



Available online at : <http://bit.ly/InfoTekJar>

InfoTekJar : Jurnal Nasional Informatika dan Teknologi Jaringan

ISSN (Print) 2540-7597 | ISSN (Online) 2540-7600



Artikel Penelitian

Prediksi Jeda pada Ucapan Kalimat Bahasa Melayu Pontianak Menggunakan Informasi Parafrasa

Arif Bijaksana Putra Negara, Rudy Dwi Nyoto, Wega Marda Fitri

Universitas Tanjungpura, Jalan Prof.Dr.Hadari Nawawi, Pontianak, 78124, Indonesia

KEYWORDS

Informasi Parafrasa, Kata Kunci, Bahasa Melayu Pontianak

CORRESPONDENCE

Phone: +62 85750010371

E-mail: wegamf@gmail.com

A B S T R A K

Jeda dalam kalimat sangat penting karena jeda yang terjadi dapat memperjelas informasi dari makna suatu teks kalimat, maka dari itu penelitian ini akan melakukan prediksi jeda Bahasa Melayu Pontianak menggunakan informasi parafrasa yang bertujuan untuk menghasilkan prediksi jeda pada Bahasa Melayu Pontianak untuk mendapatkan nilai akurasi yang tepat sehingga dapat digunakan pada sistem TTS. Rumusan masalah pada penelitian ini adalah bagaimana mengambil informasi parafrasa dalam suatu teks kalimat yang digunakan untuk memprediksi jeda. Parafrasa yaitu ungkapan kembali maksud atau isi tulisan dengan menggunakan kata-kata sendiri. Informasi parafrasa diambil berdasarkan kata kunci dan tipe PoS. Hasil dai penelitian ini adalah kalimat yang telah terpenggal memberntuk frasa ucapan jeda. Pengujian dilakukan dengan menggunakan *precision*, *recall*, dan *f-measure* terhadap 500 kalimat untuk frasa singkat dan frasa panjang. Hasil dari pengujian frasa singkat tanpa PoS diperoleh nilai *precision*, *recall*, dan *f-measure* adalah 0.59, 0.58 dan 0.59 sedangkan frasa singkat menggunakan PoS bernilai 0.59, 0.58 dan 0.59 serta untuk frasa panjang tanpa PoS diperoleh nilai *precision* 0.86, *recall* 0.78, dan *f-measure* 0.82 sedangkan frasa panjang menggunakan PoS bernilai 0.84, 0.77, dan 0.80. Berdasarkan pengujian yang telah dilakukan, didapatkan hasil bahwa frasa panjang lebih tinggi dikarenakan frasa yang menghasilkan indeks jeda panjang menghasilkan jeda pasti pada setiap ucapan dibandingkan frasa singkat karena frasa yang menghasilkan indeks jeda singkat akan menghasilkan ucapan yang berbeda pada setiap orang karena jeda singkat merupakan jeda tipis atau pendek.

PENDAHULUAN

Latar Belakang

Hasil Sensus Penduduk pada tahun 2010 menunjukkan dari berbagai bahasa daerah yang terdapat di Indonesia, Persentase penggunaan Bahasa Melayu yang digunakan oleh masyarakat di Kalimantan Barat mencapai 1.615.978 juta jiwa dari total penduduk Kalimantan Barat [1].

Bahasa Melayu Pontianak merupakan dialek bahasa Melayu yang dituturkan oleh masyarakat Kota Pontianak, Kabupaten Kubu Raya dan Kabupaten Mempawah serta memiliki kesamaan dengan Bahasa Melayu Semenanjung Malaysia (Johor-Riau) [2]. Upaya pelestarian bahasa Melayu Pontianak agar tidak punah dan ditinggalkan karena pengaruh globalisasi tetap harus dilakukan. Salah satu upaya dengan penerapan teknologi *Text To Speech* (TTS).

Teknologi *Text To Speech* (TTS) adalah suatu sistem yang dapat mengkonversi teks menjadi ucapan. Kualitas ucapan pada TTS dapat dinilai dari kejelasan dan kealiamian ucapan yang dihasilkan. Indikator ucapan yang jelas dan alami adalah nada, tekanan (keras-lemah), durasi, intonasi, dan jeda (kesenyapan). Pemberian jeda dalam kalimat sangat penting karena jeda yang terjadi dapat memperjelas informasi dari makna atau maksud pada suatu teks kalimat dapat tersampaikan dengan benar. Penelitian sebelumnya pernah dilakukan oleh M. Iqbal Kamiludin (2017) tentang Prediksi Jeda Dalam Ucapan Kalimat Bahasa Melayu Pontianak Menggunakan Metode *Shallow Parsing* dan hasilnya tidak sepenuhnya mampu melakukan proses pemrediksian jeda pada Kalimat Bahasa Melayu Pontianak secara akurat.

Pemberian prediksi jeda bisa juga dilakukan dengan mengambil informasi parafrasa. Parafrasa yaitu ungkapan kembali maksud atau isi tulisan dengan menggunakan kata-kata sendiri [3]. Diharapkan dengan menggunakan informasi parafrasa dapat

menghasilkan prediksi jeda dengan lebih baik dan dengan proses yang lebih mudah.

Rumusan Masalah

Berdasarkan latar belakang yang telah diatas, maka rumusan masalah yang muncul dalam penelitian ini adalah bagaimana mengambil informasi parafrasa dalam suatu teks kalimat yang digunakan untuk memprediksi jeda.

Informasi parafrasa diambil berdasarkan kata kunci dan tipe PoS. Kata kunci adalah kata yang akan ditetapkan dan digunakan dalam prediksi jeda dengan menggunakan tipe PoS.

Tujuan Penelitian

Adapun tujuan dari penelitian yang dilakukan adalah menghasilkan prediksi jeda dalam bahasa Melayu Pontianak menggunakan informasi parafrasa.

TINJAUAN PUSTAKA

Informasi Parafrasa

Informasi parafrasa diambil berdasarkan kata kunci dan tipe PoS. Kata kunci adalah kata yang akan ditetapkan dan digunakan dalam prediksi jeda dengan menggunakan tipe PoS. Kunci digunakan untuk memasukan jeda antara kalimat yang meningkatkan kejelasan dan kealamiannya. Kata kunci adalah kata yang telah ditetapkan yang akan digunakan untuk pemberian jeda. Pemberian jeda pada kata kunci dapat terjadi sebelum atau sesudah kata yang berada pada korpus, kemudian kata tersebut yang akan dimasukan dalam tabel kata kunci.

Aturan kata kunci dilakukan untuk mendapatkan hasil dari penandaan jeda. Penandaan jeda didapatkan dari penandaan *wavesurfer*. Kriteria analisis untuk pembuatan aturan kata kunci sebagai berikut:

- Analisis dilakukan dengan menggunakan 500 kalimat Bahasa Melayu Pontianak yang sudah ditandai indeks jeda sesuai dengan kejadian jeda dari penutur.
- Analisis dilihat dari kejadian jeda di tiap frasa dalam kalimat yang sudah ditandai indeks jeda sesuai dengan kejadian jeda dari penutur.
- Analisis dilihat berdasarkan kata pertama yang terdapat kejadian jeda.
- Analisis dilakukan dengan menghitung seberapa banyak jumlah kemunculan kata setiap jeda yaitu dengan melihat jumlah kemunculan kata setelah jeda, jumlah kemunculan kata sebelum jeda, dan jumlah kemunculan kata tidak ada jeda.
- Analisis tidak menggunakan minimal berapa banyak kemunculan jeda yang terjadi.
- Hasil analisis ditentukan dari jumlah terbanyak kemunculan jeda.

Part Of Speech Tagging

Part of Speech (PoS Tagging), yang juga disebut sebagai pelabelan kelas kata, adalah suatu proses yang memberikann label kelas kata secara otomatis pada suatu kata dalam kalimat [4]. Kelas kata terdiri dari kata sifat, kata benda, kata kerja, kata

keterangan, kata depan, kata kepunyaan, kata penghubung, dll. Suatu kata bisa diklasifikasikan ke dalam kelas kata yang berbeda [5]. Berbagai set PoS telah digunakan dalam penelitian-penelitian bidang pemrosesan bahasa alami (PBA), diantaranya lewat PAN Localization Project, khusus untuk bahasa indonesia telah dikembangkan PoS khusus untuk terjemahan ke bahasa Inggris pada tahun 2009 [6] Kelas kata tersebut dibangun berdasarkan Penn Treebank PoS *tag sets* [7] terdiri dari 29 PoS tag. PoS utama bahasa indonesia adalah kata kerja (*verb*), kata sifat (*adjective*), kata tugas (*function words*), kata keterangan (*adverb*), dan kata benda (*noun*), berdasarkan 5 (lima) PoS utama tersebut dan observasi data, [8] mendefinisikan 37 tag hasil dari modifikasi tag set yang di hasilkan oleh Adriani,(2009) dan Pisceldo dkk. (2009).

Natural Language Processing

Manipulasi teks telah dikenal sebagai sebuah bidang penelitian yang penting dalam NLP. Sebuah sistem NLP yang mengolah teks dimulai dengan analisis morfologi. Teks dikonversi, dalam kueri atau dokumen, untuk mendapatkan varian morfologi kata-kata yang terlibat. Pengolahan leksikal dan sintaktis melibatkan pemanfaatan kamus untuk menentukan karakteristik dari kata-kata, pengenalan *part of speech*, menentukan kata-kata dan frasa, serta untuk penguraian kalimat [9].

WaveSurfer

WaveSurfer adalah sebuah aplikasi *Open Source* untuk visualisasi dan manipulasi suara. Aplikasi ini telah dirancang untuk memenuhi kebutuhan pemula dan pengguna tingkat lanjut. WaveSurfer memiliki antarmuka yang sederhana dan menyediakan fungsi dalam cara yang intuitif dan yang dapat disesuaikan dengan berbagai tugas. WaveSurfer dapat digunakan sebagai alat yang berdiri sendiri untuk berbagai macam tugas penelitian ucapan dan pendidikan. Jenis pemakaian yang umum digunakan adalah analisis ucapan/suara dan anotasi/transkripsi suara. [10].

METODE PENELITIAN



Gambar 1. Diagram Alir Penelitian

Persiapan Data

Semua data yang diperlukan berupa korpus bahasa Melayu Pontianak yang diambil dari buku Sepok 1 karya Pay Jarrot Sujarwo. Tidak semua teks dari buku tersebut digunakan. Beberapa hal yang perlu disaring dalam teks cerita tersebut diantaranya tidak menyertakan kalimat yang terdapat kata atau istilah selain bahasa Melayu Pontianak. Korpus Bahasa Melayu Pontianak yang diambil sebanyak 500 kalimat.

Penandaan Kejadian Jeda di WafeSurver

File suara yang yang digunakan merupakan rekaman dari teks yang ada pada buku Sepok Satu karangan Pay Jarot Sujarwo dan dilakukan oleh penutur ahli. *File* suara ini berformat .wav dengan resolusi 16 bit. *File* suara tersebut kemudian dilakukan penandaan jeda kata per kata menggunakan aplikasi WafeSurver dengan mendengarkan jeda yang terjadi pada ucapan penutur.

Pengkategorian Indeks Jeda dan Penandaan Jeda di Teks Kalimat Bahasa Melayu Pontianak

Setelah semua file suara ditandai, selanjutnya setiap kejadian jeda dikategorikan jeda menggunakan indeks jeda. Penelitian ini menggunakan 5 kategori indeks jeda, yaitu “0” untuk indeks yang menyatakan tidak ada jeda, “1” untuk indeks yang menyatakan terjadinya jeda singkat, “2” untuk indeks yang menyatakan terjadinya jeda panjang, “,” untuk indeks yang menyatakan tanda baca koma, dan “.” untuk indeks yang menyatakan akhir kalimat.

Pengembangan Set PoS Bahasa Melayu Pontianak

Pada tahap ini digunakan set PoS yang baru yang dibentuk berdasarkan Set PoS Bahasa Melayu Pontianak. Pengelompokan kelas kata dalam Bahasa Melayu dilakukan secara manual dengan merujuk pada referensi set PoS pada penelitian M.Iqbal Kamiludin (2017), dan beberapa referensi set PoS yang ada yakni dari Set Pos Setyaningsih (2017) dan Set PoS Adriani (2009). *Set* PoS baru dibuat dengan tujuan untuk memperjelas penandaan tipe kata dalam kalimat dan meningkatkan akurasi prediksi jeda.

Tagging PoS Bahasa Melayu Pontianak

Pada dasarnya, IPOSTagger menggunakan korpus Bahasa Indonesia. Maka dari itu, korpus yang ada harus diganti dengan korpus Bahasa Melayu Pontianak. Teks kalimat untuk korpus latihan IPOSTagger yang berasal dari buku Sepok Satu diberikan *tag* PoS sesuai dengan kelas katanya secara manual sesuai set PoS yang telah dikembangkan. Penandaan PoS berdasarkan makna dari kata tersebut ketika diucapkan dan dari Kamus Besar Bahasa Indonesia (KBBI). Korpus latihan tersebut berisikan “kata/PoS”.

Analisis dan Menentukan Aturan Kata Kunci

Setelah melakukan *Tagging* PoS pada kalimat Bahasa Melayu Pontianak, selanjutnya kalimat Bahasa Melayu Pontianak dianalisis untuk mendapatkan *aturan* kata kunci yang diperlukan dalam melakukan proses prediksi jeda.

Aturan kata kunci dilakukan untuk mendapatkan hasil dari penandaan jeda. Penandaan jeda didapatkan dari penandaan *wavesurfer*.

Pemenggalan Frasa Jeda Ucapan Menggunakan Aturan Kata Kunci

Tahap yang selanjutnya dilakukan adalah melakukan pemenggalan frasa jeda menggunakan aturan kata kunci yang telah dibuat. Cara pemenggalan frasa ucapan jeda, yakni dengan memproses kalimat inputan menjadi penggalan frasa sesuai *aturan* kata kunci.



Gambar 2. Diagram Alir Pemenggalan Jeda dengan Aturan Kata Kunci

Pengujian

Pengujian dilakukan dengan jumlah 500 kalimat dengan aturan kata kunci yang dibandingkan dengan kalimat jeda dari penutur. Pengujian akurasi jeda dilakukan agar didapatkan nilai acuan pada *aturan* kata kunci dalam menghasilkan potongan jeda. Pengujian dilakukan untuk mendapatkan nilai pengujian yang akan dianalisis menjadi kesimpulan. Pengujian dilakukan dengan membandingkan kalimat asli dengan kalimat hasil prediksi sistem menggunakan *Precision*, *Recall* dan *F-Score*. Total kalimat yang akan diuji adalah 500 kalimat korpus. Korpus ini yang akan dibandingkan 500 korpus hasil olahan dari *aturan* kata kunci.

Analisa Pengujian

Pada tahap ini, hasil pengujian sistem akan di analisis secara keseluruhan untuk mempermudah penarikan kesimpulan.

Penarikan Kesimpulan

Kesimpulan dirumuskan berdasarkan tahapan-tahapan yang telah dilakukan sebelumnya apakah sistem yang dirancang dan dibangun dapat berjalan baik.

PEMBAHASAN DAN HASIL

Hasil Pengembangan Set PoS

Hasil dari pengembangan PoS bahasa Melayu Pontianak yaitu tabel PoS dengan jumlah sebanyak 46 set PoS. Set PoS bahasa Melayu Pontianak. Tabel Pengembangan PoS Bahasa Melayu Pontianak. Hasil pengembangan set PoS dapat dilihat pada tabel 1 dan 2.

Tabel 1. Set PoS Bahasa Melayu Pontianak

No.	PoS	Deskripsi	Contoh Kata
1	VBT	Verba Transitif (Transitive Verb)	Makai, nenggek, njajah
2	VBI	Verba Intransitif (Intransitive Verb)	Banyak, balek, turon
3	IN	Preposition (Kata Depan)	Di, ke, dari, pade
4	UH	Interjection (Kata Seru)	Oi, woi, alamak
5	RP	Partikel	Pon, lah, jak
6	JJ	Adjective	Kaye, lawar, pandai, budoh
7	OP	Open Parenthesis	{ { [
8	CP	Close Parenthesis) }]
9	.	Sentence Terminator	. ! ? ...
10	,	Comma	,
11	SYM	Symbol	# % ^ @
12	MD	Modal	nak, haros
13	NEG	Negation	Bukan, jangan, tadak
14	QT	Quotation	" "
15	WP	WH-Pronoun	Ape, siape, berape
16	DT	Determiner	ini, ni, tu, tuh
17	FW	Foreign Word	Wonderfull, story
18	CDP	Primary Numeral	Satu, duak, tige
19	CDO	Ordinal Numeral	Kesatu, keduak, ketige
20	CDI	Irregular Numeral	Beberape, segale, semue
21	CDC	Collective Numeral	Ratusan, ribuan, pulohan
22	RB	Adverb	Paleng, sementare
23	PRP	Personal Pronoun	Aku, saye, kau, die
24	PRL	Locative Pronoun	Sanak, sinik, situk
25	PRN	Number Pronoun	Satu-satunye, dua-duanye, semue-mue-e
26	NNP	Proper Noun	Eropa, Indonesia, Pontianak
27	NNG	Genitive Common Noun	Bukunye, rumahnye
28	NN	Common Noun	Martabat, Janji

Tabel 2. Set PoS Baru

No.	PoS	Deskripsi	Contoh Kata
1.	AR	Artikulus	Sang, Si
2.	CON	Konjungsi	Dan, kalok
3.	CR	Currency	Rp, \$
4.	WDT	WH-Determiner	Ape, Siape, Barangsiape
5.	US	Unit Simbol	Gr, Kg, Cm

No.	PoS	Deskripsi	Contoh Kata
6.	PDF	Fraction Numeral	Setengah, Seperempat
7.	CDA	Kata Bantu Bilangan	Biji, Ekor, Orang, Buah
8.	WPRB	WH-Adverb	Cemane, Ngape
9.	:	Colon	::
10.	NNC	Countable Common Noun	Buku, Rumah, Karyawan
11.	NNU	Uncountable Common Noun	Aek, Gule, Nasi, Ujan
12.	VBR	Verba Reduplikasi	Jalan-jalan, Potong-potong
13.	VBK	Verba Berkonjungsi	Bersalam-salam, Berputar-putar
14.	SL	Slash	/
15.	DS	Dash	-
16.	TRB	Adverb Of Time	Udah, Belom, Dudok, Sekarang
17.	FRB	Adverb Of Frequency	Jarang, Sering, Kadang-kadang
18.	DRB	Adverb Of Degree	Agak, Hamper, Cukop

Hasil Aturan Kata Kunci

Kalimat Bahasa Melayu Pontianak yang sudah ditandai jeda dianalisis kejadian jeda sesuai jeda penutup pada korpus jeda yang terjadi pada setiap penggalan frasa sehingga dihasilkan aturan kata kunci. Aturan kata kunci yang dihasilkan digunakan untuk proses pemenggalan frasa untuk mendapatkan frasa jeda. Berdasarkan analisis yang telah dilakukan terdapat 122 aturan kata kunci dari 217 analisis kata kunci tanpa PoS dan 126 aturan kata kunci dari 223 analisis kata kunci menggunakan PoS melalui 500 kalimat. Tabel aturan kata kunci Bahasa Melayu Pontianak seperti pada tabel 3, 4, 5, dan 6.

Tabel 3. Anaisis Kemunculan Jeda Aturan Kata Kunci

No.	Kata Kunci	a	b	c	Jeda 1	Jeda 2
1.	kalok	22		18	17	5
2.	gak		19	16	19	
3.	gak-e		1		1	
4.	aher-e		1			1
5.	tak	17		39	14	3
6.	dah	3	2	17	5	
7.	dahlah		1			1
8.	itu	1	8	6	6	3
9.	tu	11	48	25	41	8
10.	tang	63		26	61	2

Keterangan :

a = Jumlah Kemunculan Kata Setelah Jeda

b = Jumlah Kemunculan Kata Sebelum Jeda

c = Jumlah Kemunculan Kata Tidak Ada Jeda

Tabel 4. Aturan Kata Kunci

No.	Kata Kunci	Kemunculan Kata Sesudah Jeda	Kemunculan Kata Sebelum Jeda
1.	kalok	/1	
2.	gak		/1
3.	gak-e		/1
4.	aher-e		/2
5.	dahlah		/2
6.	itu		/1
7.	tuh		/1
8.	tu		/1
8.	tang	/1	
9.	ni		/1
10.	jak		/1

Tabel 5. Analisis Kemunculan Jeda pada Aturan Kata Kunci Menggunakan PoS

No.	Kata Kunci	a	b	c	POS
1.	tak	17		39	
2.	bukan	7		6	NEG
3.	tadak	8		7	
4.	ndak		6	3	
5.	dah	3	2	17	
6.	sekarang		2	2	
7.	dulok	1	2	1	TRB
8.	dolok		1		
9.	udah	2	2	13	
10.	selamak		1	2	
11.	dahlah		1		
12.	woi		1		UH
13.	eh		1		

Tabel 6. Atruran Kata Kunci Menggunakan PoS

No.	Kata Kunci	Kemunculan Kata Setelah Jeda	Kemunculan Kata Sebelum Jeda	PoS
1.	bukan	/1		
2.	tadak	/1		NEG
3.	ndak		/1	
4.	dulok		/1	TRB
5.	dolok		/1	
6.	dahlah		/2	
7.	woi		/1	
8.	Eh		/1	UH
9.	yarab bih		/1	
10.	e-eh		/2	

Implementasi Prediksi Jeda Kata Kunci

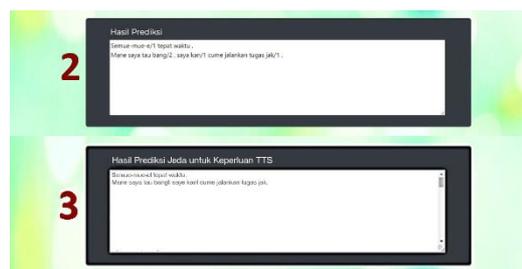
Prediksi jeda ditampilkan dalam halaman *web* untuk memudahkan pengguna melakukan prediksi jeda. Antarmuka dari prediksi jeda dapat dilihat pada gambar 3.



Gambar 3. Antarmuka Sistem Prediksi Jeda Kata Kunci



Gambar 4. Tampilan Proses Aturan Kata Kunci



Gambar 5. Tampilan Hasil Akhir Prediksi Jeda

Implementasi Prediksi Jeda Kata Kunci PoS

Prediksi jeda ditampilkan dalam halaman *web* untuk memudahkan pengguna melakukan prediksi jeda. Antarmuka dari prediksi jeda dapat dilihat pada gambar 6,7,8, dan 9.



Gambar 6. Antarmuka Sistem Prediksi Jeda Kata Kunci PoS



Gambar. 7 Tampilan Proses Tagging PoS



Gambar 8. Tampilan Hasil Kata Kunci



Gambar 9. Tampilan Hasil Prediksi Jeda

Hasil dan Analisis Pengujian

Hasil Pengujian Kecocokan Frasa Jeda dalam Satu Kalimat Penuh

Pengujian kemunculan jeda dilakukan untuk melihat persamaan kemunculan kejadian jeda pada kalimat asli korpus yang sudah ditandai kejadian jeda sesuai ucapan penutur dan kalimat hasil prediksi dari aturan kata kunci. Total kalimat yang diuji sebanyak 500 kalimat dari kalimat penutur dan 500 kalimat hasil aturan kata kunci.

Tabel 7. Perbandingan Jeda Penutur dan Jeda Aturan Kata Kunci

Tipe Pengujian	Jumlah Kalimat	Akurasi (%)
Kemunculan jeda singkat dan panjang	500	41.6%
Kemunculan jeda panjang	500	82.2%

Tabel 8. Perbandingan Jeda Penutur dan Jeda Aturan Kata Kunci PoS

Tipe Pengujian	Jumlah Kalimat	Akurasi (%)
Kemunculan jeda singkat dan panjang	500	54%
Kemunculan jeda panjang	500	80.2%

Tabel 9. Hasil Prediksi Jeda *Shallow Parsing* oleh (Kamiludin,2017)

Tipe Pengujian	Jumlah Kalimat	Akurasi (%)
Kemunculan jeda singkat dan panjang	500	10.6%

Hasil Analisa Pengujian Kecocokan Frasa Jeda dalam Satu Kalimat Penuh

Hasil pengujian menyatakan bahwa nilai akurasi terbaik pada tipe pengujian kemunculan jeda panjang. Tingginya nilai akurasi kemunculan jeda panjang dikarenakan frasa jeda panjang menghasilkan jeda pasti atau memiliki frasa yang sama pada penutur dari 500 kalimat. Sedangkan nilai pengujian kemunculan jeda singkat dan panjang menghasilkan nilai akurasi yang rendah, karena frasa jeda singkat menghasilkan pola ucapan berbeda pada setiap orang atau penutur memiliki jeda singkat yang bervariasi dari 500 kalimat sehingga aturan kata kunci dibuat belum dapat menangkap jeda singkat dengan baik. Perbandingan hasil pengujian menggunakan *shallow parsing* menyatakan bahwa nilai akurasi sebesar 10.6% pada pengujian jeda singkat dan jeda panjang. Pengujian jeda panjang tidak dilakukan karena pada penelitian sebelumnya, tidak dilakukan prediksi untuk jeda panjang, sehingga semua jeda dikategorikan sama. Rendahnya nilai akurasi dikarenakan *rule grammar* yang dibuat berdasarkan pola struktur kalimat S-P-O-K. Hal ini yang menyebabkan *rule grammar* belum bisa membentuk frasa jeda yang sama dengan frasa jeda penutur.

Hasil Pengujian Precision, Recall, dan F-Measure

Pengujian *recall*, *precision*, dan *F-Measure* dilakukan dengan membandingkan korpus frasa jeda penutur dan korpus hasil prediksi sistem. Pengujian dilakukan sebanyak 2 kali yakni pada frasa jeda singkat (1) dan frasa jeda panjang(2). Pada pengujian terhadap 5 tahap yaitu membandingkan korpus uji sebanyak 100 kalimat, 200 kalimat, 400 kalimat, dan 500 kalimat.

Tabel 10. Hasil Rangkuman Nilai Potongan Frasa Kata Kunci Terhadap Jeda Singkat dan Panjang

No.	Kalimat Korpus Uji	Precision $(\frac{a}{a+c})$	Recall $(\frac{a}{a+b})$	F-Measure $(2. \frac{p \cdot r}{p+r})$
1.	100	0.6068	0.6135	0.6100
2.	200	0.5429	0.6097	0.6066
3.	300	0.6026	0.6083	0.6054
4.	400	0.6026	0.6082	0.6053
5.	500	0.6026	0.6081	0.6053

Tabel 11. Hasil Rangkuman Nilai Potongan Frasa Kata Kunci Menggunakan PoS Terhadap Jeda Singkat dan Panjang

No.	Kalimat Korpus Uji	Precision $(\frac{a}{a+c})$	Recall $(\frac{a}{a+b})$	F-Measure $(2. \frac{p \cdot r}{p+r})$
1.	100	0.6029	0.6140	0.6082
2.	200	0.5998	0.5492	0.6049
3.	300	0.5811	0.5949	0.5895

No.	Kalimat Korpus Uji	Precision $(\frac{a}{a+c})$	Recall $(\frac{a}{a+b})$	F-Measure $(2 \cdot \frac{p \cdot r}{p+r})$
4.	400	0.5683	0.5732	0.5709
5.	500	0.5987	0.6088	0.6037

Analisa Hasil Pengujian Frasa Kata Kunci Terhadap Jeda Singkat dan Panjang

Nilai pada pengujian yang berada dikisaran 0.60 disebabkan jeda singkat memiliki jeda yang tak pasti atau selalu berubah pada setiap pengucapan sehingga banyak kalimat yang tidak akurat. Frasa jeda yang salah ini dikarenakan aturan kata kunci membentuk frasa sesuai banyaknya kemunculan kata yang memiliki jeda yang berbeda pada penutur. Jeda singkat menghasilkan pola ucapan berbeda pada setiap orang atau penutur atau memiliki jeda singkat yang bervariasi dari 500 kalimat sehingga aturan kata kunci dibuat belum dapat menangkap jeda singkat dengan baik.

Analisa Hasil Pengujian Frasa Kata Kunci Menggunakan PoS Terhadap Jeda Singkat dan Panjang

Nilai pada pengujian yang berada dikisaran 0.59 disebabkan jeda singkat memiliki jeda yang tak pasti atau selalu berubah pada setiap pengucapan sehingga banyak kalimat yang tidak akurat. Frasa jeda yang salah ini dikarenakan aturan kata kunci membentuk frasa sesuai banyaknya kemunculan kata yang memiliki jeda yang berbeda pada penutur. Jeda singkat menghasilkan pola ucapan berbeda pada setiap orang atau penutur atau memiliki jeda singkat yang bervariasi dari 500 kalimat sehingga aturan kata kunci dibuat belum dapat menangkap jeda singkat dengan baik.

Hasil Pengujian Penggalan Frasa Kata Kunci Terhadap Jeda Panjang

Tabel 12. Hasil Rangkuman Nilai Potongan Frasa Kata Kunci Terhadap Jeda Panjang

No.	Kalimat Korpus Uji	Precision $(\frac{a}{a+c})$	Recall $(\frac{a}{a+b})$	F-Measure $(2 \cdot \frac{p \cdot r}{p+r})$
1.	100	0.864	0.782	0.821
2.	200	0.864	0.781	0.820
3.	300	0.864	0.781	0.820
4.	400	0.864	0.780	0.819
5.	500	0.864	0.781	0.820

Tabel 13. Hasil Rangkuman Nilai Potongan Frasa Kata Kunci Menggunakan PoS Terhadap Jeda Panjang

No.	Kalimat Korpus Uji	Precision $(\frac{a}{a+c})$	Recall $(\frac{a}{a+b})$	F-Measure $(2 \cdot \frac{p \cdot r}{p+r})$
1.	100	0.8513	0.7786	0.8136
2.	200	0.8480	0.7749	0.8090
3.	300	0.8217	0.7498	0.7836
4.	400	0.8467	0.7734	0.8084
5.	500	0.8467	0.7733	0.8083

Analisa Hasil Pengujian Frasa Kata Kunci Terhadap Jeda Panjang

Nilai pengujian *precision* menunjukkan angka yang sama di nilai 0.86 yang berarti aturan kata kunci berhasil memprediksi penggalan frasa yang tepat untuk semua kalimat baik dari 100 kalimat sampai 500 kalimat. Nilai *recall* lebih rendah, karena masih ada penggalan frasa yang tidak sesuai dengan frasa jeda penutur karena aturan yang belum cocok. Nilai *f-measure* memiliki nilai yang hampir sama dan tergolong baik yakni 0.82. Prediksi jeda panjang memiliki nilai yang lebih baik dan tinggi, karena berdasarkan jeda penutur, letak jeda panjang cenderung memiliki pola berhenti di suatu frasa yang sama, sehingga aturan kata kunci dan aturan jeda yang dibuat dapat memprediksi jeda dengan baik.

Analisa Hasil Pengujian Frasa Kata Kunci Menggunakan PoS Terhadap Panjang

Nilai pengujian *precision* menunjukkan angka yang sama di nilai 0.84 yang berarti aturan kata kunci berhasil memprediksi penggalan frasa yang tepat untuk semua kalimat baik dari 100 kalimat sampai 500 kalimat. Nilai *recall* lebih bervariasi, karena masih ada penggalan frasa yang tidak sesuai dengan frasa jeda penutur karena aturan yang belum cocok. Nilai *f-measure* memiliki nilai yang hampir sama dan tergolong baik yakni 0.80. Prediksi jeda panjang memiliki nilai yang lebih baik dan tinggi, karena berdasarkan jeda penutur, letak jeda panjang cenderung memiliki pola berhenti di suatu frasa yang sama, sehingga aturan kata kunci yang dibuat dapat memprediksi jeda dengan baik.

KESIMPULAN DAN SARAN

Kesimpulan

Berdasarkan hasil analisis dan pengujian terhadap prediksi jeda ucapan kalimat Bahasa Melayu Pontianak dengan menggunakan pemenggalan *lookup table*, maka dapat ditarik kesimpulan sebagai berikut:

1. Berdasarkan analisis yang telah dilakukan untuk aturan kata kunci terdapat 217 kata yang dianalisis dan 122 yang masuk dalam aturan kata kunci tanpa PoS, sedangkan aturan kata kunci menggunakan PoS terdapat 224 kata yang dianalisis dan 126 yang masuk dalam aturan kata kunci.
2. Berdasarkan pengujian *precision*, *recall*, dan *f-measure*, belum sepenuhnya akurat. Pada pengujian pemenggalan terhadap frasa singkat tanpa PoS nilai *precision* 0.58, *recall* 0.59, dan *f-measure* 0.60, sedangkan pengujian frasa terhadap frasa singkat menggunakan PoS nilai *precision* 0.63, *recall* 0.61 dan *f-measure* 0.59. Pada pengujian pemenggalan terhadap frasa panjang tanpa PoS nilai *precision* 0.86, *recall* 0.78, dan *f-measure* 0.82, sedangkan pengujian frasa terhadap frasa panjang menggunakan PoS nilai *precision* 0.84, *recall* 0.77 dan *f-measure* 0.80. Frasa panjang cenderung lebih tinggi karena frasa panjang merupakan jeda pasti atau yang memiliki jeda yang sama dengan penutur.
3. Berdasarkan pengujian kecocokan frasa pada satu kalimat, jeda panjang tanpa menggunakan PoS dari 500 kalimat yang diuji menghasilkan nilai 82.2% frasa jeda yang sesuai,

sedangkan jeda panjang dengan menggunakan PoS menghasilkan nilai 80% .Kemudian, untuk jeda 1+2 tanpa menggunakan PoS dari 500 kalimat yang diuji menghasilkan nilai 41.6% frasa jeda yang sesuai, sedangkan jeda 1+2 dengan menggunakan PoS menghasilkan nilai 54%.

4. Berdasarkan penelitian yang telah dilakukan, penandaan jeda menggunakan kata kunci tanpa PoS dan menggunakan PoS mendapatkan hasil bahwa penandaan jeda kata kunci menggunakan PoS cenderung lebih baik dikarenakan penggunaan kata kunci menggunakan PoS melakukan pemneggalan frasa lebih signifikan dibanding tanpa PoS. Perbedaan nya adalah ada satu kata yang memiliki 2 PoS yang berbeda contoh kalimatnya:
 - a. Name/NNC aku/PRP bekincah/VBI macam/RP **orang**/NNC kenak/VBT masak-ek/VBT antu/NNC.
 - b. Padahal/RP yang/CON betandang/VBI tu/DT cume/DRB sebelas/CDP **orang**/CDA.
 Penggunaan kata “orang” disini berbeda karena menyesuaikan kalimat dan kata sebelum dan sesudahnya. Pada kalimat a kata orang diartikan sebagai seseorang yang mengalami kejadian tersebut, sedangkan b diartikan sebagai sekumpulan orang yang mengalami kejadian tersebut. Oleh karena itu, ada perbedaan hasil antara penggunaan PoS dan tidak.
5. Nilai *precision*, *recall*, dan *F-measure* belum mampu menghasilkan nilai akurasi yang baik dikarenakan aturan kata kunci yang dirancang masih belum sempurna untuk menghasilkan potongan frasa jeda. Hal ini disebabkan penentuan aturan kata kunci yang belum sempurna pada kalimat sehingga belum bisa memotong frasa jeda secara akurat sesuai frasa jeda penutur.

Saran

Adapun hal yang menjadi saran dalam pengembangan sistem ini agar lebih baik lagi yaitu perlu pengembangan sistem *text to speech* bahasa melayu pontianak dengan memanfaatkan informasi berupa jeda yang dihasilkan oleh aturan kata kunci dalam penelitian ini agar suara yang dihasilkan oleh sistem TTS lebih mudah dipahami.

DAFTAR PUSTAKA

- [1] Badan Pusat Statistik 2011. Kewarganegaraan, Suku Bangsa, Agama, dan Bahasa Sehari-hari Penduduk Indonesia Hasil Sensus 2010.
- [2] Putra, Muhammad Dwi Etsa.2018. Pengaruh Metode Dictionary Lookup Pada Proses Cleaning Korpus Terhadap Akurasi Mesin Penerjemah Statistik Bahasa Indonesia-Bahasa Melayu Pontianak. Pontianak: Universitas Tanjungpura
- [3] Sabarti Akhadiah, Maidar G. Arsjad, dan Sakura H. Ridwan. 1996. Pembinaan Kemampuan Menulis Bahasa Indonesia. Jakarta: Erlangga.
- [4] Jurafsky, Daniel, dan James H. Martin. 2009. Speech and Language Processing: An Introduction to Natural Language Processing, Speech Recognition, and Computational Linguistics. Edisi kedua. Prentice-Hall
- [5] Kamaludin, Muhammad Iqbal.2014. Prediksi Jeda Dalam Ucapan Kalimat Bahasa Melayu Pontianak Menggunakan Metode Shallow Parsing. Pontianak: Universitas Tanjungpura.

- [6] Adriani M, Riza H. 2008. Research Report on Local Language Computing: Development of Indonesian Language Resources and Translation System. PAN Localization.
- [7] Marcus M.P, Marcinkiewicz M.A, Santoroni B.1993. Building a Large Annotated Corpus of English: The Penn Treebank. Computational Linguistics – Special issue on using large corpora: II.
- [8] Pisceldo F, Adriani M, Manurung R. 2009. Probabilistic Part of Speech Tagging for Bahasa Indonesia. Third International Workshop on Malay and Indonesian Language Engineering. Singapore
- [9] Chowdhury, G. 2003 Natural language processing. Annual Review of Information Science and Technology.
- [10] Nugraha, Adhitya Teguh.2014. Prediksi Jeda Dalam Ucapan Kalimat Bahasa Indonesia Dengan Hidden Markov Model. Pontianak: Universitas Tanjungpura.