

Pendekatan Rule Handmade untuk Menentukan Klausa Bahasa Indonesia

Abd Wahab Syahroni¹⁾, Joan Santoso²⁾, Endang Setyati³⁾

^{1,2,3)}Teknologi Informasi - Sekolah Tinggi Teknik Surabaya

¹⁾Teknik Informatika - Universitas Madura

Surabaya, Indonesia

roney@unira.ac.id, joan@stts.edu, endang@stts.edu

Abstrak

Penelitian yang dilakukan dalam paper ini adalah pembentukan klausa sebuah kalimat Bahasa Indonesia, dengan memanfaatkan teknik chunking RegexpParser dari Natural Language Toolkit dan Rule Based System. Langkah pertama yang dilakukan dalam penelitian ini adalah mengubah level kata hasil part of speech menjadi 5 level frasa, yaitu: NP (Noun Phrase), VP (Verb Phrase), PP (Preposition Phrase), AP (Adjective Phrase) dan ADVP (Adverb Phrase). Mengingat klausa merupakan bagian kalimat yang minimal memiliki unsur Subjek dan Predikat, maka langkah kedua adalah merubah 5 level frasa dengan melihat fungsi dan kedudukannya terhadap kalimat menjadi Subjek (Subj), Predikat (P), Objek (O), Keterangan (K) dan Pelengkap (PEL). Penelitian ini mencapai akurasi hingga 95,3% dan error 4,7% dari 100 kalimat (820 kata) yang telah di ujicoba dengan menggunakan 101 rule. Terjadinya error ini disebabkan tidak adanya rule yang tepat untuk memproses hasil part of speech.

Kata kunci: Klausa, frase, chunking, rule based system, part of speech.

1. Pendahuluan

Klausa merupakan satuan gramatik yang terdiri dari: Subjek (Subj), Predikat (P), baik disertai Objek (O), Pelengkap (PEL) dan Keterangan (KET) ataupun tidak. Dengan kata lain, klausa ialah Subj-P-(O)-(PEL)-(KET). Tanda kurung menandakan bahwa satuan gramatik yang terletak dalam kurung boleh ada atau tidak. Penentuan satuan gramatik Subjek, Predikat, Objek, Pelengkap dan Keterangan dalam sebuah kalimat membutuhkan frasa. Frasa merupakan merupakan satuan gramatik yang terdiri dari dua kata atau lebih. Frasa merupakan satuan yang tidak melebihi batas fungsi unsur klausa, maksudnya frasa itu selalu terdapat dalam satu fungsi unsur klausa, yaitu: Subj, P, O, PEL, atau KET.

Dalam menentukan frasa dan klausa dalam sebuah kalimat Bahasa Indonesia dibutuhkan sebuah alat yang dapat memberi tag terhadap setiap kata dalam kalimat (POS tagger). Beberapa peneliti telah mengembangkan aplikasi POS tagger untuk Bahasa Indonesia, seperti: Wicaksono [1], Pisceldo [2], Triastuti [3], dan Sari [4]. Mengingat hasil akhir dari POS tagger berupa level kata sehingga untuk membuat sebuah frasa harus dilakukan penelitian lanjutan, demikian juga dengan klausa, mengingat klausa memiliki unsur subjek dan predikat, maka hasil dari frasa dapat dijadikan acuan untuk menentukan fungsi frasa tersebut apakah sebagai subjek, predikat, objek, keterangan atau pelengkap. Adapun penelitian mengenai frasa pernah dilakukan oleh Ali dalam Bahasa Inggris[5] dan Bahasa Urdu [6].

Oleh karena itu, dalam penelitian ini juga akan dilakukan pengolahan hasil dari POS tagger yang pernah dilakukan oleh Wicaksono [1] dengan menggunakan bantuan Teknik Chunking dari NLTK [7] dengan membuat *rule handmade* sehingga dapat menentukan frasa dan klausa dalam sebuah kalimat Bahasa Indonesia.

2. Metode Penelitian

Frasa adalah gabungan dua kata atau lebih yang bersifat non predikatif atau lazim juga disebut gabungan kata yang mengisi salah satu fungsi sintaksis di dalam kalimat [8, 9]. Frasa tidak memiliki predikat dalam strukturnya. Frasa merupakan satuan linguistik yang lebih besar dari kata dan lebih kecil dari klausa dan kalimat.

Klausa adalah satuan sintaksis berupa runtunan kata-kata berkonstruksi predikatif, artinya, di dalam konstruksi itu ada komponen berupa kata atau frasa yang berfungsi sebagai predikat dan yang lain berfungsi sebagai subjek, objek, dan sebagai keterangan. Klausa merupakan bagian dari kalimat. Klausa memiliki unsur subjek dan predikat. Klausa merupakan unsur kalimat yang mewajibkan adanya dua

fungsi sintaksis, yakni subjek (Subj) dan predikat (P) sedang yang lainnya yaitu objek (O), pelengkap (PEL), dan keterangan (K) tidak wajib ada.

Berikut ini diberikan 2 contoh kalimat, yang terdiri dari: (1) contoh frasa dan (2) contoh klausa.

- (1) a. Para guru
- b. Dosen itu
- c. Kedua orang itu
- d. Sedang mengajar
- e. di kelas
- (2) Para wanita cantik sedang mengikuti lomba

Kalimat (1) memiliki satuan bahasa “Para guru”, “Dosen itu”, “Kedua orang itu”, “sedang mengajar”, “di kelas” adalah frasa, karena satuan bahasa tersebut tidak membentuk hubungan subjek dan predikat. Kalimat (2) diatas merupakan klausa dari “para wanita cantik sedang mengikuti lomba”, yang terdiri dari tiga unsur fungsional, yaitu: unsur “para wanita cantik” menduduki fungsi subjek (Subj), “sedang mengikuti” menduduki fungsi predikat (P), dan “lomba” menduduki fungsi objek (O).

2.1. Penentuan Frasa

Frasa dapat ditentukan dengan mengolah output POS tagger menggunakan teknik phrase chunking dengan membuat rule handmade. Rule handmade akan membuat tagset baru dari hasil POS tagger. Tagset baru yang dibentuk adalah sebagai berikut:

Tabel 1. Tagset Baru Hasil Rule Hand Made

Tagset	Arti
NP	Noun Phrase
VP	Verb Phrase
PP	Preposition Phrase
AP	Adjective Phrase
ADVP	Adverb Phrase

Adapun alur proses penentuan frasa dapat dilihat pada flowchart yang tersedia di Gambar 1.



Gambar 1. Proses Penentuan Frasa

Pada Gambar 1, input kalimat Bahasa Indonesia akan di proses oleh POS tagger yang menghasilkan output level kata, kemudian phrase chunker akan merubah level kata menjadi level frasa. Jika input kalimat seperti contoh kalimat (2), maka POS tagger menghasilkan output sebagai berikut:

para/DT wanita/NN cantik/JJ sedang/RB mengikuti/VBT lomba/NN

Phrase chunker akan memproses output level kata menjadi level frasa dengan mencocokkan rule handmade yang telah dibuat dengan memanfaatkan teknik chunking RegexpParser dari Natural Language Toolkit[7]. Di bawah ini contoh rule handmade:

Rule NP : {<DT><NN><JJ>}

Kalimat : **para/DT wanita/NN cantik/JJ sedang/RB mengikuti/VBT lomba/NN**

Frasa : **para_wanita_cantik/NP**

Rule VP : {<RB><VBT>}

Kalimat : para/DT wanita/NN cantik/JJ **sedang/RB mengikuti/VBT** lomba/NN

Frasa : **sedang_mengikuti/VP**

Rule NP : {<NN>}

Kalimat : para/DT wanita/NN cantik/JJ sedang/RB mengikuti/VBT **lomba/NN**

Frasa : **lomba/NP**

Sehingga output level frasa dari level kata pada kalimat (2) adalah sebagai berikut:

para_wanita_cantik/NP sedang_mengikuti/VP lomba/NP

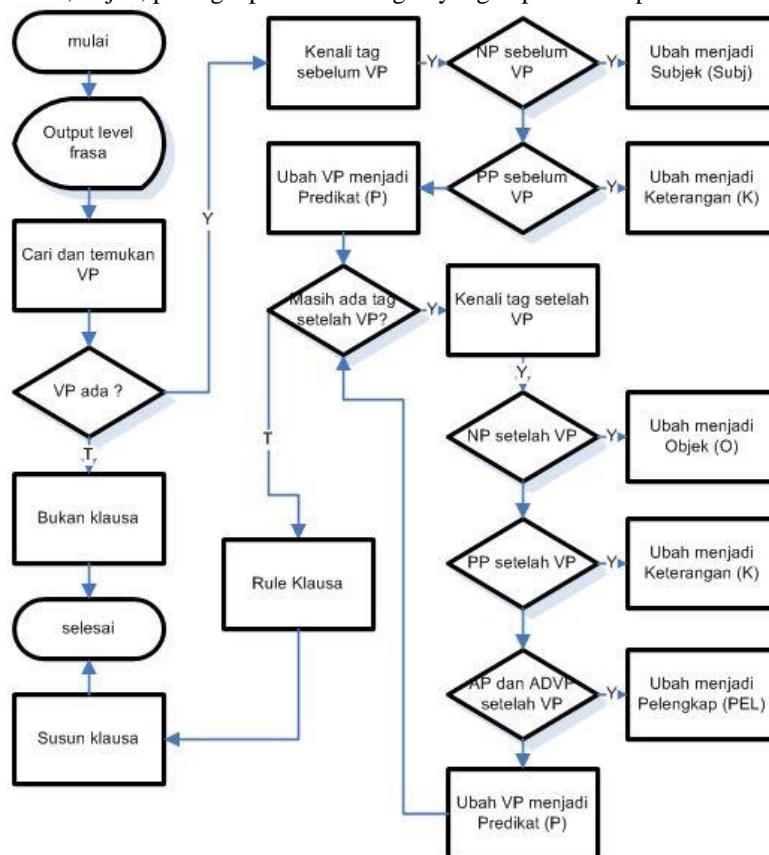
dan bentuk tree dari hasil level frasa dapat dilihat pada Gambar 2.



Gambar 2. Struktur Tree Frasa

2.2. Penentuan Klausa

Klausa ditentukan dengan mengolah output level frasa dengan cara menentukan terlebih dahulu posisi subjek, predikat, objek, pelengkap dan keterangan yang dapat dilihat pada flowchart Gambar 3.



Gambar 3. Proses Penentuan Klausa

Pada Gambar 3, langkah pertama yang dilakukan adalah mencari VP dari input level frasa, jika kalimat mengandung VP, maka kenali tag sebelum VP, ubah NP sebelum VP menjadi Subjek, jika ada PP sebelum VP, ubah PP menjadi keterangan, ubah VP menjadi predikat. Jika setelah VP masih ada tag, maka kenali tag setelah VP, ubah NP setelah VP menjadi Objek, jika ada PP ubah menjadi Keterangan, jika ada AP dan ADVP ubah menjadi Pelengkap. Jika VP sudah tidak ditemukan lagi, susun klausa berdasarkan rule klausa yang telah dibuat. Maka hasil klausa berdasarkan output level frasa di atas adalah sebagai berikut:

Input level frasa : para_wanita_cantik/NP sedang_mengikuti/VP lomba/NP
 Proses IF-THEN: **para_wanita_cantik/Subj sedang_mengikuti/P lomba/O**
 Rule KLAUSA: {(<K>+)?<Subj>(<K|PEL>+)?<P>(<O|PEL|K>+)?}
 dan bentuk tree dari hasil klausa dapat dilihat pada Gambar 4.



Gambar 4. Struktur Tree Klausa

2.3. Kategori Kalimat

Dalam penelitian ini, kategori kalimat dalam Bahasa Indonesia yang dibahas meliputi:

1. kalimat yang terdiri dari
 - Subj-P
 - Subj-P-O
 - Subj-P-PEL
 - Subj-P-K
 - Subj-P-O-PEL
 - Subj-P-O-PEL-K
 - Subj-P-O-K
 - Subj-P-PEL-K
2. Kalimat fakta/berita
3. Kalimat yang terdiri dari lebih satu kata kerja dan kata benda (kompleks).
4. Kalimat yang memiliki lebih dari satu klausa.
5. Kalimat yang telah di uji sebanyak 100 kalimat (820 kata).

Data kalimat yang digunakan bisa didapatkan dari berbagai sumber seperti buku atau berita yang menggunakan bahasa baku.

2.4. Metode Evaluasi

Adapun metode evaluasi yang digunakan adalah metode accuracy, yakni mencocokkan antara hasil dari pakar atau buku dengan hasil program mengenai susunan klausa, seperti pada Persamaan (1).

$$Rumus Accuracy = (sum(program benar/ pakar benar) / n Kalimat) * 100 \% \quad (1)$$

Tabel 2. Contoh Perhitungan Metode Accuracy

Kalimat	Pakar/Buku	Program	Program benar / Pakar benar
Roney makan	Subj-P	Subj-P	2/2 = 1
Ayah membaca koran	Subj-P-O	Subj-P-O	3/3 = 1
Budi menyanyi merdu	Subj-P-PEL	Subj-P-K	2/3 = 0.67
	Jumlah		2.67
	Banyak Kalimat		3
	Accuracy		89 %

3. Hasil dan Pembahasan

Algoritma dalam penelitian ini di implementasikan menggunakan bahasa pemrograman Python dengan bantuan Natural Language Toolkit (NLTK) [7]. Algoritma ini diujikan pada kalimat Bahasa Indonesia dengan memberi tag secara manual ke dalam 5 level frasa yaitu NP, VP, PP, AP dan ADVP. Kemudian menentukan kedudukan fungsi sintaksisnya, yaitu: subjek, predikat, objek, keterangan dan pelengkap.

Algoritma mencapai rata-rata sukses 95,3%. Algoritma rule handmade untuk frasa Bahasa Indonesia ini bergantung pada hasil identifikasi dari tagset part of speech. Jika part of speech tidak akurat mengidentifikasi input teks atau kalimat, maka artinya akan mempengaruhi efisiensi dan akurasi dari rule handmade yang telah dibuat.

Pada awalnya, rule yang sangat terbatas digunakan untuk memproses hasil part of speech menjadi bentuk frasa. Oleh karena itu, akurasi yang diperoleh menjadi rendah, namun ketika jumlah rule di tambah, maka level akurasi menjadi meningkat. Setelah rule mencapai hingga 101 rule, akurasi menjadi 95,3%. Hasil uji coba rule dapat di lihat pada Tabel 3 dibawah ini.

Tabel 3. Hasil Rule Akurasi

No	Jumlah Rule	Akurasi (%)
1	20	50
2	30	60
3	50	70
4	101	95,3

Dari Tabel 3 di atas, menunjukkan bahwa tingkat akurasi terus meningkat seiring dengan meningkatnya jumlah rule handmade yang dibuat untuk memproses hasil part of speech menjadi bentuk frasa. Algoritma ini mendapatkan 4,7% error. Error ini disebabkan tidak adanya rule yang tepat untuk memproses hasil part of speech sehingga mempengaruhi hasil selanjutnya. Daftar beberapa hasil rule handmade dapat dilihat pada Lampiran 1.

4. Simpulan

Penelitian ini menyajikan sebuah algoritma rule handmade untuk mengolah hasil part of speech tagging kalimat Bahasa Indonesia yang masih berupa level kata menjadi 5 level frasa, yaitu: NP, VP, PP, AP dan ADVP. Algoritma rule handmade untuk frasa Bahasa Indonesia ini bergantung pada hasil identifikasi dari tagset part of speech. Jika part of speech tidak akurat mengidentifikasi input teks atau kalimat, maka artinya akan mempengaruhi efisiensi dan akurasi dari rule handmade yang telah dibuat.

Langkah selanjutnya menentukan kedudukan fungsi sintaksis dari 5 level frasa terhadap kalimat menjadi subjek (Subj), predikat (P), objek (O), keterangan (K) dan pelengkap (PEL). Algoritma ini mencapai akurasi hingga 95,3%. Untuk penelitian selanjutnya dapat menggunakan sebuah proses penentuan level frasa yang lebih baik dari rule based system seperti dependency parsing.

Daftar Pustaka

- [1] A.F. Wicaksono and A. Purwarianti. HMM Based Part-of-Speech Tagger for Bahasa Indonesia. On Proceedings of 4th International MALINDO (Malay and Indonesian Language) Workshop, 2nd August 2010.
- [2] Femphy Pisceldo, Manurung, R., Adriani, Mirna. Probabilistic Part-of- Speech Tagging for Bahasa Indonesia. Third International MALINDO Workshop, Colocated Event ACL-IJCNLP 2009, Singapore, August 1, 2009.
- [3] Chandrawati, Triastuti. 2008. "Indonesian Part-of-Speech Tagger based on Conditional Random Fields and Transformation based Learning Methods". Undergraduate Thesis. Depok: Fasilkom University of Indonesia, 2008.
- [4] Sari, Syandra, Herika Hayurani, Mirna Adriani, and Stephane Bressan. Developing Part-of-Speech Tagger for Bahasa Indonesia Using Brill Tagger. The International Second MALINDO Workshop, 2008.
- [5] A. Ali and M. A. Khan, "Selecting Predicate Logic for Knowledge Representation by Comparative Study of Knowledge Representation Schemes", 2009 International Conference on Emerging Technologies.
- [6] A. Ali and M. A. Khan, " Knowledge Representation of Urdu Text Using Predicate Logic", 2010 6th International Conference on Emerging Technologies (ICET).
- [7] S. Bird, E. Klein, and E. Loper, Natural Language Processing with Python. USA: O'Reilly Media, Inc., 2009.
- [8] Ramlan. M., Ilmu Bahasa Indonesia : Sintaksis, CV. Karyono, Yogyakarta. 2005.
- [9] Supriyadi, Dr. M.Pd. Sintaksis Bahasa Indonesia. UNG Press. Gorontalo. 2014.

Lampiran 1

Tabel 4. Daftar Beberapa Hasil Rule Handmade

No	POS Tagging	Frasa	Klausa	
1	Para/DT guru/NN sedang/RB berkumpul/VBT ./.	NP=Para/DT guru/NN VP=sedang/RB berkumpul/VBT	NP VP	Subj P
2	para/DT wanita/NN cantik/JJ sedang/RB mengikuti/VBT lomba/NN ./.	NP=para/DT wanita/NN cantik/JJ VP=sedang/RB mengikuti/VBT NP=lomba/NN	NP VP NP	Subj P O
3	semua/CDI murid/NN berkumpul/VBT di/IN lapangan/NN ./.	NP=semua/CDI murid/NN VP=berkumpul/VBT PP=di/IN lapangan/NN	NP VP PP	Subj P K
4	Mereka/PRP berjalan/VBI bakmi/JJ di/IN pasar/NN ./.	NP=Mereka/PRP VP=berjalan/VBI AP=bakmi/JJ PP=di/IN pasar/NN	NP VP AP PP	Subj P PEL K
5	aku/PRP sudah/MD menghadap/VBT komandan/NNP tadi/RB ./.	NP=aku/PRP VP=sudah/MD menghadap/VBT NP=komandan/NNP ADVP=tadi/RB	NP VP NP ADVP	Subj P O PEL
6	saya/PRP memerintahkan/VBT murid/NN untuk/IN mengerjakan/VBT tugas/NN fisika/NN di/IN halaman/NN 36/CDP ./.	NP=saya/PRP VP=memerintahkan/VBT NP=murid/NN VP=untuk/IN mengerjakan/VBT NP=tugas/NN fisika/NN PP=di/IN (NP halaman/NN 36/CDP)	NP VP NP VP NP PP	Subj P O P O K
7	Petugas/NN bendera/NN mengibarkan/VBT bendera/NN merah/JJ putih/JJ ./.	NP=Petugas/NN bendera/NN VP=mengibarkan/VBT NP=bendera/NN merah/JJ putih/JJ	NP VP NP	Subj P O
8	tukang/NN kayu/NN sedang/RB memotong/VBT kayu/NN dengan/IN mesin/NN gergaji/NN ./.	NP=tukang/NN kayu/NN VP=sedang/RB memotong/VBT NP=kayu/NN PP=dengan/IN mesin/NN gergaji/NN	NP VP NP PP	Subj P O K
9	si/RP anton/NNP membantu/VBT menulis/VBT surat/NN ./.	NP= si/RP anton/NNP VP=membantu/VBT menulis/VBT NP=surat/NN	NP VP NP	Subj P O
10	guru/NN dan/CC murid/NN telah/MD meninggalkan/VBT sekolah/NN ./.	NP=guru/NN dan/CC murid/NN VP=telah/MD meninggalkan/VBT NP=sekolah/NN	NP VP NP	Subj P O
11	semua/CDI anak-anak/NN dan/CC guru/NN di/IN sekolah/NN berkumpul/VBT di/IN aula/NN ./.	NP=semua/CDI anak-anak/NN dan/CC guru/NN PP=di/IN sekolah/NN VP=berkumpul/VBT PP=di/IN aula/NN	NP PP VP PP	Subj K P K
12	roni/NN makan/VBT dan/CC minum/VBT dan/CC anton/NNP melihat/VBT televisi/NN dan/CC ubed/FW belajar/VBT komputer/NN dan/CC jaringan/NN ./.	NP=roni/NN VP=((VP makan/VBT) dan/CC (VP minum/VBT)) CC dan/CC NP anton/NNP VP melihat/VBT NP televisi/NN CC dan/CC NP=ubed/FW VP=belajar/VBT NP=komputer/NN dan/CC jaringan/NN	CLAUSE: NP VP CC CLAUSE: NP VP NP CC CLAUSE:NP VP NP	KLAUSA: Subj P CC KLAUSA: Subj P O CC KLAUSA: Subj P O