



## Ekstraksi Kata Dasar Secara Berjenjang (*Incremental Stemming*) Berbasis Aturan Morfologi untuk Teks Berbahasa Indonesia

Wahyu Hidayat<sup>1</sup>

<sup>1</sup>D3 Manajemen Informatika, Fakultas Ilmu Terapan, Universitas Telkom

<sup>1</sup>Jl. Telekomunikasi Terusan Buah Batu, Bandung 40257, Indonesia

Email korespondensi: [wahyuhidayat@tass.telkomuniversity.ac.id](mailto:wahyuhidayat@tass.telkomuniversity.ac.id)

Dikirim 18 April 2017, Direvisi 28 April 2017, Diterima 3 Mei 2017

Abstrak – Ekstraksi kata dasar atau *stemming* pada Bahasa Indonesia adalah proses yang kompleks di mana beberapa partikel awal dan beberapa partikel akhiran dari 13 awalan, 3 sisipan, dan 19 akhiran yang dikenal dapat digunakan secara sekaligus pada sebuah kata. Selain itu, proses *stemming* tidak terlalu menghasilkan 1 kata dasar (non-deterministik) karena terdapat beberapa kata dalam Bahasa Indonesia yang memiliki 2 kemungkinan, yaitu sebagai kata dasar maupun kata berimbuhan, misalnya pada kata “beruang”. Penelitian yang telah ada sebelumnya menggunakan kombinasi awalan dan akhiran yang tidak mungkin dan menerapkan *heuristik* untuk memilih kata dasar. Dalam penelitian ini diusulkan sebuah metode *stemming* secara berjenjang di mana berdasarkan urutan tertentu, secara bergantian partikel akhiran dan awalan dilepaskan dari sebuah kata sehingga dihasilkan sebuah kata dasar. Jika ditemukan beberapa kandidat kata dasar maka salah satu kata dasar akan dipilih. Metode ini diuji pada 6464 dokumen Al-Quran Terjemahan Indonesia dengan menggunakan kamus berukuran 5000 kata yang dipilih secara acak dari Kamus Besar Bahasa Indonesia. Dari 3432 kata unik yang diproses, diperoleh 94,7% kata dasar yang dapat diekstrak secara langsung dan hanya 5,3% yang perlu diproses lebih lanjut karena kandidat kata dasar yang ditemukan lebih dari satu. Dibandingkan dengan melakukan pemilihan kata dasar secara manual, metode ini dapat memilih kata dasar yang tepat hingga 79,12%.

Kata kunci – ekstraksi kata dasar, morfologi, non-deterministik, akurasi

Abstract—Stemming in Indonesian Language is a complex process where few particles among 13 known prefixes, 3 known infixes, and 19 known suffixes can be used simultaneously in a single word. Moreover, stemming process does not guarantee a single word result (hence, non-deterministic), because there are few words in Indonesian Language that has 2 possibilities, either it is a morpheme itself or it is an affixed word, e.g. the word “beruang”. Previous researches have tried to deal with this issue by listing impossible prefix and suffix combination while at the same time implementing heuristic to identify final stem. In this paper, we present an incremental stemming method, where based on particular order, suffixes and prefixes are alternately stripped until a single morpheme (stem) is found. In case there are several stem candidates were found, then one stem will be chosen. This method is tested against 6464 text files from Holy Quran Indonesian Translation using 5000-word dictionary randomly chosen from Kamus Besar Bahasa Indonesia. From 3432 unique words, 94.7% single stem can be identified immediately and only 5.3% that need to be processed further since there are more than one candidate stem were found. Compared to manually stem selection, this method is able to accurately choose 79.12% correct stem.

Keywords-stemming; morphology; non-deterministic; accuracy

### I. PENDAHULUAN

Ekstraksi kata dasar atau *stemming* adalah salah satu tahapan yang memiliki peran penting dalam bidang *information retrieval* [1]. *Stemming* memiliki

tiga tujuan utama, yaitu untuk *clustering* berdasarkan topik, meningkatkan kualitas hasil pencarian pada sebuah *information retrieval system* dan untuk memperkecil ukuran indeks pada sebuah *information*

*retrieval system* [2]. Dalam surveinya [2] membagi algoritma *stemming* menjadi pendekatan klasikal dan pendekatan modern. Pendekatan klasikal lebih fokus kepada bentuk kata dan aturan morfologi dan sangat dipengaruhi oleh bahasa di mana *stemming* tersebut akan diterapkan. Adapun pendekatan modern berupaya mendapatkan kata dasar dengan mengidentifikasi konteks dan domain dari dokumen atau kalimat tempat kata tersebut ditemukan. Beberapa penelitian yang menggunakan pendekatan modern adalah [3] dan [4].

Metode *stemming* perlu terus dikembangkan dan disempurnakan mengingat *stemming* adalah salah satu proses awal di tahap *indexing* pada sebuah *information retrieval engine*. Peningkatan kualitas proses *stemming* akan pada akhirnya turut meningkatkan kualitas *information retrieval engine* secara keseluruhan.

Beberapa penelitian sebelumnya telah mencoba mengusulkan berbagai metode *stemming* untuk teks berbahasa Bahasa Indonesia. Salah satu penelitian terbaru yang melakukan pendekatan klasikal dengan memanfaatkan aturan morfologi adalah metode *Flexible Affix Classification* [5]. Namun demikian, penelitian tentang *stemming* pada Bahasa Indonesia yang menjadi banyak rujukan bagi penelitian lain adalah [6]. Penelitian ini menjadi dasar algoritma *stemming* pada Bahasa Indonesia yang cukup populer, yaitu *Confix Stripping* (*ECS Stemmer*) [8]. Penelitian lain berupaya menemukan aturan morfologi secara otomatis melalui proses *learning* [9], memodifikasi *stemming* untuk proses lematisasi [10] atau menemukan manfaat lain dari proses *stemming*, misalnya untuk kompresi teks [11] dan untuk mendeteksi kesalahan pengetikan (*typo*) [12].

*Stemming* pada Bahasa Indonesia memiliki tantangan tersendiri di mana terdapat 13 awalan, 3 sisipan, dan 19 akhiran standar [13]. Hal ini menjadi semakin kompleks karena beberapa partikel awalan dan beberapa partikel akhiran dapat digunakan sekaligus dalam sebuah kata. *Stemming* dalam Bahasa Indonesia tidak selalu menghasilkan 1 kata dasar (*stem*) karena terdapat beberapa kata dalam Bahasa Indonesia yang memiliki 2 kemungkinan, yaitu sebagai kata dasar maupun kata berimbuhan. Salah satu contohnya adalah kata “beruang”, di mana kata ini dapat berupa kata dasar “uang” yang diberi awalan ber- maupun kata dasar “beruang” itu sendiri. Contoh kata lain dalam Bahasa Indonesia yang dapat menghasilkan lebih dari 1 stem adalah kata “beribu” di mana kata ini memiliki 2 kemungkinan, yaitu awalan ber- yang bertemu dengan kata dasar “ibu” (“ber-“ + “ibu”) atau awalan be- yang bertemu dengan kata dasar “ribu” (“be-“ + “ribu”).

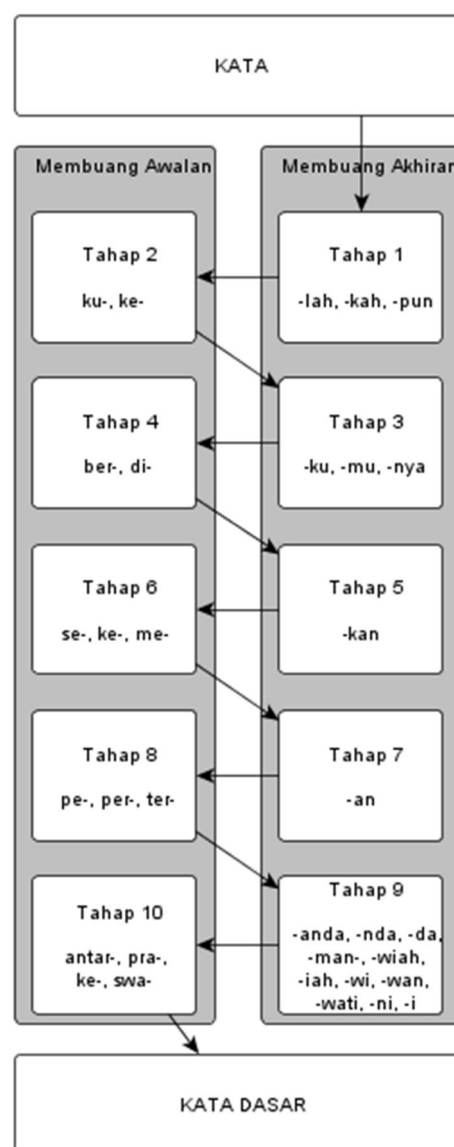
Penelitian tentang *stemming* pada Bahasa Indonesia yang mencoba memberikan solusi untuk masalah ini salah satunya adalah [14] yang memanfaatkan daftar kombinasi awalan dan akhiran yang tidak mungkin sekaligus menerapkan *heuristic*

untuk pemilihan kata dasar (*stem*) dari beberapa kemungkinan kata dasar yang dihasilkan. Namun demikian, [14] belum memproses beberapa akhiran standar dalam Bahasa Indonesia, diantaranya yaitu akhiran *-wi*, *-wan*, *-wati*, dan *-wiah*.

## II. METODE PENELITIAN

### A. Perancangan Stemmer

Metode *stemming* berjenjang yang diusulkan bekerja dengan cara membuang akhiran dan awalan secara bertahap. Akhiran dibuang terlebih dahulu kemudian disusul oleh awalan. Proses ini terbagi menjadi 10 tahap di mana tahap ganjil adalah tahap pembuangan akhiran dan tahap genap adalah tahap pembuangan awalan. Detil partikel-partikel awalan dan akhiran yang dibuang pada setiap tahapnya dapat dilihat pada Gambar 1 berikut ini.



Gambar 1. Sepuluh Tahap Pembuangan Akhiran Dan Awalan Secara Bertahap Dan Bergantian

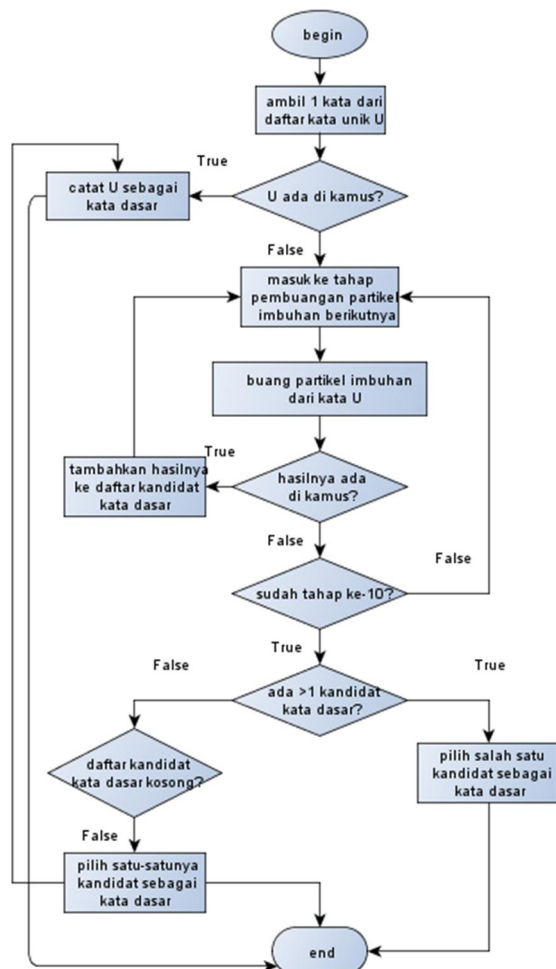
Dalam setiap tahap pembuangan imbuhan, diterapkan aturan morfologi dan dilakukan

pencocokan kata yang dihasilkan dengan daftar kata pada kamus. Jika ditemukan kecocokan maka kata yang dihasilkan akan dimasukkan ke dalam daftar kandidat kata dasar.

Tabel 1. Aturan Morfologi

Partikel Asal	Huruf yang Mengikuti	Berubah Menjadi Partikel
me-	d, c, j, t	men-
	s	meny-
	b, p	mem-
ber-	a,i,u,e,o, g, h, k, kh	meng-
	r atau suku pertama diakhiri huruf r	be
	kata "ajar"	bel-
pe-	d, c, j, t	pen-
	s	peny-
	b, p	pem-
ter-	a, i, u, e, o, g, h, k, kh	peng-
	r	te-

Tabel 1 di atas menunjukkan aturan morfologi yang diterapkan pada proses pembuangan partikel imbuhan sesuai dengan aturan yang dijelaskan pada [13]. Adapun Gambar 2 berikut ini menunjukkan *flowchart* yang menggambarkan logika proses pembuangan partikel imbuhan.



Gambar 2. Logika Proses Pembuangan Partikel Imbuhan

Jika kandidat kata dasar yang ditemukan lebih dari satu maka setelah tahap ke-10 dilakukan 1 tahap tambahan, yaitu pemilihan kata dasar dari daftar kandidat kata dasar yang ditemukan. Ada 3 metode memilih kata dasar dari daftar kandidat kata dasar, yaitu:

- a) Memilih kandidat kata dasar yang paling pertama ditemukan sebagai kata dasar
- b) Memilih kandidat kata dasar yang terpanjang (memiliki jumlah karakter paling banyak) sebagai kata dasar
- c) Memilih kandidat kata dasar yang terpendek (memiliki jumlah karakter paling sedikit) sebagai kata dasar

### B. Pengujian Stemmer

Untuk menguji performa metode *stemming* yang diusulkan, digunakan 6464 *file* teks berbahasa Indonesia dari Alquran Terjemahan Bahasa Indonesia. Kamus yang digunakan untuk proses *stemming* adalah kamus *sampling* berukuran 5000 kata yang berisi 2 sampai 3 kata yang dipilih secara acak dari setiap halaman pada Kamus Besar Bahasa Indonesia.

Adapun langkah-langkah pengujiannya adalah sebagai berikut.

1. Setiap *file* teks mengalami proses *parsing* untuk mengubahnya menjadi potongan-potongan kata dalam format *lowercase* dengan ukuran kata minimal 3 karakter.
2. Proses *stopping* dilakukan terhadap potongan-potongan kata hasil dari tahap *parsing*. Setiap kata dibandingkan dengan daftar kata yang tidak penting (*stopwords*), jika ditemukan kecocokan maka kata tersebut tidak diteruskan ke proses *stemming*. Hasil dari proses *stopping* adalah kata-kata yang dianggap penting.
3. Menghilangkan duplikasi kata-kata yang dianggap penting sehingga dihasilkan daftar kata yang unik.
4. Setiap kata unik menjadi input bagi proses *stemming*, baik kata yang menjadi input maupun kata yang menjadi output proses *stemming* dicatat.
5. Jika terdapat kata yang memiliki lebih dari satu kandidat kata dasar, maka baik kata yang menjadi input, daftar kandidat kata dasar maupun kata dasar yang dipilih oleh sistem akan dicatat.
6. Untuk kata-kata yang memiliki kandidat kata dasar lebih dari satu, output pilihan kata dasar dari sistem dibandingkan dengan hasil pemilihan kata dasar secara manual untuk mengukur akurasi metode yang diusulkan.

### III. HASIL PENELITIAN

Setelah dilakukan pengujian sesuai dengan langkah-langkah yang telah dijelaskan pada sub bab sebelumnya, hasilnya diperoleh sebagai berikut.

Tabel 2. Hasil Tiap Tahap Pengujian

Proses	Input	Output
Parsing	6464 file teks	227.160 kata
Stopping	227.160 kata, daftar kata tidak penting ( <i>stopwords</i> )	126.094 kata penting
Menghilangkan Duplikasi	126.094 kata penting	3432 kata unik
Stemming	3432 kata unik, kamus kata dasar	3250 kata dasar (langsung ditemukan) 182 kata dasar (dipilih dari >1 kandidat kata dasar)

Adapun terhadap 182 kata yang memiliki lebih dari satu kandidat kata dasar, dilakukan pemilihan kata dasar dengan 3 metode yang berbeda dan diperoleh hasil pengujian seperti pada Tabel 3 berikut.

Tabel 3. Hasil Pengujian Metode Pemilihan Kata Dasar

Metode Pemilihan Kata Dasar	Jumlah Kandidat Kata Dasar	Jumlah kata dasar yang dipilih dengan tepat
Pilih kata dasar yang pertama ditemukan	2	137
	3	7
Pilih kata dasar terpanjang	2	123
	3	8
Pilih kata dasar terpendek	2	49
	3	3

Berikut ini adalah tabel yang menunjukkan beberapa contoh kata asal, daftar kandidat kata dasar, kata dasar yang dipilih oleh sistem dan kata dasar dan perbandingannya dengan kata dasar yang dipilih secara manual.

Tabel 4. Contoh Kasus Ditemukan &gt;1 Kandidat Kata Dasar

Input kata	Kandidat Kata Dasar	Metode Pemilihan Kata Dasar	Kata Dasar yang Dipilih Sistem	Kata Dasar Dipilih Manual
berikanlah	beri, ikan	pertama	beri	beri
		terpanjang	beri	
		terpendek	beri	
kebalikan	balik, kebal	pertama	balik	balik
		terpanjang	balik	
		terpendek	balik	
memulai	mula, pula	pertama	mula	mula
		terpanjang	mula	
		terpendek	mula	
memohon	mohon, pohon	pertama	mohon	mohon
		terpanjang	mohon	
		terpendek	mohon	
berimankah	beri, iman, rim	pertama	iman	iman
		terpanjang	iman	
		terpendek	rim	

Adapun secara keseluruhan, hasil uji tingkat akurasi metode memilih kata dasar dari beberapa kandidat kata dasar terlihat pada tabel berikut ini.

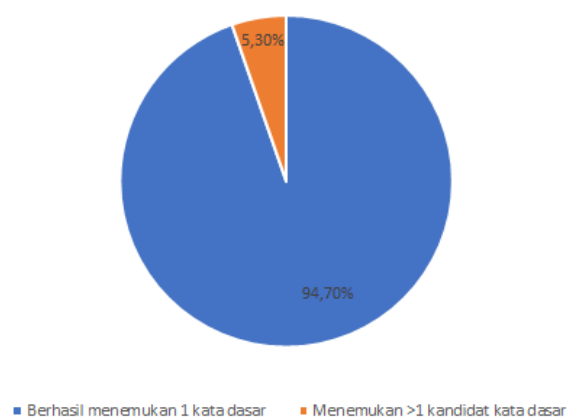
Tabel 5. Perbandingan Tingkat Akurasi Metode Pemilihan Kata Dasar

Metode Pemilihan Kata Dasar	Jumlah Kandidat Kata Dasar	Persentase Jumlah Kata Dasar yang Dipilih dengan Tepat
Pilih kata dasar yang pertama ditemukan	2 kandidat (166 kasus)	82,53%
	3 kandidat (16 kasus)	43,75%
	<b>Rata-rata</b>	<b>79,12%</b>
Pilih kata dasar yang terpanjang	2 kandidat (166 kasus)	74,10%
	3 kandidat (16 kasus)	50%
	<b>Rata-rata</b>	<b>71,98%</b>
Pilih kata dasar yang terpendek	2 kandidat (166 kasus)	29,52%
	3 kandidat (16 kasus)	18,79%
	<b>Rata-rata</b>	<b>28,57%</b>

### IV. PEMBAHASAN

Dari data-data yang diperoleh melalui proses pengujian terlihat bahwa metode *stemming* secara berjenjang (*incremental stemming*) sebagian besar berhasil menemukan kata dasar tanpa perlu melalui tahap pemilihan kata dasar. Hanya sedikit kasus saja yang membutuhkan proses pemilihan kata dasar di mana proses *stemming* menemukan lebih dari satu kandidat kata dasar. Adapun perbandingan persentase kasusnya dapat dilihat pada Gambar 3 berikut ini.

Hasil Proses Stemming Secara Berjenjang

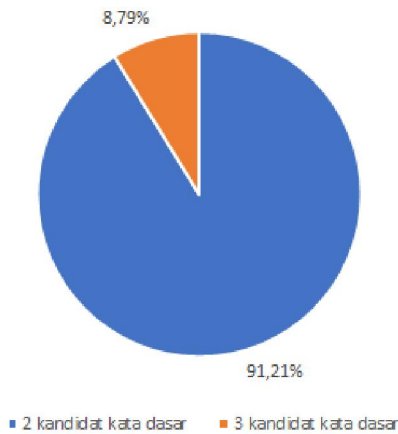


Gambar 3. Perbandingan Persentase Kasus Stemming Langsung Menghasilkan Satu Kata Dasar yang Memerlukan Proses Lebih Lanjut

Dari 5,3% kasus yang membutuhkan proses pemilihan kata dasar karena terdapat lebih dari satu kandidat kata dasar, terdapat jauh lebih banyak kasus di mana jumlah kandidat kata dasarnya adalah dua buah, yaitu sebanyak 91,21% kasus. Selebihnya pada 8,79% kasus, proses *stemming* menghasilkan 3 kandidat kata dasar. Berdasarkan eksperimen yang

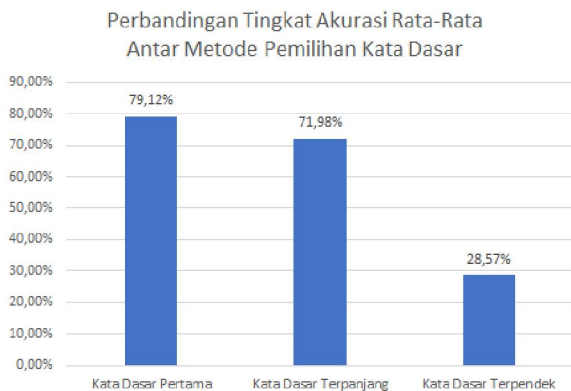
dilakukan belum ditemukan adanya kasus di mana jumlah kandidat kata dasar lebih dari 3 buah. Hal ini dinilai positif karena semakin sedikit jumlah pilihan maka peluang membuat pilihan yang benar akan semakin besar. Adapun perbandingan persentase antara kasus di mana ditemukan dua kandidat kata dasar dan ditemukan tiga kandidat kata dasar dapat dilihat pada Gambar 4 berikut ini.

Perbandingan Jumlah Kandidat Kata Dasar



Gambar 4. Perbandingan Persentase Jumlah Kasus Ditemukan Dua Kandidat Kata Dasar dan Ditemukan Tiga Kandidat Kata Dasar

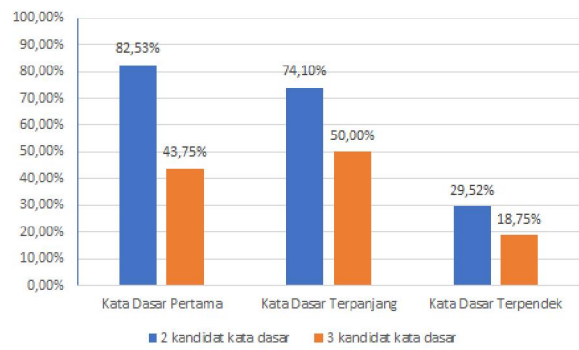
Untuk melakukan pemilihan kata dasar, dari tiga metode yang diujikan, ternyata secara rata-rata, metode pemilihan kata dasar dengan memilih kata dasar yang pertama ditemukan mengungguli dua metode yang lain di angka 79,12%. Adapun perbandingannya dapat dilihat pada Gambar 5 berikut ini.



Gambar 5. Sepuluh Tahap Pembuangan Akhiran dan Awalan Secara Bertahap dan Bergantian

Jika dilihat lebih detail, semakin sedikit jumlah pilihan kandidat kata dasar, maka semakin besar pula tingkat akurasi yang ditunjukkan oleh ketiga metode pemilihan kata dasar. Secara umum, tiap metode pemilihan kata dasar menunjukkan performa yang cukup baik saat dihadapkan pada dua pilihan kandidat kata dasar. Namun, begitu kandidat pilihan kata dasar bertambah menjadi tiga buah maka semua metode menunjukkan penurunan akurasi yang signifikan. Adapun detail perbandingannya terlihat pada Gambar 6 berikut ini.

Perbandingan Tingkat Akurasi Antar Metode Pemilihan Kata Dasar



Gambar 6. Sepuluh Tahap Pembuangan Akhiran dan Awalan Secara Bertahap dan Bergantian

Gambar 6 di atas menunjukkan bahwa dengan menurunnya tingkat akurasi seiring dengan bertambahnya jumlah kandidat kata dasar yang ditemukan maka semakin menegaskan pentingnya meminimalisir jumlah kandidat kata dasar.

Adapun perbedaan antara metode *incremental stemming* dengan penelitian sebelumnya dapat dilihat pada Tabel 6 berikut ini.

Tabel 6. Perbandingan Dengan Metode Lain

Metode Stemming	Data Uji	Akurasi	Kemampuan menangani kemungkinan ditemukan kandidat kata dasar > 1
<i>Incremental Stemming</i>	6464 file teks dari Alquran Terjemahan Indonesia	94,7%	Ya, akurasi 79.12%
<i>Flexible Affix Classification</i> [5]	1074 file teks dari forum diskusi Universitas Bina Nusantara	93.7%	Tidak
<i>Confix Stripping</i> [8]	9898 artikel berita dari Kompas (hanya mengambil satu kata tiap artikel, yaitu kata ke-5)	95.05%	Tidak
<i>Enhanced Confix Stripping</i> [9]	253 dokumen berita berbahasa Indonesia	Akurasi tidak diukur, Peningkatan efisiensi ukuran index sebesar 32.66%	Tidak

Dari Tabel 5 terlihat bahwa beberapa metode *stemming* untuk Bahasa Indonesia yang sudah pernah diusulkan belum dapat dibandingkan secara *apple to apple* karena dibangun untuk tujuan yang berbeda dan diuji dengan data yang berbeda-beda. Oleh karena itu, untuk mengukur performa dalam rangka memilih metode *stemming* yang paling optimal, dibutuhkan data dan metode uji yang seragam.

## V. PENUTUP

### A. Kesimpulan

Berdasarkan hasil pengujian maka dapat disimpulkan bahwa metode *stemming* berjenjang (*incremental stemming*) yang dipaparkan secara umum telah berhasil mengidentifikasi sebagian besar kata dasar (94,7%) tanpa perlu melakukan pemilihan kata dasar dan hanya 5,3% yang memerlukan proses lebih lanjut karena ditemukan lebih dari 1 kata dasar.

Dari 5,3% kata yang perlu diproses lebih lanjut, metode ini dapat menghasilkan 2-3 pilihan kata dasar, di mana 91,21% diantaranya memiliki 2 kemungkinan kata dasar dan 8,79% sisanya memiliki 3 kemungkinan kata dasar.

Adapun metode pemilihan kata dasar yang memberikan akurasi tertinggi adalah dengan memilih kata dasar yang paling pertama ditemukan, yaitu dengan tingkat akurasi rata-rata 79,12% disusul kemudian dengan memilih kata dasar yang paling panjang (71,98%) dan yang paling rendah akurasinya adalah dengan cara memilih kata dasar yang paling pendek (28,57%).

Semua metode pemilihan kata dasar yang telah diuji cenderung menunjukkan performa yang semakin buruk seiring dengan bertambahnya jumlah kandidat kata dasar yang harus dipilih. Oleh karena itu, sangat penting untuk mengembangkan metode *stemming* yang mampu meminimalisir jumlah kandidat kata dasar. Salah satu cara yang mungkin dapat ditempuh adalah dengan memperbanyak koleksi kata dasar dalam kamus yang digunakan.

### B. Saran

Penelitian lebih lanjut dibutuhkan untuk mengukur pengaruh ukuran kamus terhadap jumlah kandidat kata dasar yang dihasilkan serta akurasi secara keseluruhan. Hipotesis sementara adalah semakin besar ukuran kamus maka diharapkan akurasi akan meningkat, namun tentu saja hal ini perlu dibuktikan melalui suatu percobaan. Adapun agar dapat mengukur akurasi untuk memilih metode *stemming* yang paling optimal, dibutuhkan data dan metode uji yang seragam.

Selain itu, penelitian lebih lanjut juga dibutuhkan untuk menangani kata yang terlihat seperti kata berimbuhan namun sebenarnya bukan, misalnya singkatan dan sebagainya. Konteks kalimat perlu diidentifikasi sehingga proses *stemming* dapat menghasilkan hasil yang lebih akurat, sesuai dengan konteks kalimatnya.

## DAFTAR PUSTAKA

- [1] Manning, C.D., Raghavan, P., and Schütze, H. (2009): *An Introduction to Information Retrieval*, Cambridge University Press, England, p. 117-120
- [2] Moral, Cristian, et al. "A survey of stemming algorithms in information retrieval." *Information Research: An International Electronic Journal* 19.1 (2014): n1.
- [3] Mayfield, J. & McNamee, P. (2003). Single n-gram stemming. In *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, (pp. 415-416). New York, NY: ACM Press.
- [4] Peng, F., Ahmed, N., Li, X. & Lu, Y. (2007). Context sensitive stemming for Web search. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, (pp. 639-646). New York, NY: ACM Press.
- [5] Setiawan, Reina, Aditya Kurniawan, Widodo Budiharto, Iman Herwidiana Kartowisastro, and Harjanto Prabowo. "Flexible affix classification for stemming Indonesian Language." In *Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 2016 13th International Conference on*, pp. 1-6. IEEE, 2016
- [6] Indradjaja, Lily Suryana, and Stephane Bressan. "Automatic learning of stemming rules for the Indonesian language." *Proc. of the The 17th Pacific Asia Conference on Language, Information and Computation (PACLIC)*. 2003.
- [7] Asian, Jelita, Hugh E. Williams, and Seyed MM Tahaghoghi. "Stemming Indonesian." *Proceedings of the Twenty-eighth Australasian conference on Computer Science-Volume 38*. Australian Computer Society, Inc., 2005.
- [8] Adriani, M., Asian, J., Nazief, B., Tahaghoghi, S.M. and Williams, H.E., 2007. Stemming Indonesian: A confix-stripping approach. *ACM Transactions on Asian Language Information Processing (TALIP)*, 6(4), pp.1-33.
- [9] Arifin, A., Ciptaningtyas, H., & Mahendra, I. (2009). Enhanced Confix Stripping Stemmer And Ants Algorithm For Classifying News Document In Indonesian Language. *The International Conference on Information & Communication Technology and Systems*, 5, pp. 149-158.
- [10] Suhartono, Derwin. "Lemmatization Technique in Bahasa: Indonesian." *Journal of Software* 9.5 (2014): 1203
- [11] Sinaga, Ardiles, and Hertog Nugroho. "Development of word-based text compression algorithm for Indonesian language document." In *Information and Communication Technology (ICoICT), 2015 3rd International Conference on*, pp. 450-454. IEEE, 2015.
- [12] Widjaja, Marsel, and Seng Hansun. "Implementation of Porter's Modified Stemming Algorithm in an Indonesian Word Error Detection Plugin Application." *International Journal Of Technology* 6, no. 2 (2015): 139-150
- [13] Suhendar, M. E., and Pien Supinah. "MKDU (Mata Kuliah Dasar Umum) Bahasa Indonesia." *Balai Pustaka* (1995).
- [14] Purwarianti, A., 2011, July. A non deterministic Indonesian stemmer. In *Electrical Engineering and Informatics (ICEEI), 2011 International Conference on* (pp. 1-5). IEEE