

Silabifikasi Menggunakan Metode *Rule-Based* dalam Bahasa Indonesia

Aditya Rahman¹, Danang Triantoro Murdiansyah², Kemas Muslim Lhaksana³

^{1,2,3} Universitas Telkom, Bandung

¹adityaok@students.telkomuniversity.ac.id, ²kemasmuslim@telkomuniversity.ac.id,

³danangtri@telkomuniversity.ac.id

Abstrak

Silabifikasi kata adalah suatu proses pembagian kata menjadi suku kata. Silabifikasi kata penting dilakukan pada proses *Text to Speech System* dan *Continuous Speech Recognizer* agar memperoleh hasil yang optimal. Pada penelitian ini penulis akan mengembangkan algoritma silabifikasi kata dengan menerapkan metode *Rule-Based*. *Rule-Based* adalah Metode yang diciptakan untuk menyelesaikan suatu masalah dengan cara menjabarkan pola dari suatu masalah. Hasil pengujian algoritma silabifikasi kata menggunakan metode *Rule-Based* dalam bahasa Indonesia pada dataset "*Trainset Syllabification 01 40K 1 2 3 4*" dengan menggunakan *WER (Word Error Rate)* memperoleh nilai 0,029. Algoritma silabifikasi yang didapat dibuat menjadi aplikasi silabifikasi kata dasar dalam bahasa Indonesia. Kekurangan dari aplikasi yang dibuat adalah masih terdapat nya eror pada beberapa kata dalam bahasa Indonesia.

Kata kunci : Silabifikasi, *Rule-Based*, Bahasa Indonesia

Abstract

Word syllabification is a process of dividing words into syllables. It is important to do word syllabus in the *Text to Speech System* and *Continuous Speech Recognizer* processes in order to obtain optimal results. In this study the author will develop a word syllabic algorithm by applying the *Rule-Based* method. *Rule-Based* is a method created to solve a problem by describing the pattern of a problem. The results of testing the word syllabus algorithm using the *Rule-Based* method in Indonesian on the "*Trainset Syllabification 01 40K 1 2 3 4*" dataset using *WER (Word Error Rate)* obtained a value of 0.029. The syllabic algorithm obtained is made into an application of basic word syllabus in Indonesian. The drawback of the application made is that there are still errors in some words in Indonesian.

Keywords: Syllabification, *Rule-Based*, Indonesian language

1. Pendahuluan

Latar Belakang

Komponen terpenting dalam interaksi antara individu satu dengan individu lain nya adalah Bahasa. Dalam kehidupan manusia pasti ada interaksi antara manusia satu dengan manusia lain nya [3]. Bahasa resmi bangsa Indonesia adalah bahasa Indonesia [1]. Identitas bangsa Indonesia tercermin dari bahasa Indonesia [2]. Bahasa Indonesia berperan untuk menyampaikan suatu informasi [7]. Kata adalah salah satu bagian dari bahasa. Untuk memahami cara membaca suatu kata kita dapat melakukan silabifikasi kata. Silabifikasi kata adalah proses membagi kata menjadi suku kata menurut aturan tertentu[4]. *Syllabification algorithm* adalah salah satu cara untuk melakukan silabifikasi kata [8]. Salah satu contoh penerapan *Syllabification algorithm* yaitu penerapan nya pada bahasa Spanyol [3]. Aturan tata bahasa Spanyol digunakan sebagai acuan pada penerapan *Syllabification algorithm* pada bahasa Spanyol [3]. Nilai error rate dari *Syllabification algorithm* pada bahasa Spanyol adalah 1,6%[3].

Salah satu penerapan silabifikasi dimasa kini adalah penerapan nya pada komponen *Text to Speech System* [5]. Dengan memanfaatkan silabifikasi kita dapat memperoleh hasil dari suatu sistem *Text to Speech* yang lebih baik dan alami [5]. Penerapan lain dari silabifikasi yaitu pada *Continuous Speech Recognizer* [9]. Hasil menunjukkan *Continuous Speech Recognizer* menggunakan silabifikasi memperoleh hasil yang lebih baik ketimbang menggunakan HMM (Hidden Markov Model) [9].

Pada penelitian ini penulis akan mengembangkan suatu algoritma silabifikasi menggunakan metode *Rule-Based*, setelah algoritma berhasil di buat penulis akan membuat aplikasi yang dapat melakukan silabifikasi dalam bahasa Indonesia. Dataset yang dipakai pada pengujian kali ini adalah "*Trainset Syllabification 01 40K 1 2 3 4*". Pada metode *Rule-Based* algoritma silabifikasi dibuat dari proses pembelajaran dataset, pola yang didapat dari proses pembelajaran tersebut akan menjadi aturan silabifikasi kata. Aturan silabifikasi kata yang didapat akan diterapkan untuk melakukan silabifikasi kata dasar bahasa Indonesia yang terdapat dalam dataset. Evaluasi

algoritma silabifikasi akan dilakukan dengan cara menghitung *Word Error Rate (WER)*[6]. Semakin rendah nilai *Word Error Rate* maka semakin bagus algoritma yang dibuat.

Topik dan Batasannya

Berdasarkan pada apa yang telah di jelaskan pada bagian latar belakang, permasalahan yang akan ditelisik pada penelitian ini adalah bagai mana menerapkan metode *Rule-Based* kedalam algoritma silabifikasi bahasa Indonesia. Sedangkan batasan masalah dari penelitian ini adalah algoritma silabifikasi *Rule-Based* hanya akan digunakan pada bahasa indonesia, aplikasi yang dibuat dari algoritma silabifikasi berbasis python dan hanya dapat di akses menggunakan windows.

Tujuan

Tujuan dari dilakukan nya penelitian yaitu mempergunakan metode *Rule-Based* pada algoritma silabifikasi dan membangun aplikasi dari algoritma silabifikasi yang telah di buat.

Organisasi Tulisan

Bab2 menjelaskan tentang penelitian yang relevan dan menjadi acuan bagi penelitian kali ini. Bab 3 menjelaskan sistem yang dibangun untuk penelitian kali ini. Bab 4 menjelaskan hasil dan analisis dari hasil yang telah di dapat. Bab 5 menjelaskan kesimpulan atas hasil penelitian yang di lakukan.

2. Studi Terkait

2.1 Kata

Suatu satuan bahasa yang memiliki arti dapat disebut sebagai kata [10]. Kata merupakan satuan bahasa terkecil dalam bahasa Indonesia [10]. Kosakata dalam bahasa indonesia terbagi menjadi 2 jenis yakni: kosakata umum dan kosa kata khusus. Katadasar atau kata turunan yang sudah lumrah dipakai dalam berinteraksi antar individu disebut kosakata umum, sedangkan kata dasar atau kata turunan yang jarang atau tidak lumrah digunakan dalam berinteraksi antar individu dapat disebut sebagai kosakata khusus [11].

2.2 Suku kata

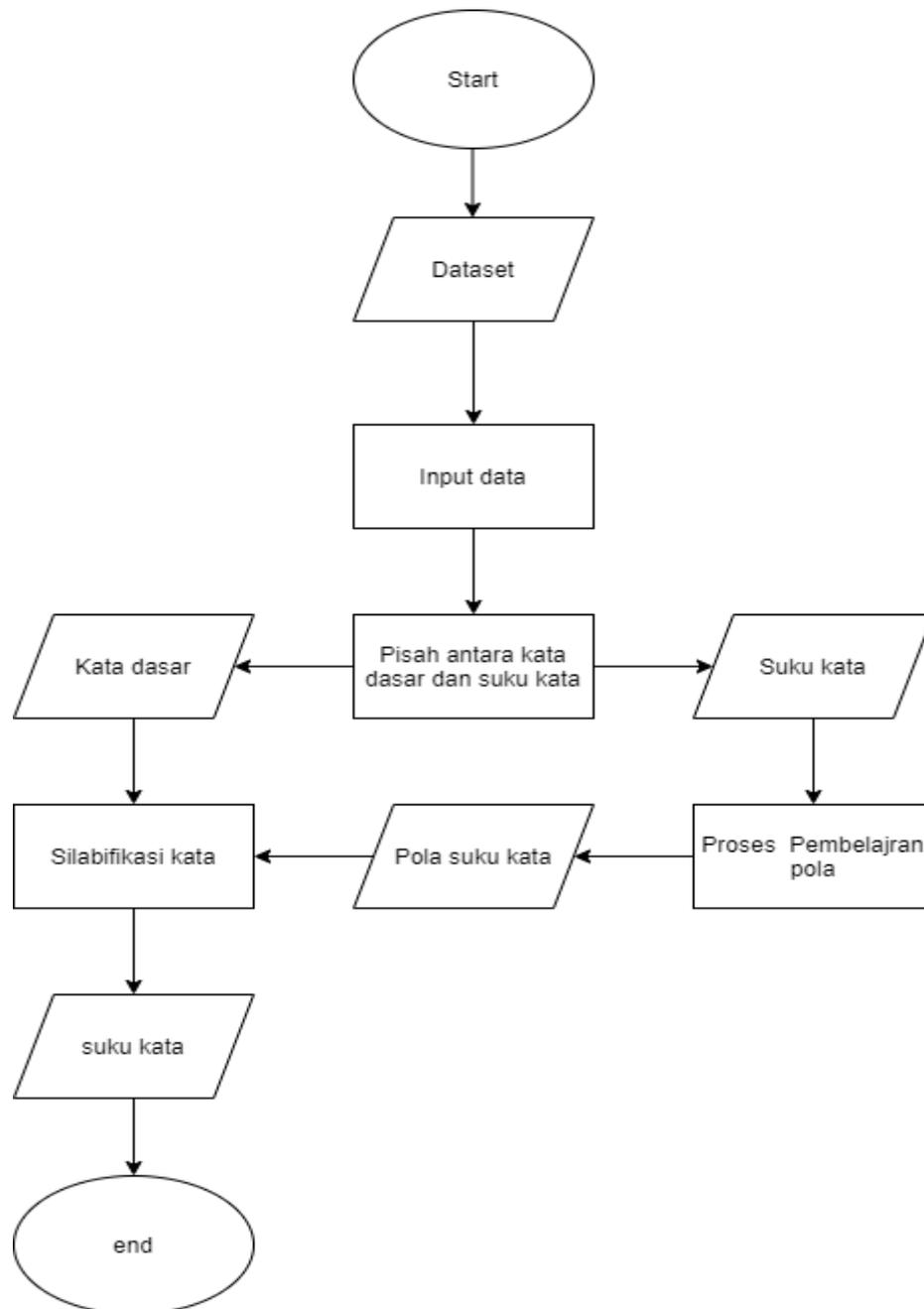
Kumpulan huruf yang merupakan bagian dari suatu kata yang pelafalan nya dilakukan dalam suatu desiran dan memiliki suatu lambang bunyi atau beberapa lambang bunyi dapat disebut sebagai suku kata[7]. Aturan yang diterapkan dalam suatu suku kata dari suatu kata umumnya semenjana [8]. Adapun contoh suku kata dari kata “amerta” adalah “amer” dan “ta”, hal ini dikarnakan kata amerta dilafalkan dalam dua desiran yaitu: “amer” untuk desiran pertama dan “ta” untuk desiran kedua.

2.3. Rule-Based

Metode yang diciptakan untuk menyelesaikan suatu masalah dengan cara menjabarkan pola dari suatu masalah disebut sebagai metode *Rule-based* [12]. Pada umumnya suatu aturan dalam *Rule-Based* terdiri dari duabagian yaitu suatu kondisi dan tindakan yang dilakukan bila suatu kondisi terpenuhi seluruhnya atau sebagian [12]. Peraturan-peraturan yang di dapat akan di implementasikan ke dalam suatu algoritma [12]. Algoritma akan menentukan peraturan yang dianggap tepat terhadap suatu permasalahan [12]. Kekurangan dari *Rule-Based* adalah tidak dapat melahirkan aturan anyar, hal ini dikarnakan *Rule-Based* tidak dapat melakukan *Learning*[13].

3. Sistem yang Dibangun

Sistem yang akan di bangun pada penelitian kali ini bertujuan melakukan silabifikasi pada suatu kata. Sumber data yang dipakai pada penelitian ini adalah dataset “*Trainset Syllabification 01 40K 1 2 3 4*”. Dataset berisi kata dalam bahasa Indonesia dan silabifikasi kata itu sendiri. Kata dalam dataset tersebut akan dilakukan proses silabifikasi dengan metode *Rule-Based* berbasis aturan pemenggalan kata hasil pembelajaran pola dataset itu sendiri. Berikut ini adalah skenario sistem yang akan di bangun:



Gambar 1 *Flowchart* Diagram Algoritma Silabifikasi Menggunakan Metode *Rule-Based* dalam Bahasa Indonesia

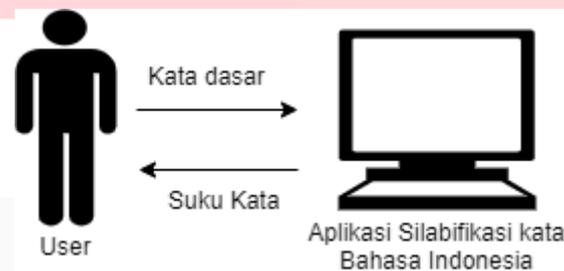
Brikut ini adalah penjabaran dari sistem yang akan dibangun:

1. Input data
Sumber data yang akan dipakai adalah dataset “*Trainset Syllabification 01 40K 1 2 3 4*” Dataset berisi kata dalam bahasa indonesia dan silabifikasi kata itu sendiri.
2. Pemisahan kata dan suku kata
Kata dan suku kata yang terdapat dalam dataset akan dipisahkan satu sama lain.
3. Pembelajaran pola
Pada tahap ini akan dilakukan proses pembelajaran pola suku kata. Hasil dari proses ini adalah pola suku kata.
4. Silabifikasi kata
Silbifikasi akan dilakukan berdasarkan pola yang telah didapat pada proses pembelajaran pola suku kata. Hasil dari proses ini adalah suku kata.

Setelah algoritma silabifikasi di dapat maka akan dibuat program untuk silabifikasi kata dalam bahasa Indonesia. Pada program yang akan dibuat *user* dapat menginputkan kata dasar dalam bahasa Indonesia dengan catatan untuk setiap penulisan huruf diftong dan gabungan konsonan makan akan dirubah menjadi suatu simbol. Simbol yang digunakan terdapat pada tabel 1. Setelah *user* menginputkan kata dasar dalam bahasa Indonesia, sistem akan melakukan proses silabifikasi kata. Setelah proses silabifikasi selesai makan akan muncul suku kata dari kata dasar yang diinputkan *user*. Gambaran sistem terdapat pada gambar 2.

No	Diftong dan gabungan konsonan	Simbol
1	au	@
2	ai	\$
3	ei	%
4	oi	^
5	kh	(
6	ng)
7	sy	~
8	ny	+

Tabel 1 Simbol Diftong dan Gabungan Konsonan



Gambar 2 Gambaran Sistem Aplikasi Silabifikasi Kata

4. Evaluasi

4.1 Hasil Pengujian

Dataset yang digunakan dalam pengujian algoritma silabifikasi *Rule-Based* dalam bahasa Indonesia adalah “*Trainset Syllabification 01 40K 1 2 3 4*”. Dataset berisi kata dasar dan suku kata dari kata tersebut. Dataset berjumlah 40456 kata. Berikut ini adalah contoh kata dasar yang di uji dan suku kata yang diperoleh dari proses silabifikasi:

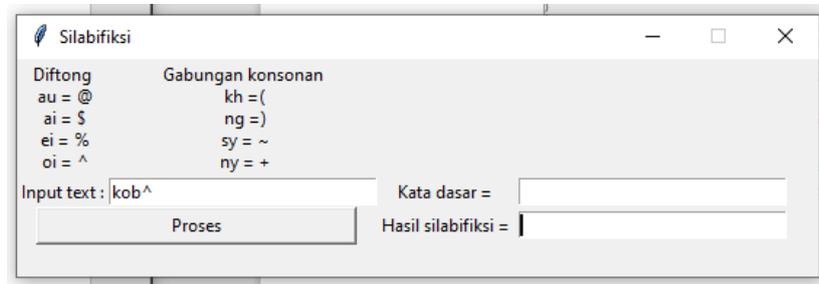
No	Kata dasar	Suku kata hasil silabifikasi
1	abadi	a-ba-di
2	abduksi	ab-duk-si
3	amprok	am-prok
4	ampuh	am-puh
5	anabasis	a-na-ba-sis
6	amuba	a-mu-ba
7	album	al-bum

Tabel 2 Contoh Kata Dasar yang di Uji dan Suku Kata yang Diperoleh dari Proses Silabifikasi

Prose evaluasi dilakukan dengan cara menghitung *Word Error Rate (WER)* yang didapat dari hasil pembagian jumlah kata yang salah di silabifikasi dibagi dengan total kata. Hasil silabifikasi suatu kata dinyatakan benar bila sesuai dengan suku kata yang terdapat dalam dataset. Semakin kecil nilai *WER* maka semakin baik model silabifikasi yang dibuat. Pada penelitian ini *WER* yang diperoleh sebesar 0,029. Dibawah ini adalah formula yang dipakai untuk menentukan nilai *WER* :

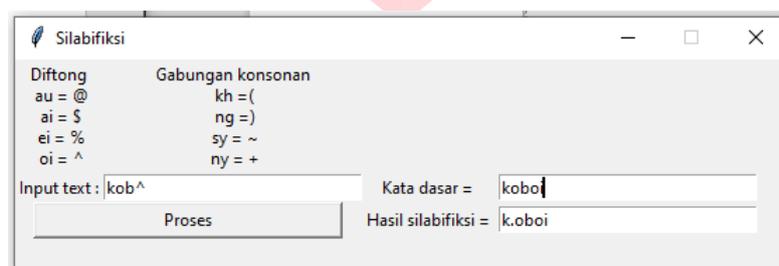
$$WER = \frac{\text{Jumlah suku kata yang salah}}{\text{Total jumlah kata}}$$

Pengujian selanjutnya adalah tentang program silabifikasi yang dibuat. Pengujian dilakukan dengan cara *user* menginputkan sutau kata dasar dalam bahasa Indonesia, huruf diftong dan gabungan konsonan di input menggunakan simbol yang telah di tentukan.



Gambar 3 Proses Input Kata Dasar

Ketika *user* mengklik tombol proses maka program silabifikasi akan menampilkan kata dasar dan hasil silabifikasi kata tersebut.



Gambar 4 Hasil Proses Silabifikasi Kata

4.2 Analisis Hasil Pengujian

Algoritma silabifikasi menggunakan metode *Rule-based* mampu melakukan silabifikasi pada kata dalam bahasa Indonesia. Pada pengujian yang telah dilakukan teknik silabifikasi *Rule-Based* berhasil memperoleh nilai *WER* 0,029 (terdapat 1173 kata yang salah disilabifikasi dari 40456 kata bahasa Indonesia). Pada pengujian aplikasi silabifikasi terbukti bahwa program yang dibuat dapat melakukan silabifikasi pada kata dasar bahasa Indonesia yang di inputkan oleh *user*. Namun pada beberapa kata dasar aplikasi masih mengalami eror.

5. Kesimpulan

Dari hasil pengujian dapat disimpulkan:

1. Metode *Rule-Based* dapat diimplementasikan pada algoritma silabifikasi kata dalam bahasa Indonesia.
2. Algoritma silabifikasi yang diperoleh dapat diimplementasikan untuk membuat aplikasi silabifikasi kata.
3. Aplikasi silabifikasi kata masih terdapat eror bila *user* menginputkan beberapa kata dalam bahasa Indonesia.

Referensi

- [1] Arifin, H. Haryanto, B Pratama, 2019, SEGMENTASI SUKU KATA BAHASA INDONESIA BERBASIS ATURAN UNTUK PENGEMBANGAN SISTEM TEXT-TOAUDIOVISUAL, Indonesia, Prosiding SNATIF.
- [2] M. Ekawati, 2017, KESANTUNAN SEMU PADA TINDAK TUTUR EKSPRESIF MARAH DALAM BAHASA INDONESIA, Indonesia, Adabiyāt: Jurnal Bahasa dan Sastra, Vol. I , No. 1, Juni 2017.
- [3] K. T. Martono, 2011, Augmented Reality Sebagai Metafora Baru dalam Teknologi Interaksi Manusia dan Komputer, Indonesia, JURNAL SISTEM KOMPUTER-VOL.1, NO.2.
- [4] S. Bartlett, G. Kondrak, C. Cherry, Automatic Syllabification with Structured SVMs for Letter-To-Phoneme Conversion, Canada, Department of Computing Science Research University of Alberta.
- [5] Gelbukh, Alexander and Heriberto Cuayáhuitl. A Syllabification Algorithm for Spanish. Heidelberg: Springer Berlin, 2004.
- [6] I. Ramlia, N. Jamila, N. Semana, N. Ardib, 2015, An Improved Syllabification for a Better Malay Language Text-toSpeech Synthesis (TTS), Malaysia, Elsevier B.V.
- [7] R. P. Suminar, 2016, PENGARUH BAHASA GAUL TERHADAP PENGGUNAAN BAHASA INDONESIA MAHASISWA UNSWAGATI, Indonesia, JURNAL LOGIKA, Vol XVIII, No 3, Desember 2016.
- [8] N. Yusliani, R. Apriliansyah, T. W. Apriani, Pemenggalan Kata Dasar Bahasa Indonesia menggunakan Syllabification Algorithm, Indonesia.
- [9] Lakshmi A., H. A. Murthy, 2006, A SYLLABLE BASED *CONTINUOUS SPEECH RECOGNIZER* for TAMIL, USA, Conference: INTERSPEECH 2006 - ICSLP, Ninth International Conference on Spoken Language Processing, Pittsburgh, PA, USA, September 17-21, 2006.
- [10] M. Rohim, Suprapti, I. Baehaqie, 2013, ANALISIS KONTRASTIF BAHASA INDONESIA DAN BAHASA ARAB BERDASARKAN KALA, JUMLAH, DAN PERSONA, Indonesia, Jurnal Sastra Indonesia 2 (1) (2013).
- [11] Y. Mulyati, 2017, KOSAKATA SISWA SEKOLAH DASAR KELAS RENDAH (KAJIAN JENIS KATA, BENTUK KATA, JENIS MAKNA, DAN MEDAN MAKNA), Indonesia.
- [12] E. Pangkatodi, Liliana, G. Setiabudhi, Implementasi Rule Base System dan Fuzzy Logic Artificial Intelligence pada Game Kartu Capsa, Indonesia.
- [13] Chen H., Jakeman A. J., Norton J. P., 2008, Artificial Intelligence techniques: An introduction to their use for modeling environmental systems, *Mathematics and Computers in Simulation* 78, p(379–400).