

## PENERAPAN KAMUS DASAR PADA ALGORITMA PORTER UNTUK MENGURANGI KESALAHAN STEMMING BAHASA INDONESIA

Nurul Justina Mahardianing Verdaningroem<sup>1</sup>, Aries Saifudin<sup>2,\*</sup>

<sup>1,2</sup>Program Studi Teknik Informatika, Fakultas Teknik, Universitas Pamulang, Tangerang

Jl. Puspipetek No.23, Buaran, Serpong, Kota Tangerang Selatan, Banten 15310

\*Email: [aries.saifudin@yahoo.co.id](mailto:aries.saifudin@yahoo.co.id)

Diterima: 15 Maret 2018

Direvisi: 26 April 2018

Disetujui: 21 Juni 2018

### ABSTRAK

Proses *stemming* porter memerlukan waktu lebih sedikit dibanding dengan algoritma *stemming* lainnya, tetapi algoritma tersebut mempunyai kekurangan yaitu pada hasil akhir *stemming* menghasilkan kata yang selalu tepat. Proses *stemming porter* masih berdasarkan dari aturan morfologi bahasa Indonesia. Pada penelitian ini diusulkan penambahan kamus pada algoritma *stemming Porter* agar dalam proses *stemming* dapat mengurangi keambiguan dan meningkatkan ketepatan hasil. Kamus yang ditambahkan akan digunakan untuk pengecekan pada setiap kata yang telah melewati proses pemotongan imbuhan, kemudian hasil dari pemotongan langsung dicocokkan dengan kamus yang berisi kata dasar. Penggunaan kamus dasar pada algoritma Porter dapat meningkatkan ketepatan *stemming* sebesar 13,333%. Tetapi kecepatan proses *stemming* menggunakan kamus jauh lebih lambat dibandingkan tanpa kamus.

**Kata kunci:** *Stemming, Porter, Kamus, Bahasa Indonesia*

### ABSTRACT

*Porter stemming process requires less time than the other stemming algorithms, but these algorithms have the disadvantage that the final results stemming generate word is always right. Porter stemming process is still based on the morphology of Indonesian rule. In this study, the proposed addition of the dictionary at the Porter stemming algorithm so that in the process of stemming can reduce ambiguous and improve the accuracy of the results. Dictionaries are added to be used for checking on every word that has gone through the process of cutting the suffix, then the result of direct deductions are matched to a dictionary containing the basic words. The use of the dictionary base on algorithms stemming Porter can increase the accuracy of 13.333%. But the speed of the process stemming the use of a dictionary is much slower than without a dictionary.*

**Keywords:** *Stemming, Porter, Dictionary, Indonesian Language*

## PENDAHULUAN

Kebutuhan terhadap informasi sangat meningkat. Banyak informasi dibutuhkan untuk dapat diperoleh dengan cepat, dan relevan sesuai kebutuhan. Lebih penting mencari informasi yang relevan sesuai kebutuhan daripada menyimpan dan melupakan (Pontis, Blandford, Greifeneder, Attalla, & Neal, 2017), tetapi mengambil informasi yang relevan dan berarti sangat sulit (Vivekavardhan, 2015). Berdasarkan hal tersebut dibutuhkan sebuah sistem untuk memenuhi kebutuhan informasi.

*Information retrieval* merupakan metode mencari material (dokumen), meliputi proses mencari dan mengambil informasi berbasis pengetahuan dari kumpulan dokumen (Roshdi & Roohparvar, 2015). *Information retrieval* didefinisikan sebagai tindakan, metode, dan prosedur untuk menemukan kembali data yang tersimpan, kemudian menyediakan informasi mengenai subyek yang dibutuhkan (Karyono & Utomo, 2012). *Information retrieval* digunakan untuk menemukan kembali dokumen teks yang dibutuhkan oleh pengguna, dokumen teks tersebut berupa dokumen yang berbentuk digital (Afuan, 2013). Dokumen digital merupakan dokumen yang dibuat, diteruskan, dikirim, atau disimpan dalam bentuk analog, elektromagnetik, atau sejenisnya yang dapat dilihat, ditampilkan atau didengar melalui media komputer atau sistem elektronik.

*Information retrieval* merupakan bagian penting dalam pencarian informasi untuk mendapatkan informasi yang tepat. Mesin pencarian web adalah contoh paling dikenal yang menerapkan *information retrieval* (Ruban, Sam, Serrao, & Harshitha, 2015). Sebagian besar mesin pencari konvensional bekerja berdasarkan *text queries*, yaitu pengguna memasukkan sejumlah kata sebagai *query* dan mesin pencari mencari dokumen yang berhubungan dengan *query* tersebut (Rahimi & Zahedi, 2014).

Bahasa Indonesia adalah bahasa nasional Negara Indonesia yang merupakan bahasa pemersatu. Saat ini Bahasa Indonesia memiliki peranan penting bukan hanya dalam skala nasional, tetapi juga dalam skala global sebagai bahasa asing, sehingga banyak dipelajari oleh penutur asing (Siroj, 2015). Bukti yang menunjukkan bahwa bahasa Indonesia masih berperan penting adalah

penggunaannya dalam berbagai kegiatan seperti rapat, siaran radio, pidato kenegaraan, dan bahasa pengantar pada berbagai tingkat pendidikan. Bahasa Indonesia terus diupayakan sebagai bahasa pengantar ilmu pengetahuan dan teknologi dengan memperkaya kosakata sesuai perkembangan ilmu pengetahuan dan teknologi (Samsudin, 2015). Bahasa Indonesia sudah dipakai sebagai bahasa perantara bukan saja di Kepulauan Nusantara, tetapi juga hampir di seluruh Asia Tenggara (Bahtiar & Fatimah 2014). Di internet terdapat banyak sekali dokumen berbahasa Indonesia dalam bentuk artikel, jurnal, dan lain-lain.

Bahasa Indonesia berbeda dari bahasa lain karena setiap bahasa memiliki karakteristik sendiri (Apriyanti, Wulandari, Safitri, & Dewi, 2016). Dalam Bahasa Indonesia, afiks mengambil peran penting dan sangat sensitif, karena sedikit perbedaan afiks dapat menyebabkan perbedaan makna yang jauh (Zainuddin, 2016). Afiks terdiri dari *prefixes*, *infixes*, *suffixes*, dan *confixes*. Suatu kata dapat digabungkan dengan *prefixes*, *infixes*, *suffixes* dan *confixes*, hal tersebut menyebabkan sulitnya pencocokan kata yang berhubungan. Beberapa penggabungan menyebabkan perubahan pada akar kata, sebagai contoh kata sapu akan berubah menjadi apu bila digabungkan dengan imbuhan meny-

Dalam *information retrieval*, sebelum pencarian informasi dokumen teks dilakukan prapengolahan terlebih dulu yang dikenal dengan *stemming*, yaitu proses untuk mendapatkan kata dasar dengan menghapus afiks dari kata yang dicari (Setiawan, Kurniawan, Budiharto, Kartowisastro, & Prabowo, 2016). Tugas *stemming* kata merupakan salah satu teknik *preprocessing* dasar dalam NLP (*Natural Language Processing*) (Brychcín & Konopík, 2015). *Stemming* adalah proses heuristik yang memotong ujung kata untuk menemukan kata dasar, dan seringkali mencakup pembuangan afiks (Al Ajeeli, 2016). *Stemming* digunakan untuk mengubah kata yang mengandung afiks menjadi kata dasar (Fadlullah, Kamudi, Nasir, Arifin, & Purwitasari, 2016).

Banyak algoritma yang dikembangkan untuk melakukan proses *stemming* Bahasa Indonesia, di antaranya algoritma Nazief dan Andriani, algoritma Porter, serta algoritma Arifin dan Setiono (Novitasari, 2016). Proses *stemming*

untuk Bahasa Indonesia menggunakan algoritma berbasis aturan memiliki tingkat kesalahan yang tinggi (Utomo, 2013). Algoritma *stemming* porter merupakan yang paling banyak digunakan untuk *stemming* bahasa Inggris (Gupta & Lehal, 2013). Algoritma *stemming* Porter mencari kata dasar dengan menghilangkan afiks/imbuan, kata yang dihasilkan kadang ambigu atau membingungkan karena aturan morfologi bahasa Indonesia yang tidak konsisten.

Agar dalam proses *stemming* algoritma Porter dapat mengurangi keambiguan atau meningkatkan akurasi maka diusulkan penggunaan kamus kata dasar. Setiap kata yang telah melewati proses pemotongan imbuhan dibandingkan dengan kata yang tersimpan dalam kamus kata dasar. Dengan melewati proses validasi menggunakan kamus, diharapkan dapat meningkatkan ketepatan.

## PENELITIAN TERKAIT

Pada penelitian (Tala, 2003) menyatakan bahwa penelitian ini ditujukan untuk mengetahui efek dari *stemming* pada bahasa Indonesia. Penelitian ini dilakukan dengan membandingkan stemmer bahasa Indonesia yang sudah ada, dengan stemmer murni berbasis aturan yang diusulkan. Stemmer berbasis aturan ini dikembangkan berdasarkan struktur kata-kata bahasa Indonesia secara morfologis. Sedangkan metode evaluasi yang digunakan adalah Paice. Penelitian ini juga mengevaluasi pengaruh ukuran kamus dengan kualitas stemmer. Teks yang dijadikan sebagai sample diambil berasal dari 2 media koran online Indonesia, yaitu kompas dan tempo. Dan dihasilkan bahwa proses *stemming* menggunakan algoritma porter menghasilkan 11,8% kata-kata tidak bisa dimengerti dalam teks kompas. Kemudian *stemming* menggunakan algoritma porter yang dibuat proses yang dihasilkan lebih cepat dibandingkan dengan algoritma Nazief dan Adriani.

Pada penelitian Agusta (Agusta, 2009) menyatakan bahwa algoritma-algoritma *stemming* memiliki kelebihan dan kekurangan masing-masing. Efektifitas algoritma *stemming* dapat diukur berdasarkan beberapa parameter, seperti kecepatan proses, keakuratan, dan kesalahan. Tujuan penelitian ini adalah membandingkan kemampuan dan ketepatan algoritma Nazief dan Adriani dengan

algoritma porter. Proses perbandingan algoritma porter dan Nazief dan Adriani dilakukan dengan membuat program sederhana yang memproses dokumen teks Bahasa Indonesia sehingga diketahui stem, waktu proses, presisi dari hasil *stemming* dokumen tersebut. Pengujian algoritma dilakukan pada 30 dokumen teks Bahasa Indonesia dengan ukuran dokumen yang bervariasi. Hasil dari penelitiannya adalah proses *stemming* dokumen teks berbahasa Indonesia menggunakan algoritma Porter membutuhkan waktu yang lebih singkat dibanding dengan *stemming* menggunakan algoritma Nazief dan Adriani. Kemudian proses *stemming* algoritma Porter memiliki persentase keakuratan (presisi) lebih kecil dibanding dengan algoritma Nazief dan Adriani.

Pada penelitian Utomo (Utomo, 2013) stemmer yang dilakukan menggunakan algoritma *stemming* Tala yang berbasis web menyatakan bahwa *stemming* adalah proses untuk mencari kata dasar. Proses *stemming* dilakukan dengan 2 cara yaitu dengan menggunakan kamus dan menggunakan aturan-aturan imbuhan. Algoritma *stemming* berbahasa Indonesia tanpa menggunakan kamus diperkenalkan oleh Vega dan Bressan, dengan menghilangkan imbuhan-imbuhan pada kata-kata berbahasa Indonesia. Selain Vega, Tala juga memperkenalkan Porter stemmer untuk bahasa Indonesia. Tala menggunakan *rule base* untuk menghilangkan imbuhan kata, *stemming* Tala sama sekali tidak menggunakan kamus sebagai acuan mempunyai tingkat kesalahan yang tinggi, sehingga dapat mempengaruhi akurasi hasil akhir.

Untuk mendapatkan tingkat kebenaran hasil yang tinggi biasanya menggunakan kamus seperti yang dikenalkan oleh Nazief dan Adriani, *stemming* bahasa melayu yang diperkenalkan oleh Ahmad, Yuso, dan Sembok. Tetapi teknik penggunaan kamus membutuhkan waktu komputasi yang tinggi Karena adanya proses pengambilan data pada database.

Stemer bahasa Indonesia Tala memiliki struktur pembentukan kata dalam bahasa Indonesia adalah sebagai berikut:

**[awalan-1] + [awalan-2] + dasar + [akhiran]  
+ [kepunyaan] + [sandang]**

Masing-masing bagian tersebut (yang dalam kotak bisa ada atau tidak), digabungkan dengan kata dasar membentuk kata berimbuhan. Stemming Tala merupakan adopsi dari algoritma stemming berbahasa Inggris terkenal yaitu stemming Porter. Algoritma stemming porter merupakan yang paling banyak digunakan untuk stemming bahasa Inggris (Gupta & Lehal, 2013). Stemming ini menggunakan analisa aturan dasar untuk mencari kata dasar sebuah kata. Stemming ini sama sekali tidak menggunakan kamus sebagai acuan, seperti halnya Stemming Ahmad, Vega dan Jelita.

Pada proses stemming dilakukan evaluasi pada 1000 kata terbanyak. Kata-kata tersebut telah bebas dari *stopword*. Hasil evaluasi stemmer Tala diperlihatkan kesalahan karena *overstemming* paling banyak terjadi, dari 1000 kata terdapat 177 kata yang salah karena *overstemming*. Kemudian untuk bahasa asing tidak terjadi perubahan karena akhiran dan awalan tidak dikenali oleh sistem. Kemudian kesalahan juga terjadi pada nama orang/istilah/singkatan. Kebanyakan dengan akhiran ‘i’ akan terpotong oleh sistem huruf ‘i’ terakhirnya, karena tidak ada mekanisme pendeteksian apakah ‘i’ tersebut akhiran atau bagian dari kata. Kesalahan juga terjadi pada kesalahan ketik/masukan kata, susunan imbuhan yang salah atau imbuhan asing secara statistik, dari 1000 kata tersebut terdapat 256 kata yang mengalami kesalahan pencarian kata dasar, sehingga tingkat keberhasilannya adalah 74,4%. Proses Stemming Bahasa Indonesia menggunakan algoritma berbasis aturan mempunyai tingkat kesalahan tinggi, sehingga dapat mempengaruhi akurasi hasil akhir. Tetapi performa stemming berbasis aturan relatif stabil dengan jumlah dokumen yang berkembang.

## METODE PENELITIAN

Stemming adalah cara yang digunakan untuk meningkatkan performa *information retrieval*, kegunaan stemming yaitu untuk mencari kata dasar dari bentuk kata yang berimbuhan (Fitri, 2013). Stemming adalah suatu proses yang terdapat dalam sistem *information retrieval* yang mengubah kata yang terdapat dalam suatu dokumen menjadi kata dasar menggunakan aturan-aturan tertentu (Agusta, 2009).

Proses stemming adalah proses yang digunakan untuk mencari kata dasar dari proses penyaringan, proses penyaringan adalah proses pengambilan kata yang dianggap penting atau memiliki makna (Utomo, 2013). Permasalahan utama dalam proses stemming adalah bagaimana cara mendapatkan kata dasar yang benar dari hasil pemotongan imbuhan, factor yang mempengaruhi efektifitas algoritma stemming adalah (Mandala, Koryanti, Munir, & Harlili, 2004):

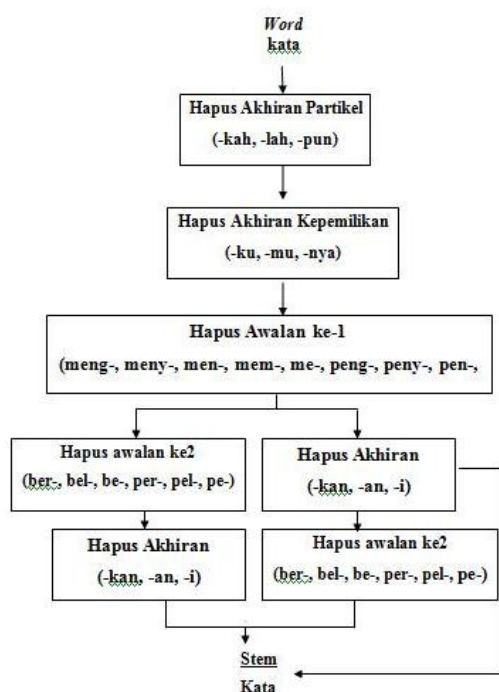
- a. Kesalahan pada proses pemenggalan imbuhan dari kata dasarnya, kesalahan ini dapat berupa:
  - *Overstemming*, yaitu pemenggalan imbuhan yang melebihi dari seharusnya. Contoh: kata *masalah* menjadi *masa*. Kesalahan ini dapat timbul karena bentuk kata dasarnya menyerupai imbuhan.
  - *Understemming*, pemenggalan imbuhan yang terlalu sedikit dari yang seharusnya. Contoh: kata *belajar* menjadi *lajar*. Kesalahan ini dapat timbul akibat kekurangan pada aturan pola imbuhan yang didefinisikan.
  - *Unchange*, kasus khusus dari *understemming*, di mana tidak terjadi pemenggalan imbuhan. Contoh: kata *telapak*, setelah pemenggalan kata dasar hasil yang didapat tetap *telapak*. Kesalahan ini dapat timbul akibat kekurangan aturan pola imbuhan yang didefinisikan.
  - *Spelling exeption*, yaitu huruf pertama kata dasar yang didapat tidak benar akibat dari pemenggalan awalan. Contoh: kata *memukul* menjadi *ukul*, ini diakibatkan beberapa imbuhan yang berubah bentuk ketika ditempelkan pada suatu kata dasar.
- b. Kekurangan pada perumusan aturan penambahan imbuhan pada kata dasar. Hal ini dapat terjadi karena morfologi Bahasa Indonesia yang kompleks, sehingga sangat sulit atau bahkan tidak mungkin untuk merumuskan aturan yang sempurna.
- c. Jumlah total aturan imbuhan yang didapat berhubungan dengan efektifitas proses temu kembali, di mana semakin banyak pola penambahan imbuhan yang

dapat dirumuskan, maka proses temu kembali semakin efektif.

Analisa sistem stemming ini mengacu pada stemming porter yang telah dikembangkan oleh Fadila Z Tala (Tala, 2003) sistem stemming porter yang ada dalam penelitian sebelumnya merupakan sistem stemming yang masih menggunakan aturan dalam algoritma stemming porter belum menggunakan kamus. Dalam analisa sistem ini akan dijelaskan aturan stemming pada stemming porter bahasa Indonesia. Implementasi, aturan *affixs* dan kondisi tindakan akan dijelaskan pada sub berikutnya.

### Implementasi

Pada implementasi stemming porter ini sebelumnya hanya digunakan pada kata atau kalimat bahasa Inggris, karena bahasa Inggris dan bahasa Indonesia datang dari 2 kelas yang berbeda, beberapa modifikasi dilakukan agar performa yang dibuat untuk algoritma stemming porter dapat digunakan untuk bahasa Indonesia. Berikut adalah desain standar dari stemming porter Bahasa Indonesia:



**Gambar 1.** Desain Stemming Porter Untuk Bahasa Indonesia

Pada Gambar 1 dijelaskan bahwa proses stemming porter pertama adalah:

- 1) Menghapus akhiran partikel
- 2) Menghapus akhiran kepemilikan
- 3) Menghapus awalan pertama, jika tidak ada maka lanjut untuk menghapus awalan ke-2, dan jika tidak ditemukan maka menghapus akhiran.
- 4) Menghapus awalan ke-2 dan menghapus akhiran dan kata akhir diartikan sebagai kata dasar.
- 5) Menghapus akhiran, jika tidak ada maka kata tersebut diartikan sebagai kata dasar, namun jika ditemukan maka dilanjutkan dengan menghapus awalan ke-2 dan kata akhir diartikan sebagai kata dasar.

### Aturan Affix (Imbuan)

Berdasarkan aturan morfologi dari Gambar 1, terdapat 5 aturan *affix* yang telah dibuat untuk stemming bahasa Indonesia dan akan diterapkan dalam sistem aplikasi ini. 5(lima) dari aturan tersebut akan didefinisikan dalam beberapa tabel. Yang pertama adalah aturan penghapusan partikel yang berupa akhiran (*suffix*), pada Tabel 1 menjelaskan aturan penghapusan partikel akhiran.

**Tabel 1.** Kelompok Rule Pertama: Akhiran Partikel

Akhiran	Pengganti	Kondisi Tindakan	Kondisi Tambahan	Contoh
Kah	NULL	2	NULL	Bukukah—buku
Lah	NULL	2	NULL	Adalah—ada
pun	NULL	2	NULL	Bukupun--buku

Tabel 1 menjelaskan akhiran kata yang harus dihilangkan ada pada aturan tersebut adalah -kah, -lah, dan -pun.

### Kondisi Tindakan

Dalam rangka pengulangan dengan pengejaan dari kata bahasa Indonesia, tindakan kondisi yang akan dipilih untuk digunakan pada algoritma stemming porter sudah dimodifikasi. Dalam bahasa Indonesia, bagian terkecil dari kata disebut dengan suku kata. Pada Tabel 2 berikut akan menjelaskan aturan penghapusan bagian akhir.

**Tabel 2.** Kelompok Rule Kedua: Akhiran Kepemilikan

Akhiran	Pengganti	Kondisi Tindakan	Kondisi Tambahan	Contoh
Ku	NULL	2	NULL	Bukuku—buku
Mu	NULL	2	NULL	Bukumu—buku
nya	NULL	2	NULL	Bukunya—buku

Tabel 2 menjelaskan aturan akhiran yang harus dihilangkan berupa akhiran -ku, -mu, dan -nya. Kemudian aturan ke 3 dalam stemming porter bahasa Indonesia ini adalah penghilangan awalan pada kata. Pada Tabel 3 akan menjelaskan beberapa awalan yang ada pada aturan.

**Tabel 3.** Kelompok Rule Ketiga: Awalan Pertama

Awalan	Pengganti	Kondisi Tindakan	Kondisi Tambahan	Contoh
Meng	NULL	2	NULL	Mengukur—ukur
Meny	S	2	V ...	Menyapu—sapu
men	NULL	2	NULL	Menduga—duga Menuduh—uduh
mem	P	2	V ...	Memilah—pilah
Mem	NULL	2	NULL	Membaca—baca
Me	NULL	2	NULL	Merusak—rusak
Peng	NULL	2	NULL	Pengukur—ukur
Peny	S	2	V ...	Penyapu—sapu
Pen	NULL	2	NULL	Penduga—duga Penuduh—uduh
Pem	P	2	V ...	Pemilah—pilah
Pem	NULL	2	NULL	Pembaca—baca
Di	NULL	2	NULL	Diukur—ukur
Ter	NULL	2	NULL	Tersapu—sapu
ke	NULL	2	NULL	Kekasih—kasih

Tabel 3 merupakan awalan-awalan yang harus dihilangkan pada kata yang akan dilakukan stemming. Kemudian aturan yang keempat untuk porter stemming adalah penghapusan awalan yang kedua dijelaskan pada Tabel 4.

**Tabel 4.** Kelompok Rule Keempat yang Mencakup Awalan yang Kedua

Awalan	Pengganti	kondisi Tindakan	Kondisi Tambahan	Contoh
Ber	NULL	2	NULL	Berlari—lari
Bel	NULL	2	Ajar	Belajar—ajar
Be	NULL	2	K*er	bekerja—kerja
Per	NULL	2	NULL	Perjelas—jelas
Pel	NULL	2	Ajar	Pelajar—ajar
pe	NULL	2	NULL	Pekerja—kerja

Sebuah suku kata terdiri dari setidaknya satu huruf vokal. Untuk contohnya dapat dilihat pada Tabel 5.

**Tabel 5.** Contoh Suku Kata Dalam Kata Berbahasa Indonesia

Kondisi	Contoh	Suku Kata
0	kh, ng, ny	kh, ng, ny
1	ma, af, nya, nga	ma, af, nya, nga
2	maaf, kami, rumpun, kompleks	ma-af, ka-mi, rum-pun, kom-pleks
3	mengapa, menggunung, tandai	meng-a-pa, meng-gu-nung, tan-da-i

Ukuran kata yang dirancang pada Tabel 5 tidak dapat menangkap semua ukuran sebenarnya dalam kata, ini karena bahasa Indonesia selalu mengenal dua vokal yang diucap sekaligus, itulah urutan dua huruf vokal yang dianggap sebagai sebuah vokal bebas dipisahkan. Sebagai contohnya adalah ai (pantai), au (atau), dan oi. Aturan yang kelima adalah aturan untuk akhiran, dapat dilihat pada Tabel 6.

**Tabel 6.** Aturan Untuk Akhiran

Akhiran	Pengganti	Kondisi Tindakan	Kondisi Tambahan	Contoh
-kan	NULL	2	Prefix bukan anggota (ke, peng)	tarikan—tarik mengambilkan—ambil
-an	NULL	2	prefix bukan anggota (di, meng, ter)	makanan—makan perjanjian—janji
-i	NULL	2	prefix bukan anggota (ber, ke, peng)	tandai—tanda mendapati—dapat

Basis data yang digunakan dalam aplikasi ini adalah basis data kamus, kamus yang dimaksud merupakan kamus bahasa Indonesia yang berisi tabel `tb_katadasar` yang memuat kata dasar sesuai dengan kamus besar bahasa Indonesia. Struktur table `tb_katadasar` ditunjukkan pada Tabel 7.

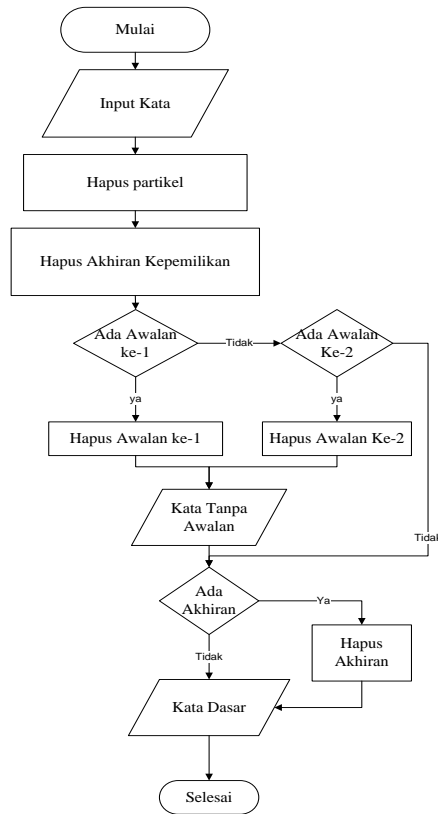
**Tabel 7.** Struktur Table `tb_katadasar`

No	Nama Field	Tipe	Lebar	Kunci
1	<code>id_ktdasar</code>	Integer	10	Primary
2	<code>katadasar</code>	Varchar	20	
3	<code>tipe_katadasar</code>	Varchar	20	

*Flowchart* dibuat untuk menggambarkan alur sebuah sistem pada aplikasi. Gambar 2 menunjukkan *flowchart* dari aplikasi stemming tanpa menggunakan kamus.

Pada Gambar 2 dapat dijelaskan proses stemming dimulai dari memasukan kata kemudian mulai dengan penghapusan partikel, dilanjutkan dengan penghapusan akhiran kepemilikan setelah itu bila kata memiliki awalan pertama makan akan diproses penghapusan awalan pertama jika tidak ada maka akan langsung masuk ke proses penghapusan awalan ke-2 kemudian kata tanpa

awalan akan dihapus akhirnya jika terdapat akhirnya, bila tidak ada maka kata tersebut adalah kata dasar dan proses stemming tanpa kamus selesai.

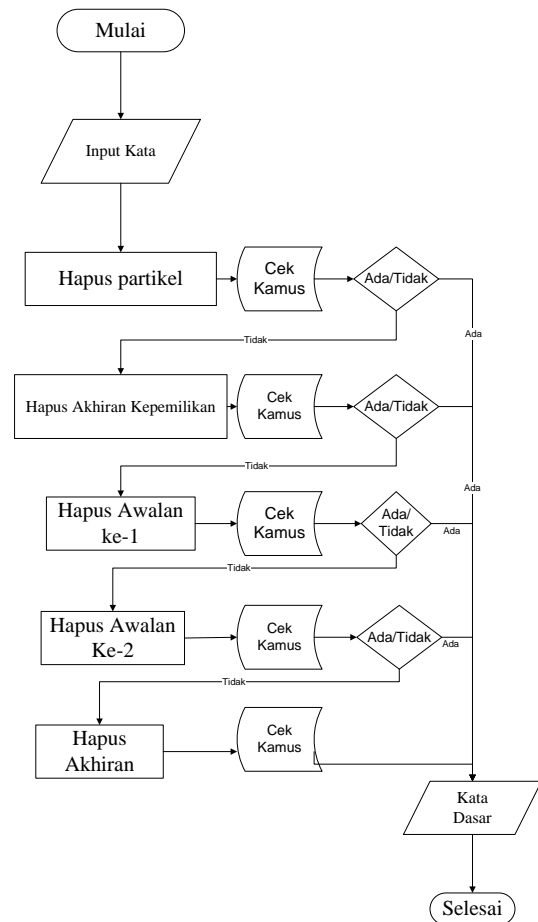


Gambar 2. Flowchart Stemming Tanpa Kamus

Sedangkan flowchart proses stemming dengan kamus ditunjukkan pada Gambar 3. Pada Gambar 3 dapat dijelaskan bahwa setiap proses penghilangan imbuhan selalu melewati proses pengecekan kamus, agar ketepatan untuk menghasilkan kata dasar yang sesuai dengan kamus kata dasar bahasa Indonesia yang telah disediakan.

## HASIL DAN PEMBAHASAN

Pada penelitian ini dilakukan eksperimen dengan menggunakan komputer untuk melakukan proses stemming. Spesifikasi perangkat keras dan sistem operasi yang digunakan pada penelitian ini adalah menggunakan sebuah laptop dengan prosesor Intel® core™ i3-3217U CPU @ 1.80GHz (4 CPUs), ~1.7GHz, memori (RAM) 2 GB, dan menggunakan sistem operasi Windows 8 Pro 64-bit (6.2, build 9200). Sedangkan perangkat lunak yang digunakan untuk menyimpan database kamus adalah MySQL.



Gambar 3. Flowchart Stemming Porter dengan Kamus

Tabel 8. Hasil Pengujian 30 Kata yang di pilih Secara Acak

No	Kata Berimbuhan	Hasil Stem dengan Kamus	Hasil stem tanpa kamus	Waktu Stemming	
				Stem Tanpa Kamus /detik	Stem Dengan Kamus /detik
1	Terangsang	Angsang	Angsang	0,00024	0,51293
2	Mengenal	Kenal	Kenal	0,00026	0,50676
3	Mengail	Ail	Ail	0,00026	0,47243
4	Mengunci	Kunci	Kunci	0,00047	0,48950
5	Mengembalikan	Kembali	Kembali	0,00026	0,51464
6	Mengelana	Kelana	Kelana	0,00027	0,47553
7	Masukan	Masu	Masu	0,00030	0,53291
8	Memeriksa	Periksa	Iksa	0,00023	0,53811
9	Mengendarai	Kendara	Kendar	0,00026	0,47217
10	Memenjarakan	Njara	Njara	0,00035	0,56486
11	Segalanya	Segala	gala	0,00036	0,46224
12	Mempengaruhi	Ngaruh	Ngaruh	0,00030	0,51812
13	Proses	Proses	Proses	0,00029	0,46083
14	Analisa	Analisa	Analisa	0,00024	0,46314
15	Diketahui	Ketahu	Ketahu	0,00033	0,42627
16	Seberapa	Berapa	Berapa	0,00040	0,57328
17	Keterhubungan	Terhubung	Terhubung	0,00023	0,47685
18	Tingkat	Tingkat	Tingkat	0,00037	0,48108
19	Dokumen	Dokumen	Dokumen	0,00036	0,48968
20	Antar	Antar	Antar	0,00022	0,51307
21	Jauh	Jauh	Jauh	0,00030	0,58590
22	Ada	Ada	Ada	0,00032	0,53438
23	Menanak	Anak	anak	0,00047	0,60775
24	Berakal	Akal	akal	0,00023	0,51729
25	Nasihat	Nasihat	nasihat	0,00022	0,49057
26	Membacakan	Baca	baca	0,00021	0,49193
27	Terkenal	Kenal	kenal	0,00031	0,49009
28	Berlari	Lari	Lar	0,00032	0,48746
29	Berguman	Gum	Gum	0,00032	0,49607
30	Pemrograman	Rogram	Rogram	0,00021	0,48454

Dari Tabel 8 percobaan di atas menyatakan bahwa ada beberapa kata imbuhan yang distemming dengan kamus menghasilkan kata dasar yang rata-rata masih memiliki makna, namun pada hasil kata dasar yang distemming tanpa kamus menghasilkan beberapa kata dasar yang ambigu atau membingungkan seperti “gum”, ”rogram”, “gala”, “njara”, “iksa”, ”masu” dan “ail”.

Sementara untuk perbandingan waktu menunjukkan bahwa waktu yang diperlukan dalam proses stemming dengan kamus membutuhkan waktu beberapa detik lebih lama dari pada stemming tanpa menggunakan kamus. Dapat dilihat pada kolom waktu stemming, pada stemming tanpa kamus menunjukkan bahwa dari 30 kata berimbuhan yang diproses stemming rata-rata menunjukkan waktu 0,00030 detik, sementara untuk stemming dengan menggunakan kamus membutuhkan lebih dari 0,001 detik pada Tabel 8 tersebut menunjukkan bahwa rata-rata waktu stemming dengan kamus adalah 0,50767 detik. Percobaan waktu ini dicoba pada proses stemming setiap 1 kata berimbuhan.

Untuk waktu yang dihasilkan pada saat proses stemming dengan memasukkan 30 kata secara langsung dan memprosesnya secara bersamaan menghasilkan waktu 0,00813 detik untuk proses stemming tanpa kamus dan 15,50625 detik untuk proses stemming dengan menggunakan kamus. Dari waktu tersebut ada selisih waktu 15,49812 detik selisih yang sangat signifikan.

Untuk Persentase hasil kata dasar yang diproses stemming dengan kamus dapat dilihat pada perhitungan berikut:

$$((30-5))/30 \times 100\% = 83.333\%$$

Sementara hasil kata dasar yang diproses stemming tanpa kamus dapat dilihat pada perhitungan berikut:

$$((30-9))/30 \times 100\% = 70\%$$

Hasil presentase di atas menunjukkan bahwa proses stemming dengan ditambahkan kamus dasar memiliki persentase keakuratan kata yang benar lebih besar 13.333% dari pada persentase keakuratan stemming yang hanya menggunakan aturan yang sudah ada.

## KESIMPULAN

Berdasarkan penulisan dan penelitian yang telah diuraikan, maka dapat dibuat kesimpulan yaitu, penambahan kamus pada algoritma stemming dapat meningkatkan ketepatan pada kata dasar yang dihasilkan, kata berimbuhan yang berhasil distemming dengan tepat terdapat 21 kata dari 30 kata yang diujikan dengan stemming tanpa kamus dasar dan terdapat 25 kata yang berhasil dilakukan stemming dengan tepat dari 30 kata yang diujikan dengan stemming kamus. Persentase hasil kata dasar yang diproses stemming dengan kamus adalah 83,333% dan hasil persentase kata dasar yang diproses stemming tanpa kamus adalah 70%, sehingga menunjukkan proses stemming dengan menggunakan kamus mempunyai persentase ketepatan lebih besar 13,333% dari pada stemming tanpa kamus.

Pada proses stemming dengan menggunakan kamus mempengaruhi kecepatan pada proses stemming, ini dilihat dari rata-rata waktu yang dihasilkan dari proses stemming dengan kamus mempunyai rata-rata 0.50767 detik sedangkan untuk proses stemming tanpa kamus mempunyai rata-rata 0,00030 detik, sehingga pengaruh penambahan kamus terlihat sangat signifikan pada kecepatan proses stemming.

## DAFTAR PUSTAKA

- Afuan, L. (2013). Stemming Dokumen Teks Bahasa Indonesia Menggunakan Algoritma Porter. *Jurnal Telematika vol.6*, 34-40.
- Agusta, L. (2009). Perbandingan Algoritma Stemming Porter dengan Algoritma Nazief&Adriani untuk Stemming Dokumen Teks Bahasa Indonesia. *Konferensi Nasional Sistem dan Informatika*, 196-200.
- Al Ajeeli, A. T. (2016). An Intelligent Framework for Natural Language Stems Processing. *Global Journal of Computer Science and Technology*, 16(1), 22-38.
- Apriyanti, T., Wulandari, H., Safitri, M., & Dewi, N. (2016). Translating Theory of English into Indonesian and Vice-Versa. *Indonesian Journal of English Language Studies*, 2(1), 38-59.
- Bahtiar, A., & Fatimah. (2014). *Bahasa Indonesia untuk Perguruan Tinggi*. Bogor: IN-MEDIA-Bogor.
- Brychcin, T., & Konopik, M. (2015). HPS: High Precision Stemmer. *Information Processing and Management*, 68-91.



- Fadlullah, A., Kamudi, D. D., Nasir, M., Arifin, A. Z., & Purwitasari, D. (2016). Web News Documents Clustering in Indonesian Language using Singular Value Decomposition-Principal Component Analysis and Ant Algorithms. *Jurnal Ilmu Komputer dan Informasi (Journal of Computer Science and Information)*, 9(1), 17-25. doi:10.21609/jiki.v9i1.362
- Fitri, M. (2013). Perancangan Sistem Temu Balik Informasi Dengan Metode Pembobotan Kombinasi TF-IDF untuk Pencarian Dokumen Berbahasa Indonesia. *Jurnal Sistem dan Teknologi Informasi (JustIN)*, 1(1), 1-6. Retrieved from <http://jurnal.untan.ac.id/index.php/justin/article/view/1319>
- Gupta, V., & Lehal, G. S. (2013). A Survey of Common Stemming Techniques and Existing Stemmers for Indian Languages. *Journal of Emerging Technologies in Web Intelligence*, 5(2), 157-161. doi:10.4304/jetwi.5.2.157-161
- Karyono, G., & Utomo, F. S. (2012). Temu Balik Informasi pada Dokumen Teks Berbahasa Indonesia dengan Metode Vector Space Retrieval AI Model. *Journal Seminal Nasional Teknologi Informasi & Komunikasi Terapan*, 282-288.
- Mandala, R., Koryanti, E., Munir, R., & Harlili. (2004). Sistem Stemming Otomatis untuk Kata dalam Bahasa Indonesia. *Seminar Nasional Aplikasi Teknologi Informasi*, 29-35.
- Novitasari, D. (2016). Perbandingan Algoritma Stemming Porter dengan Arifin Setiono untuk Menentukan Tingkat Ketepatan Kata Dasar. *Jurnal String*, 1(2), 120-129.
- Pontis, S., Blandford, A., Greifeneder, E., Attalla, H., & Neal, D. (2017). Keeping Up to Date: An Academic Researcher's Information Journey. *Journal of the Association for Information Science and Technology*, 68(1), 22-35. doi:10.1002/asi.23623
- Rahimi, M., & Zahedi, M. (2014). Query Expansion based on Relevance Feedback and Latent Semantic Analysis. *Journal of AI and Data Mining*, 2(1), 79-84.
- Roshdi, A., & Roohparvar, A. (2015). Review: Information Retrieval Techniques and Applications. *International Journal of Computer Networks and Communications Security*, 3(9), 373-377.
- Ruban, S., Sam, S. B., Serrao, L. V., & Harshitha. (2015). A Study and Analysis of Information Retrieval Models. *International Journal of Innovative Research in Computer and Communication Engineering*, 3(7), 230-236.
- Samsudin, D. (2015). Peran Media Dalam Pemasyarakatan Istilah Bahasa Indonesia. *Metalingua*, 13(2), 151-159.
- Setiawan, R., Kurniawan, A., Budiharto, W., Kartowisastro, I. H., & Prabowo, H. (2016). Flexible affix classification for stemming Indonesian Language. 2016 13th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON) (pp. 1-6). Chiang Mai, Thailand: IEEE. doi:10.1109/ECTICon.2016.7561257
- Siroj, M. B. (2015). Pengembangan Model Integratif Bahan Ajar Bahasa Indonesia Ranah Sosial Budaya berbasis ICT bagi Penutur Asing Tingkat Menengah. *Jurnal Pendidikan Bahasa dan Sastra Indonesia*, 4(1), 74-84.
- Tala, F. Z. (2003). A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia. 1-33.
- Utomo, M. S. (2013, Januari). Implementasi Stemmer Tala pada Aplikasi Berbasis Web. *Jurnal Teknologi Informasi DINAMIK Volume 18 no 1*, 41-45.
- Vivekavardhan, J. (2015). Search Engines for User Centric Information Retrieval and Scholarly Communication. *International Journal of Advanced Library and Information Science*, 3, 201-211.
- Zainuddin. (2016). A Study on Derivational Affixes of Indonesian Noun-Formation in Newspaper Editorial: A Semantic Perspective. *International Journal of English Linguistics*, 6(3), 148-155. doi:10.5539/ijel.v6n3p148

