = 0,818$$

Kalimat 6

$$\frac{\text{jumlah kata yang berhasil di stemming}}{\text{jumlah keseluruhan kata hasil stemming}} = \frac{9}{11} = 0,818$$

Kata yang gagal di stemming:

Kalimat 1

$$\frac{\text{jumlah kata yang gagal di stemming}}{\text{jumlah keseluruhan kata hasil stemming}} = \frac{2}{16} = 0,125$$

Kalimat 2

$$\frac{\text{jumlah kata yang gagal di stemming}}{\text{jumlah keseluruhan kata hasil stemming}} = \frac{3}{12} = 0,25$$

Kalimat 3

$$\frac{\text{jumlah kata yang gagal di stemming}}{\text{jumlah keseluruhan kata hasil stemming}} = \frac{6}{17} = 0,352$$

Kalimat 4

$$\frac{\text{jumlah kata yang gagal di stemming}}{\text{jumlah keseluruhan kata hasil stemming}} = \frac{0}{7} = 0$$

Kalimat 5

$$\frac{\text{jumlah kata yang gagal di stemming}}{\text{jumlah keseluruhan kata hasil stemming}} = \frac{2}{11} = 0,181$$

Kalimat 6

$$\frac{\text{jumlah kata yang gagal di stemming}}{\text{jumlah keseluruhan kata hasil stemming}} = \frac{2}{11} = 0,181$$

Rata-rata jumlah keseluruhan kalimat yang berhasil di stemming

$$\frac{\text{jumlah semua rata-rata kata yang berhasil di stemming}}{\text{jumlah semua kalimat yang di uji}} = \frac{0,875 + 0,75 + 0,647 + 1 + 0,818 + 0,818}{6} = \frac{4,908}{6} = 0,818$$

Rata-rata jumlah keseluruhan kalimat yang gagal di stemming

$$\frac{\text{jumlah semua rata-rata kata yang gagal di stemming}}{\text{jumlah semua kalimat yang di uji}} = \frac{0,125 + 0,25 + 0,352 + 0 + 0,181 + 0,181}{6} = \frac{1,089}{6} = 0,1815$$

4.3 Hasil Pengujian Paragraf

Kata yang berhasil di stemming:

Paragraf 1

$$\frac{\text{jumlah kata yang berhasil di stemming}}{\text{jumlah keseluruhan kata hasil stemming}} = \frac{25}{30} = 0,833$$

Paragraf 2

$$\frac{\text{jumlah kata yang berhasil di stemming}}{\text{jumlah keseluruhan kata hasil stemming}} = \frac{43}{53} = 0,811$$

Paragraf 3

$$\frac{\text{jumlah kata yang berhasil di stemming}}{\text{jumlah keseluruhan kata hasil stemming}} = \frac{36}{45} = 0,822$$

Kata yang berhasil di stemming:

Paragraf 1

$$\frac{\text{jumlah kata yang gagal di stemming}}{\text{jumlah keseluruhan kata hasil stemming}} = \frac{5}{30} = 0,167$$

Paragraf 2

$$\frac{\text{jumlah kata yang gagal di stemming}}{\text{jumlah keseluruhan kata hasil stemming}} = \frac{10}{53} = 0,188$$

Paragraf 3

$$\frac{\text{jumlah kata yang gagal di stemming}}{\text{jumlah keseluruhan kata hasil stemming}} = \frac{9}{45} = 0,176$$

Rata-rata jumlah keseluruhan kalimat yang berhasil di stemming

$$\frac{\text{jumlah semua rata - rata kata yang berhasil di stemming}}{\text{jumlah semua paragraf yang di uji}} = \frac{0,833 + 0,811 + 0,822}{3} = \frac{2,466}{3} = 0,822$$

Rata-rata jumlah keseluruhan kalimat yang gagal di stemming

$$\frac{\text{jumlah semua rata - rata kata yang gagal di stemming}}{\text{jumlah semua paragraf yang di uji}} = \frac{0,167 + 0,188 + 0,176}{3} = \frac{0,531}{3} = 0,177$$

4.4 Hasil Pengujian dokumen

Kata yang berhasil di stemming:

Dokumen 1

$$\frac{\text{jumlah kata yang berhasil di stemming}}{\text{jumlah keseluruhan kata hasil stemming}} = \frac{103}{128} = 0,8046$$

Dokumen 2

$$\frac{\text{jumlah kata yang berhasil di stemming}}{\text{jumlah keseluruhan kata hasil stemming}} = \frac{98}{122} = 0,8196$$

Kata yang gagal di stemming:

Dokumen 1

$$\frac{\text{jumlah kata yang gagal di stemming}}{\text{jumlah keseluruhan kata hasil stemming}} = \frac{25}{128} = 0,1953$$

Dokumen 2

$$\frac{\text{jumlah kata yang berhasil di stemming}}{\text{jumlah keseluruhan kata hasil stemming}} = \frac{24}{122} = 0,1803$$

Rata-rata jumlah keseluruhan kalimat yang berhasil di stemming

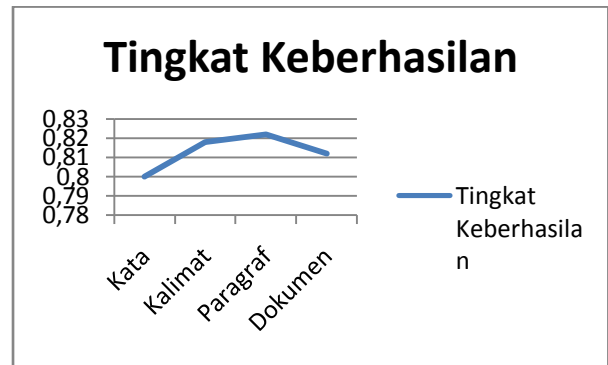
$$\frac{\text{jumlah semua rata - rata kata yang berhasil di stemming}}{\text{jumlah semua paragraf yang di uji}} = \frac{0,8046 + 0,8196}{2} = \frac{1,6242}{2} = 0,8121$$

Rata-rata jumlah keseluruhan kalimat yang gagal di stemming

$$\frac{\text{jumlah semua rata - rata kata yang gagal di stemming}}{\text{jumlah semua paragraf yang di uji}} = \frac{0,1875 + 0,1803}{2} = \frac{0,3678}{2} = 0,1839$$

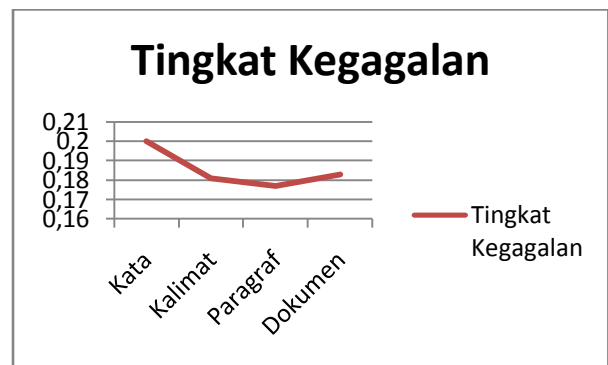
Tabel 3 Tingkat Keberhasilan dan Tingkat Kegagalan

No	Berdasarkan	Tingkat Keberhasilan	Tingkat Kegagalan
1	Kata	0.8	0.2
2	Kalimat	0.818	0.181
3	Paragraf	0.822	0.177
4	Dokumen	0.812	0.183



Gambar 4.1 Grafik Tingkat Keberhasilan

Dari gambar grafik diatas dapat disimpulkan bahwa inputan paragraf lebih banyak tingkat kebenarannya, dan tingkat kebenaran paling sedikit pada inputan kata.



Gambar 4.2 Grafik Tingkat Kegagalan

Dari gambar grafik diatas dapat disimpulkan bahwa tingkat kegagalan paling banyak terdapat pada inputan kata, dan tingkat kegagalan paling rendah terdapat pada inputan paragraf.

5. KESIMPULAN

Dari uji coba dan analisa yang telah dijelaskan dalam bab sebelumnya, maka penulis mengambil beberapa kesimpulan sebagai berikut :

1. Lama atau tidaknya waktu pemrosesan aplikasi untuk menemukan kata dasar tergantung panjang dan banyaknya kata, kalimat, paragraf, dan dokumen.
2. Dari perbandingan uji coba stemming kata, kalimat, paragraf dan dokumen tingkat keberhasilan yang paling banyak terdapat pada inputan paragraf, dan tingkat keberhasilan yang paling sedikit pada inputan kata.

6. DAFTAR PUSTAKA

- [1] Baskoro, D.O., Malik, H., Anshari, M.H., (2012), *Porter Stemmer Information Retrieval*, Universitas Gadjah Mada, Yogyakarta.
- [2] Choiroh, U., (2011), *Pembuatan Aplikasi Penerjemah kalimat Tunggal Bahasa Indonesia Kedalam Kalimat Bahasa Jawa Berbasis Kamus*, Universitas Islam Negeri (UIM) Maulana Malik Ibrahim, Malang.
- [3] Nirmala, (2012), *Pintar Bahasa Jawa*, Giri Utama, Surabaya.
- [4] Sulistyio, W., Sarno, R., (2008), *Auto Matching Antar Dokumen dengan Metode Cosine Measure*, Unversitas Kristen Satya Wacana dan Institut Teknologi Sepuluh Nopember (ITS), Jawa Tengah dan Surabaya.