

Pemenggalan Kata Dasar Bahasa Indonesia menggunakan *Syllabification Algorithm*

Novi Yusliani, M.T.¹, Rezi Apriliansyah², Tiara Windri Apriani³

Jurusan Teknik Informatika, Fakultas Ilmu Komputer, Universitas Sriwijaya

Jalan Palembang-Prabumulih Km.32 Inderalaya Ogan Ilir Kode Pos 30662

e-mail: novi_yusliani@unsri.ac.id¹, reziapriiliansyah@gmail.com², tiarawindri@outlook.com³

Abstract — This paper presents a study of Indonesian syllable structure and an algorithm for identifying syllables in Indonesian words. The algorithm implements a set of syllabification rules based on the Regulation of the Minister of Education and Culture of the Republic of Indonesia Number 50 Year 2015 on General Guidelines for Indonesian Spelling. Experiment on a random sample shows that the syllabification algorithm achieves 100% word accuracy. Random sample contains 300 Indonesian words. Weakness of this algorithm can not handle Indonesian words contain ‘ng’, ‘ny’, ‘kh’, ‘sy’, ‘ks’, and words contain three consecutive consonant letters.

Keywords — text to speech system, speech synthesis, syllabification algorithm, rule-based method

Abstrak - Makalah ini menyajikan studi tentang struktur suku kata bahasa Indonesia dan algoritma untuk mengidentifikasi suku kata dalam kata-kata bahasa Indonesia. Algoritma mengimplementasikan seperangkat aturan pembagian suku kata berdasarkan Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 50 Tahun 2015 tentang Pedoman Umum Ejaan Bahasa Indonesia. Percobaan pada sampel acak menunjukkan bahwa algoritma pembagian suku kata mencapai akurasi kata 100%. Sampel acak mengandung 300 kata bahasa Indonesia. Kelemahan dari algoritma ini tidak dapat menangani kata-kata bahasa Indonesia berisi ‘ng’, ‘ny’, ‘kh’, ‘sy’, ‘ks’, dan kata-kata mengandung tiga huruf konsonan berturut-turut.

Kata kunci - sistem text to speech, speech synthesis, algoritma pembagian suku kata, metode rule-based

I. PENDAHULUAN

Speech synthesis merupakan salah satu komponen terpenting di dalam *text-to-speech system*. Komponen ini melakukan *encoding* dan menghasilkan suara dari kata yang dimasukkan ke dalam sistem. Salah satu pendekatan yang dapat digunakan dalam proses pelafalan kata adalah *syllable*. Proses pelafalan kata dengan menggunakan pendekatan *syllable* menghasilkan kualitas suara yang lebih baik [2]. Pada pendekatan *syllable*, pelafalan kata dilakukan berdasarkan suku kata. Suku kata dihasilkan melalui proses pemenggalan kata.

Proses pemenggalan kata menjadi suku kata dapat dilakukan dengan menggunakan *syllabification algorithm*. Pengembangan *syllabification algorithm* telah banyak dilakukan, di antaranya [1][3][4]. [3] mengembangkan *syllabification algorithm* untuk bahasa Spanyol dengan menggunakan pendekatan *rule (rule-based)*. Aturan-aturan yang diimplementasikan ke dalam algoritma dibangkitkan

berdasarkan aturan tata bahasa (*grammatical rules*) dari bahasa Spanyol. Hasil pengujian menunjukkan *error rate* dari algoritma adalah 1,6%.

Syllabification algorithm untuk bahasa Sinhala, yaitu bahasa resmi negara Sri Lanka, telah dikembangkan [4]. Algoritma yang dikembangkan juga menggunakan pendekatan *rule*. Aturan-aturan tersebut dibangkitkan berdasarkan aturan linguistik dari bahasa Sinhala. Hasil pengujian menunjukkan bahwa tingkat akurasi dari algoritma ini adalah 99,95%.

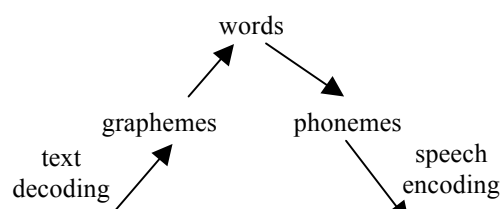
Pada bahasa Malaysia, *syllabification algorithm* dengan menggunakan pendekatan *rule* dikembangkan oleh [1]. Aturan-aturan tersebut dibangkitkan berdasarkan morfologi dan struktur suku kata dari bahasa Malaysia. Pengujian dilakukan dengan menggunakan sumber tiga koleksi data yang berbeda. Penelitian ini menghasilkan tingkat akurasi sebesar 60,7% pada koleksi data BERNAMA, 77,4% pada koleksi data Kamus Dewan Collection, dan 71,6% pada koleksi data Overlap.

Karena itu, penelitian ini bertujuan untuk mengembangkan *syllabification algorithm* untuk kata dasar pada Bahasa Indonesia. Algoritma yang dikembangkan menggunakan pendekatan *rule*. *Rule* dibangkitkan berdasarkan aturan pemenggalan kata pada peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 50 Tahun 2015 tentang Pedoman Umum Ejaan Bahasa Indonesia.

II. LANDASAN TEORI

A. Text to Speech System

Text to speech system adalah sistem yang berfungsi untuk mengubah teks menjadi suara secara otomatis (Dutoit, 1997) dalam [5]. Sistem ini terdiri dari dua komponen, yaitu *text analysis* dan *speech synthesis* [6]. *Text analysis* merupakan komponen yang bertugas melakukan *decoding* terhadap kata yang menjadi masukan sistem. *Decoding* adalah proses untuk mengubah pola tulisan (*signal written*) dari kata tersebut menjadi sebuah pesan [6]. Selanjutnya pesan yang dihasilkan akan diproses ke dalam komponen *speech synthesis*. Komponen ini akan melakukan *encoding* terhadap pesan yang diterima. *Encoding* merupakan proses pembuatan sinyal dari pesan [6]. Keluaran dari komponen ini adalah pelafalan kata (suara) dari kata yang dimasukkan ke dalam sistem. Gambar 1 memberikan gambaran tahapan perubahan kata menjadi suara (lafaz dari kata).



GBR. 1 GAMBARAN UMUM DARI *TEXT-TO-SPEECH SYSTEM* [6]

B. Kata

Kata dalam Bahasa Indonesia terdiri atas satu suku kata atau lebih, misalnya “ban”, “bantu”, “membantu”, “memperbantukan” [7]. Suku kata adalah bagian kata yang diucapkan dalam satu hembusan nafas dan umumnya terdiri dari beberapa fonem [7]. Wujud suku yang membentuknya mempunyai struktur dan kaidah pembentukan yang sederhana. Salah satu contoh adalah kata ‘sapu’. Kata ‘sapu’ diucapkan dengan dua hembusan napas: satu untuk ‘sa-’ dan satunya lagi untuk ‘-pu’. Dengan demikian, kata ‘sapu’ terdiri atas dua suku kata. Tiap suku terdiri atas dua bunyi, yaitu [sa] dan [pu].

Suku kata dalam Bahasa Indonesia dapat terdiri atas (1) satu vokal, (2) satu vokal dan satu konsonan, (3) satu konsonan dan satu vokal, (4) satu konsonan, satu vokal, dan satu konsonan, (5) dua konsonan dan satu vokal, (6) dua konsonan, satu vokal, dan satu konsonan, (7) satu konsonan, satu vokal, dan dua konsonan, (8) tiga konsonan dan satu vokal, (9) tiga konsonan, satu vokal, dan satu konsonan, (10) dua konsonan, satu vokal, dan dua konsonan, serta (11) satu konsonan, satu vokal, dan tiga konsonan. Gambar 2 menunjukkan contoh suku kata.

1. V	a-mal
2. VK	ar-ti
3. KV	pa-sar
4. KVK	pak-sa
5. KVKK	teks-til
6. KVKKK	korps
7. KKV	slo-gan
8. KKVK	trak-tor
9. KKKV	stra-ta
10. KKKVK	struk-tur
11. KKVKK	kom-pleks
Ket: V - Vokal K - Konsonan	

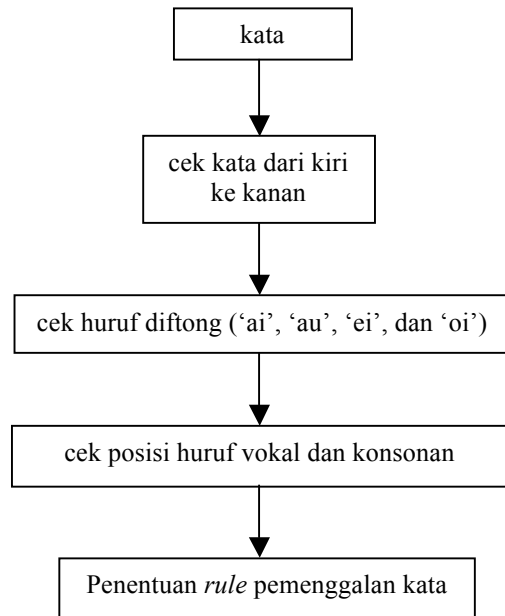
GBR. 2 CONTOH SUKU KATA [7]

Pada Bahasa Indonesia proses pemenggalan kata diatur di dalam peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 50 Tahun 2015 tentang Pedoman Umum Ejaan Bahasa Indonesia. Pemenggalan kata pada kata dasar dilakukan dengan menggunakan aturan berikut ini:

- a. Jika di tengah kata terdapat huruf vokal yang berurutan, pemenggalannya dilakukan di antara kedua huruf vokal itu.
Contoh: bu-ah ma-in

- b. Huruf diftong ‘ai’, ‘au’, ‘ei’, dan ‘oi’ tidak dipenggal.
Contoh: pan-dai au-la
sau-da-ra sur-vei
am-boi
 - c. Jika di tengah kata dasar terdapat huruf konsonan (termasuk gabungan huruf konsonan) di antara dua huruf vokal, pemenggalannya dilakukan sebelum huruf konsonan itu.
Contoh: ba-pak la-wan
de-ngan ke-nyang
mu-ta-khir mu-sya-wa-rah
 - d. Jika di tengah kata dasar terdapat dua huruf konsonan yang berurutan, pemenggalannya dilakukan di antara kedua huruf konsonan itu.
Contoh: Ap-riil cap-lok
makh-luk man-di
sang-gup som-bong
swas-ta
 - e. Jika di tengah kata dasar terdapat tiga huruf konsonan atau lebih yang masing-masing melambangkan satu bunyi, pemenggalannya dilakukan di antara huruf konsonan yang pertama dan huruf konsonan yang kedua.
Contoh: ul-tra in-fra
ben-trok in-stru-men
- Sedangkan gabungan huruf konsonan yang melambangkan satu bunyi tidak dipenggal.
Contoh: bang-krut bang-sa
ba-nyak ikh-las

III. SYLLABIFICATION ALGORITHM



GBR. 3 BAGAN ALUR SYLLABIFICATION ALGORITHM

Syllabification algorithm yang dikembangkan menggunakan pendekatan *rule (rule-based)*. Aturan-aturan yang diimplementasikan ke dalam algoritma dibangkitkan berdasarkan aturan pemenggalan kata dalam peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 50 Tahun 2015. Bagan alur algoritma pada Gbr 3 menunjukkan tahapan-tahapan pemenggalan kata berdasarkan aturan yang telah dibangkitkan.

Berdasarkan Gbr. 3 proses pengecekan keberadaan huruf diftong ('ai', 'au', 'ei', dan 'oi') pada kata dilakukan terlebih dahulu. Hal ini dikarenakan huruf diftong tidak dipenggal. Setelah pengecekan huruf diftong, dilakukan pengecekan huruf vokal dan huruf konsonan beserta posisinya pada kata. Proses pemenggalan kata (penentuan *rule* pemenggalan kata) dilakukan setelah tanda diberikan, yaitu berupa indeks. Proses pemberian tanda dilakukan setelah proses pengecekan huruf diftong dan pengecekan huruf vokal juga huruf konsonan. Gbr. 4 menunjukkan *syllabification algorithm* yang telah dikembangkan.

1. *Scan* kata dari kiri ke kanan
2. Cek huruf diftong ('ai', 'ei', 'au', dan 'oi') dalam kata, jika:
 - a. di depan, beri indeks setelah huruf diftong
 - b. di tengah, beri indeks sebelum dan setelah huruf diftong,
 - c. di belakang, beri indeks setelah huruf diftong.
3. Pengecekan huruf vokal dan konsonan:
 - a. jika huruf ke-i konsonan dan huruf ke-i+1 konsonan, beri indeks setelah huruf ke-i.
 - b. Jika huruf ke-i vokal dan huruf ke-i+1 vokal, beri indeks setelah huruf ke-i.
 - c. Jika huruf ke-i vokal dan huruf ke-i+1 konsonan dan huruf ke-i+2 vokal, beri indeks setelah huruf ke-i
4. Lakukan pemenggalan sesuai indeks yang telah ditentukan.

GBR. 4 SYLLABIFICATION ALGORITHM UNTUK KATA DASAR BAHASA INDONESIA

IV. HASIL PENGUJIAN

Data yang digunakan dalam pengujian adalah kata dasar yang telah diketahui posisi pemenggalan kata tersebut. Jumlah kata dasar yang digunakan untuk pengujian adalah 300 kata dasar. Hasil pengujian menunjukkan algoritma berhasil memenggal semua kata dasar dengan benar. Tabel 1 menunjukkan hasil pengujian terhadap *syllabification algorithm* yang dikembangkan.

TABEL 1. HASIL PENGUJIAN

Jumlah Kata Dasar	Benar	Salah	Akurasi
300	300	0	100%

300	300	0	100%
-----	-----	---	------

Hasil pengujian menunjukkan algoritma berhasil memenggal kata dasar dengan benar untuk 300 kata dasar yang diujikan. Perhitungan akurasi *syllabification algorithm* dilakukan dengan menggunakan persamaan 1. Tingkat akurasi dari *syllabification algorithm* yang dikembangkan adalah 100%.

$$\text{Akurasi} = \frac{\text{Jumlah Data Uji Benar}}{\text{Jumlah Data Uji}} \times 100\% \quad (1)$$

Contoh kata dasar yang digunakan dalam pengujian *syllabification algorithm* ini dapat dilihat pada tabel 2. Pengujian dikatakan benar apabila algoritma memberikan hasil pemenggalan kata yang sesuai dengan data yang ada.

TABEL 2. CONTOH DATA UJI

No	Kata Dasar	Pemenggalan
1	abadi	<i>aba-di</i>
2	sosiologi	<i>so-si-o-lo-gi</i>
3	standarisasi	<i>stan-dar-di-sa-si</i>
4	wawancara	<i>wa-wan-ca-ra</i>
5	teknologi	<i>tek-no-lo-gi</i>
6	akomodasi	<i>ako-mo-da-si</i>
7	akrobat	<i>ak-ro-bat</i>
8	diagnosis	<i>di-ag-no-sis</i>
9	bendahara	<i>ben-da-ha-ra</i>
10	analisis	<i>ana-li-sis</i>
11	biosfer	<i>bi-o-sfer</i>
12	artikulasi	<i>ar-ti-ku-la-si</i>
13	asimilasi	<i>asi-mi-la-si</i>
14	atmosfer	<i>at-mo-sfer</i>
15	bandara	<i>ban-da-ra</i>
16	diplomasi	<i>dip-lo-ma-si</i>
17	desain	<i>de-sa-in</i>
18	depresi	<i>de-pre-si</i>
19	bencana	<i>ben-ca-na</i>
20	dekoratif	<i>de-ko-ra-tif</i>
21	daerah	<i>da-e-rah</i>
22	bendahara	<i>ben-da-ha-ra</i>
23	beranda	<i>be-ran-da</i>
24	efisien	<i>efi-si-en</i>
25	biosfer	<i>bi-o-sfer</i>
26	cendekiawan	<i>cen-de-ki-a-wan</i>
27	brigadir	<i>bri-ga-dir</i>
28	buah	<i>bu-ah</i>

29	identifikasi	<i>iden-ti-fi-ka-si</i>
30	indonesia	<i>in-do-ne-sia</i>

V. KESIMPULAN

Syllabification algorithm adalah algoritma yang berfungsi untuk memenggal kata menjadi bagian kata yang diucapkan dalam satu hembusan nafas dan umumnya terdiri dari beberapa fonem. Salah satu pendekatan yang dapat digunakan untuk mengembangkan algoritma ini adalah pendekatan *rule (rule-based)*. Aturan yang diimplementasikan ke dalam *syllabification algorithm* adalah aturan-aturan yang diperoleh dengan menganalisis morfologi kata, struktur suku kata, atau berdasarkan aturan pemenggalan kata yang telah ditentukan.

Syllabification algorithm yang dikembangkan dalam makalah ini menggunakan aturan pemenggalan kata yang telah ditetapkan dalam peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 50 Tahun 2015 tentang Pedoman Umum Ejaan Bahasa Indonesia. Pengujian algoritma dilakukan dengan menggunakan 300 kata dasar yang telah diketahui posisi pemenggalan kata tersebut. Tingkat akurasi algoritma yang dihasilkan adalah 100%. Kelemahan dari algoritma ini adalah tidak dapat menangani kata dasar yang mengandung ‘ng’, ‘ny’, ‘kh’, ‘ks’, ‘sy’, dan kata dasar yang mengandung tiga huruf konsonan yang berurutan.

REFERENSI

- [1] M. Hafiz, K. Rabiah, A. A. Azreen, and A. M Taufik, “Syllabification Algorithm Based on Syllable Rules Matching for Malay Language,” Proc. 10th WSEAS Int. Conf. Appl. Comput. Appl. Comput. Sci., pp. 279–286, 2011.
- [2] E. Veera Raghavendra, B. Yegnanarayana, and K. Prahallad, “Speech Synthesis using Approximate Matching of Syllables,” 2008 IEEE Spok. Lang. Technol. Work., no. December, pp. 37–40, 2008.
- [3] H. Cuayáhuítl, “A Syllabification Algorithm for Spanish,” Comput. Linguist. Intell. Text, pp. 412–415, 2004.
- [4] R. Weerasinghe, A. Wasala, and K. Gamage, “A Rule based Syllabification Algorithm for Sinhala,” Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 3651 LNAI, pp. 438–449, 2005.
- [5] J. O. Onaolapo, F. E. Idachaba, J. Badejo, T. Odu, and O. I. Adu, “A Simplified Overview of Text-To-Speech Synthesis,” Proc. World Congr. Eng., vol. I, pp. 5–7, 2014.
- [6] P. Taylor, “Text-to-Speech Synthesis,” Signal Processing, p. 642, 2009.
- [7] H. Alwi, S. Dardjowidjojo, H. Lapoliwa, and A. M. Moeliono, “Tata Bahasa Baku Bahasa Indonesia.” BalaiPustaka, 2003.