

PENINGKATAN ALGORITMA STEMMING PORTER BAHASA INDONESIA BERDASARKAN METODE MORFOLOGI DENGAN MENGAPLIKASIKAN ATURAN KOMBINASI AWALAN AKHIRAN

Lutfi Ali Muharrom, M.Si,¹Erma Supitasari (1210651022)²,

Jurusan teknik informaika fakultas teknik universitas muhammadiyah jember

E-mail: supitasarierma@gmail.com

Abstract___ Stemming algorithm has advantages and disadvantages . Stemming algorithm effectiveness can be measured by several parameters , such as the speed of the process and have a weakness in terms of accuracy , and errors stemming stemming.Metode using morphological word has several drawbacks such as the right not omit the prefix on the basis of words beginning with the letter " k " , " t " , " s " , " p " and is not appropriate in removing the suffix especially for the suffix " kan " and " an " . porter stemmer for stemming the increase in the Indonesian language by utilizing two Indonesian morphology . Thus , stemming process can be expedited . In this final task performed 10 trials on a document drawn abstract journals only course and obtained an accuracy of 50.83 % , precission 50.83 % , and 75.80 % recall .

Intisari___ Algoritma *stemming* memiliki kelebihan dan kekurangannya. Efektifitas algoritma *stemming* dapat diukur berdasarkan beberapa parameter, seperti kecepatan proses dan memiliki kelemahan dalam hal keakuratan, dan kesalahan *stemming*.Metode *stemming*dengan menggunakan morfologi suatu kata memiliki beberapa kekurangan seperti tidak tepat menghilangkan awalan pada kata dasar yang berawalan huruf “k”, “t”, “s”, “p” serta tidak tepat dalam menghilangkan akhiran terutama untuk akhiran “kan” dan “an”.Modifikasi *porter stemmer* untuk *stemming* pada kata Bahasa Indonesia dengan memanfaatkan dua morfologi Bahasa Indonesia. Jadi, proses *stemming* dapat lebih cepat dilakukan. Dalam tugas akhir ini dilakukan 10x uji coba pada dokumen jurnal yang hanya diambil abstarknya saja dan didapatkan akurasi sebesar 50,83%, *Precission* 50,83%, dan *recall* 75,80%.

Kata Kunci:algoritma, stemming, morfologi, modifikasi *porter stemmer*.

1. Latar Belakang

Bahasa Indonesia merupakan bahasa utama yang digunakan masyarakat Indonesia untuk berkomunikasi, baik secara lisan maupun tertulis. Dalam kehidupan sehari-hari mungkin tidak terlalu diperhatikan penulisan

kata yang benar, namun dalam penggunaan surat resmi kata dalam Bahasa Indonesia menjadi sangat penting. Selain untuk berkomunikasi, penggunaan Bahasa Indonesia yang baik dan benar juga sangat diperlukan dalam penulisan-penulisan dokumen formal,

jurnal, laporan dan sebagainya. Penggunaan bahasa yang benar menjadi unsur penting dalam Bahasa Indonesia, kamus Besar Bahasa Indonesia menjadi acuan utama dalam penggunaan kata baku Berbahasa Indonesia.

Pencarian informasi berupa dokumen teks atau yang dikenal dengan istilah *Information retrieval* (IR) merupakan proses pemisah dokumen-dokumen yang dianggap relevan dari sekumpulan dokumen yang tersedia. Bertambahnya jumlah dokumen teks yang dapat diakses diinternet dapat diikuti dengan meningkatnya kebutuhan pengguna akan perangkat pencarian informasi yang efektif dan efisien. Efektif berarti user mendapatkan dokumen yang relevan dengan *query* yang diinputkan. Efisien berarti waktu pencarian yang sesingkat-singkatnya.

Stemming adalah salah satu cara yang digunakan untuk meningkatkan performa IR (*Information Retrieval*) dengan cara mentransformasi kata-kata dalam sebuah teks ke kata dasarnya. Algoritma *stemming* untuk bahasa yang satu berbeda dengan algoritma *stemming* untuk bahasa lainnya. Proses *stemming* pada teks Bahasa Indonesia lebih rumit karena terdapat variasi imbuhan yang harus dibuang untuk mendapatkan *root word* dari sebuah kata. Beberapa algoritma *stemming* bahasa indonesia telah dikembangkan sebelumnya. Penggunaan algoritma *stemming* yang sesuai mempengaruhi performa IR.

Algoritma *stemming* memiliki kelebihan dan kekurangannya. Efektifitas algoritma *stemming* dapat diukur berdasarkan beberapa parameter, seperti kecepatan proses dan memiliki kelemahan dalam hal keakuratan, dan kesalahan *stemming*. Hasil dari *stemming* disebut token. Salah satu keuntungan menggunakan *stemming* dalam pengembangan sistem IR adalah efisiensi dan index file yang sudah terkomputasi.

Dalam Bahasa Indonesia ada beberapa penelitian yang dilakukan yang berhubungan dengan *stemming* Bahasa Indonesia. Semua penelitian yang telah dilakukan mengubah bentuk kata menjadi kata dasar dengan berdasarkan morfologi dari suatu kata. Penelitian sebelumnya yang dilakukan oleh Baskoro (2013) dari Ilmu Komputer Universitas Gadjah Mada dengan melihat struktur dari satu kata . Penelitian yang dilakukan oleh Baskoro lebih dulu melihat akhiran partikel dan kata ganti milik yang merupakan aturan struktur morfologi Bahasa Indonesia Infleksional kata ganti dan Derivasional sufik dibandingkan melihat awalan dari suatu kata, sedangkan masih banyak lagi aturan-aturan dari morfologi Bahasa Indonesia dari kedua Struktur tersebut.

Algoritma ini menjadi tidak efisien karena diperlukan masing-masing 2 tabel di *database* untuk kata berakhiran partikel dan kata ganti milik. Selain itu kekurangan dari penelitian Baskoro adalah Algoritma ini tidak memperhitungkan kata dasar yang berawalan huruf “k”, “t”, “s”, “p” dengan benar. Dalam Bahasa Indonesia kata yang diawali dengan huruf “k”, “t”, “s”, “p” jika diberi awalan “me-“, atau “pe-“ maka huruf depannya akan melebur sebagai contoh : “mengacau” menjadi “me”+”kacau”. Jika kata “mengacau” dicari kata dasarnya dengan menggunakan *stemming porter* maka kata dasar yang muncul adalah “acau”. Kata ini tentu tidak bisa disamakan dengan kata “mengantar” menjadi “me” + “antar” walaupun secara struktur sama.

Hal ini tidak tepat jika digunakan untuk mencari kata dasar yang secara terstruktur. Metode *stemming* dengan menggunakan morfologi suatu kata memiliki beberapa kekurangan seperti tidak tepat menghilangkan awalan pada kata dasar yang berawalan huruf “k”, “t”, “s”, “p” serta tidak tepat dalam menghilangkan akhiran terutama untuk akhiran “kan” dan “an”. Untuk menyelesaikan masalah ini, penelitian ini menawarkan penggunaan 2 tingkat morfologi pada kata dasar berawalan huruf “k”, “t”, “s”, “p” serta menggunakan aturan kombinasi awalan dan akhiran, untuk menghilangkan akhiran pada suatu kata seperti awalan “di” hanya boleh dipasangkan dengan akhiran “kan” dan tidak boleh dengan akhiran “an”. Hasil dari penelitian ini adalah *stemming* yang memiliki tingkat akurasi 95,5%, lebih baik dibandingkan *stemming* sebelumnya yang menggunakan algoritma berdasarkan morfologi satu kata. *Stemming* sebelumnya yang menggunakan algoritma berdasarkan morfologi suatu kata memiliki tingkat akurasi 82,5%.

Berdasarkan penelitian yang sudah ada, penulis mencoba melakukan penelitian bagaimana menentukan kata dasar pada teks bahasa indonesia dengan menggunakan **“peningkatan algoritma *stemming porter* bahasa indonesia berdasarkan metode morfologi dengan mengaplikasikan 2 tingkat morfologi dan aturan kombinasi awalan dan akhiran”**. Penulis ingin melakukan analisis akurasi algoritma “*stemming porter* dengan mengaplikasikan morfologi dan aturan kombinasi awalan dan akhiran” pada teks bahasa indonesia.

2. Tinjauan Pustaka

2.1 Morfologi Bahasa Indonesia

Morfologi disebut juga ilmu bahasa yang mempelajari seluk beluk kata. Morfologi adalah bidang linguistik yang mempelajari

susunan bagian kata secara gramatikal (Verhaar, 1984:52). Selain itu morfologi yaitu bidang linguistik yang mempelajari morfem dan kombinasi-kombinasinya. Bagian dari struktur bahasa yang mencakup kata dan bagian-bagian kata, yaitu morfem (Kridalaksana, 1984:129).

Menurut Ramlan (1997) mendefinisikan morfologi adalah bagian dari ilmu bahasa yang membicarakan atau mempelajari tentang seluk beluk bentuk kata serta pengaruh perubahan-perubahan bentuk kata terhadap golongan dan arti kata. Morfologi dalam bahasa Indonesia terdiri dari infleksional dan derivasional (Tala, 2003). Berdasarkan definisi-definisi tersebut, dapat disimpulkan bahwa morfologi adalah bidang linguistik yang mempelajari hubungan antara morfem yang satu dengan morfem yang lain untuk membentuk sebuah kata.

Morfem adalah bentuk bahasa yang terkecil yang tidak dapat lagi dibagi menjadi bagian-bagian yang lebih kecil, misalnya, kata putus jika dibagi menjadi "pu" dan "tus", bagian-bagian itu tidak dapat lagi disebut morfem karena tidak memunyai makna, baik makna leksikal ataupun makna gramatikal. Demikian juga "me-" dan "-kan" tidak dapat kita bagi menjadi bagian yang lebih kecil (Badudu, 1958:66).

2.1.1 Infleksional

Infleksional adalah struktur sederhana yang diikuti oleh imbuhan yang tidak mempengaruhi kata dasar. Struktur ini dibagi menjadi 2, yaitu *sufiks* partikel dan kata ganti. Partikel berfungsi untuk memberi penekanan pada kata. Macam-macam dari partikel adalah "-lah", "-kah", "-tah", "-pun". Partikel ditulis dengan serangkai dengan kata yang mendahuluinya, kecuali partikel "-pun" yang ditulis terpisah. Kata ganti terdiri dari "-ku", "-mu", dan "-nya".

Tiap *sufiks* pada kelompok partikel dan kata ganti mungkin saja terjadi pada kata yang sama. Bila hal ini terjadi, maka harus mengikuti aturan *sufiks* pada kata ganti mendahului partikel. Sehingga struktur penulisan infleksional dapat disimpulkan seperti pada struktur dibawah ini.

Infleksional = (kata dasar + kata ganti) | (kata dasar + partikel) | (kata dasar + kata ganti + partikel)

2.1.2 Derivasional

Derivasional adalah struktur sederhana yang diikuti imbuhan yang dapat mempengaruhi kata dasar. Derivasional pada struktur Bahasa Indonesia terdiri dari *prefix*, *suffix*, dan *konfix*.

Prefix yang sering muncul antara lain "ber-", "di-", "ke-", "meng-", "per-", dan "ter-".

Beberapa *prefix* seperti "ber-", "meng-", "peng-", "per-", "ter-" dapat terjadi pada beberapa bentuk. Bentuk tersebut tergantung dari huruf pertama dari kata dasar. Pada struktur kata derivasional terdapat proses peluluhan yaitu mengalami perubahan pengucapan, seperti kata sapu menjadi menyapu yang terdiri dari *prefix* "meng-" dan kata dasar sapu.

suffix derivasional antara lain "i", "kan", dan "an", penambahan pada *suffix* berbeda dengan *prefix*, karena tidak pernah mengubah pengucapan pada kata turunannya. *konfix* merupakan gabungan dari *prefix* dan *suffix*, dimana *prefix* dan *suffix* disisipkan bersama kata turunan yang baru. Namun, tidak semua kombinasi dari *prefix* dan *suffix* dapat digabungkan jadi *konfix*. Ada beberapa kombinasi dari *prefix* dan *suffix* yang tidak dibolehkan seperti yang di tunjukkan pada tabel 2.1.

Tabel 2.1 Pasangan konfiks yang tidak diperbolehkan

<i>Prefix</i>	<i>Suffix</i>
Ber	I
Di	an
Ke	i kan
Meng	an
Peng	i kan
Ter	an

prefix atau *konfix* dapat ditambahkan pada kata yang sudah ditambahkan *prefix* atau *konfix* sebelumnya, sehingga akan menghasilkan struktur *prefix* yang ganda. Namun, tidak semua *prefix* atau *konfix* dapat di tambahkan pada *prefix* atau *konfix* tertentu untuk membentuk *refix* ganda. Aturan untuk menentukan *prefix* ganda ditunjukkan pada tabel 2.2.

Tabel 2.2 Aturan urutan untuk *prefix* ganda

<i>Prefix</i> 1	<i>Prefix</i> 2
Meng	Per
Di	Ber
Ter	
Ke	

Dari penjelasan diatas, dapat disimpulkan struktur derivasional didefinisikan seperti dibawah ini.

Derivasional = *prefix* | *sufiks* | *konfix* | *prefix* ganda
Dimana :

Prefiks = prefiks
 + kata dasar = kata dasar + sufiks
 Sufiks = kata dasar + sufiks
 Konfiks = prefiks + kata dasar + sufiks
 Prefiks ganda = (prefiks₂ + prefiks₁) | (prefiks + konfiks)

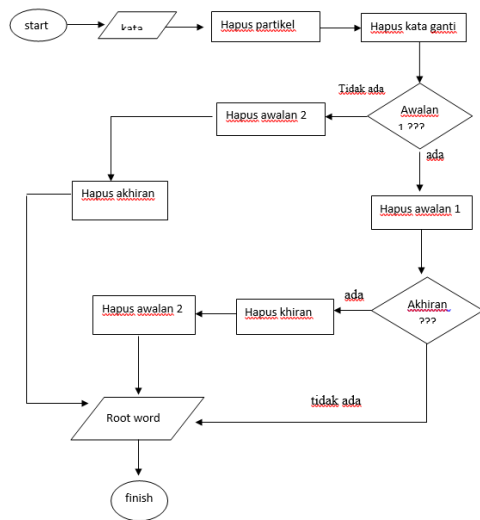
2.1.3 Gabungan Infleksional dan Derivasional

Dari uraian struktur morfologi infleksional dan derivasional, maka didapat struktur morfologi Bahasa Indonesia secara umum, yaitu:

[[Pre₁]] + [[Pre₂]] + KD + [[Suf]] + [[KG]] + [[Par]]

Dimana:
 Pre₁ : prefiks 1
 Pre₂ : prefiks 2
 KD : kata dasar
 Suf : sufiks
 KG : kata ganti
 Par : partikel

3. Diagram Alir proses stemming porter pada Bahasa Indonesia



Gambar 3.5 Flowchart stemmer porter pada Bahasa Indonesia

Berikut adalah keterangan dari langkah-langkah proses stemming pada Bahasa Indonesia:

1. Input Kata
Langkah awal pada penelitian ini yang pertama adalah memasukkan kata yang akan di *stem* pada kolom input kata.
2. Hapus partikel
Jika terdapat partikel “kah”, “tah”, “pun” maka hapus partikel.
3. Hapus kata ganti
Pada langkah ini jika terdapat kata kepemilikan “ku”, “mu”, “nya” maka, hapus kata ganti.
4. Awalan 1

Periksa apakah terdapat prefiks (awalan) “me”, “di”, “be”, “pe”, “ke”, “ter”, “se” atau tidak.

5. Hapus awalan 1
Apabila terdapat prefiks (awalan) seperti diatas maka hapus prefiks.
6. Hapus akhiran
Apabila tidak terdapat akhiran maka hapus akhiran. Selanjutnya hapus awalan 2 an akan ditemukan *root word* atau kata dasar.
7. Hapus awalan 2
Apabila tidak terdapat awalan 1 maka akan langsung ke langkah selanjutnya menghaus awalan 2 dan hapus akhiran dan akan ditemukan *root word* atau kata dasar.

4. Implementasi Sistem

4.1 Tampilan Utama Sistem

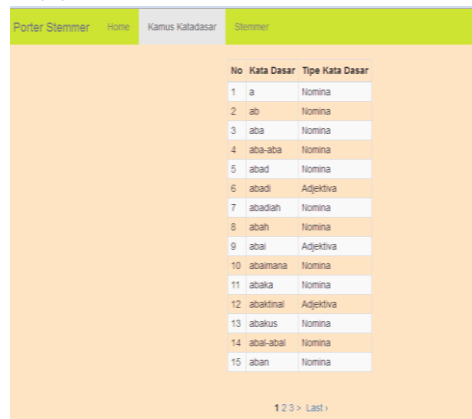
Halaman menu utama adalah halaman yang pertama kali muncul pada saat sistem dijalankan. Halaman utama menampilkan nama aplikasi dan penyusunan penelitian, halaman utama dapat dilihat seperti gambar dibawah ini:



Gambar 4.1 Halaman Utama

4.2 Halaman kamus kata dasar

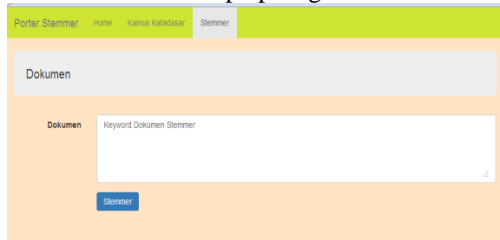
Pada halaman kamus kata dasar ini user dapat melihat kumpulan kata dasar yang sudah ditentukan oleh Kamus Besar Bahasa Indonesia, pada kamus besar ini tidak langsung semua kata dasar ditampilkan pada 1 halaman namun ada tombol *next* untuk melihat kata dasar selanjutnya. Tampilan halaman menu kamus kata dasar dapat dilihat pada gambar 4.2.



Gambar 4.2 Menu Kamus kata dasar

4.3 Halaman Menu Stemmer

padahalaman *stemmer* ini akan dilakukan pengujian pada kalimat untuk menentukan kata dasar. Halaman ini dapat dilihat pada gambar 4.3



Gambar 4.3 menu stemmer

Padahalaman menu *stemmer* di atas akan digunakan untuk menguji kalimat agar dapat ditemukan kata dasar dari setiap kata tersebut.

5. Kesimpulan dan Saran

5.1 Kesimpulan

Berdasarkan pembahasan yang telah dilakukan pada tugas akhir ini diperoleh beberapa kesimpulan, antara lain:

1. Sistem ini sudah dapat digunakan oleh pengguna untuk menentukan kata dasar pada Bahasa Indonesia yang sesuai dengan Kamus Besar Bahasa Indonesia.
2. Didapatkan ketepatan dan kecepatan dalam menentukan kata dasar tanpa menggunakan Kamus Kata Dasar secara manual.
3. Hasil dari akurasi 62,5% , *precision* 62,5%, dan *recall* 87%.

5.2 Saran

Dalam pengembangan penelitian ini penulis memberikan beberapa saran, yaitu:

1. sistem yang dikembangkan disarankan dapat dirancang di platform lainnya, seperti mobile sehingga akan menjadi lebih luas dan bisa digunakan oleh setiap orang.
2. Untuk menghasilkan akurasi stemmer dapat dibandingkan dengan algoritma stemmer yang lain seperti andriana.

DAFTAR PUSTAKA

- Abdul, Kadir. (2005). *Pengenalan Teknologi Informasi*, Yogyakarta, Penerbit Andi.
- Adisantoso, J., Wigena, A. H., & Akhmadi, C. H. (2003). *Algoritma Pemotongan Akhiran*

Baku untuk Kata dalam Bahasa Indonesia Berbasis Algoritma Porter. *Majalah Ilmiah – Ilmu Komputer*, vol.1 No., -8.

Agusta, Ledy. (2009). Perbandingan Algoritma Stemming Porter Dengan Algoritma Nazief & Adriani Untuk Stemming Dokumen Teks Bahasa Indonesia. *Konferensi Nasional Sistem dan Informatika 2009*, Bali, November 14, 2009.

Ali, N. H., & Ibrahim, N. S. (2012). Porter Stemming Algorithm for Semantic Checking. *Proceedings of International Conference on Computer and Information Technology (ICCTI) 2012*, (hal. 253-258).

Alwi, Hasan ., (1998) *Tata Bahasa Baku Indonesia*. Jakarta: Departemen Pendidikan dan Kebudayaan Indonesia.

Gupta, V. (2014). Suffix Stripping Based Verb Stemming for Hindi. *International Journal of Advanced Research in Computer Science and Software Engineering Volume 4, Issue I, January, 179*.

Haspari Kembang, R., & Santoso Juli, Y., (2015). Stemming Artikel Bahasa Indonesia Dengan Pendekatan Confix-Stripping. *Prosiding Seminar Nasional Manajemen Teknologi XXII Program Studi MMT-ITS, Surabaya 24 januari 2015*.

Kartika, H.C. (2003). Penerapan Teknik Stemming Pada Bahasa Jawa Ngoko Berbasis Algoritma Porter . . Skripsi . Jember : Universitas Muhammadiyah Jember.

Kridlksana, Hrimurti.(1996). *Pembentukan Kata dalam Bahasa Indonesia*. Gramedia Pustaka Utama.

- Lestiyanto, Dony Yudha (2014) *stemming* Bahasa Indonesia Sebagai Media Belajar Siswa Sekolah Menggunakan Algoritma Porter .
 . Universitas Dian
 Nuswantoro Semarang
- Putra Pandu Adikara.2012
<http://hikaruyuuki.lecture.ub.ac.id/kamus-kata-dasar-dan-stopword-list-bahasa-indonesia/> di unduh pada tanggal (25 Mei 2016)
- Ramlan, M. (1997). *Ilmu Bahasa Indonesia: Morfologi Suatu Tinjauan Deskriptif*. Yogyakarta: CV Karyono.
- Sembok, T. M., & Ata, B. A. (2013). Arabic Word Stemming Algorithms and Retrieval Effectiveness. *Proceedings of the Word Congress on Engineering Vol III, WCE 2013, July 3-5, 2013*. London , U.K..
- Tala, F. Z. (2003). *A Study of Stemming Effects on Information Retrieval in Bhasa Indonesia*. Amsterdam: Institute for Logic, Language and Computation, Universiteit van Amsterdam.
- Wiguna, Putu Bagus Susastra & Hartono, Bimo Sunafri. (2013). Peningkatan Algoritma Porter Stemmer Bahasa Indonesia Berdasarkan Metode Morfologi dengan Mengaplikasikan 2 Tingkat Morfologi dan Aturan Kombinasi Awalan dan Akhiran. *JNTETI*, Vol. 2, No. 2, Mei 2013.
- Zaman, B., & Winarko, E. (201). Analisis Fitur Kalimat untuk Peringkat Teks Otomatis pada Bahasa Indonesia. *Indonesian Journal of Computing and Cybernatics Systems (IJCSS)*, 60-68.
- Zaman, Badrus. (2014). Modifikasi Algoritma Porter Untuk Stemming Pada Kata Bahasa Indonesia. *Seminar Nasional Teknologi Informasi dan Komunikasi 2014 (SENTIKA 2014) Yogyakarta, 15 Maret 2014*.