

THE IMPLEMENTATION OF CYK ALGORITHM TO SEPARATE SENTENCE-FORMING ELEMENTS IN INDONESIAN

Teguh Tri Sulianto¹, Rosita Herawati²

^{1,2}Program Studi Teknik Informatika Fakultas Ilmu Komputer, Universitas Katholik Soegijapranata

¹teguhtrisulianto@gmail.com, ²rosita@unika.ac.id

Abstract

A sentence contains some constituent elements. Those elements have its own function. The four main elements in Bahasa Indonesia are subject, predicate, object, and prepositional phrase. The structure of this sentence often mixed up while determining those elements. This research aims to help people who confused to determined what are the constituent element in a sentence. The sentence can be inputted and the program identified the sentence that inputted. It started by separating the sentence into words and put it inside the cyk matrix table. After that, the program will identify if each word has some degree of similarity with any of constituent element characteristic. The program showed what are the constituent elements of the sentence. It will be shown for the words for each element and the rules that used to determined it. So the people can figure the constietuent element.

Keywords: sentence, constituent elements, cyk.

Pendahuluan

Kalimat adalah satuan bahasa terkecil dalam wujud lisan atau tulisan yang mengungkapkan suatu pikiran yang utuh [Alwi, 1998]. Kalimat memiliki beberapa elemen penyusun atau elemen sintaksis (posisi kata atau aturan kata) yang terdiri dari subjek (S), predikat (P), objek (O), keterangan (K).

Penting untuk memahami unsur-unsur tersebut agar kalimatnya menjadi benar. Kalimat yang benar sangat penting saat penulis menulis sebuah karya ilmiah. Dalam bahasa Indonesia harus memperhatikan peraturan EYD (Ejaan Yang Disempurnakan). Namun terkadang penulis sulit membuat kalimat yang benar karena kesulitan dalam menganalisa kalimat tersebut.

Sebenarnya ada algoritma yang bisa digunakan untuk membantu penulis menyelesaikan kasus diatas, yaitu menggunakan algoritma CYK. Algoritma ini membantu menganalisa kalimat untuk menemukan elemen penyusunnya. Tapi, perlu banyak waktu dan usaha jika algoritma cyk dilakukan secara manual. akan lebih membantu jika algoritma ini diimplementasikan dengan menggunakan aplikasi. Dengan menggunakan aplikasi ini, seorang penulis dapat membuat kalimat yang benar dengan mudah dan menulis sebuah karya ilmiah yang baik.

Landasan Teori

Artikel "Analisis Perbandingan Algoritma LCP (Left-Corner-Parsing) Dan Algoritma CYK (Cocke-Younger-Kasami) Untuk Memeriksa Pola Kalimat Baku Bahasa Indonesia" yaitu membandingkan algoritma LCP dan algoritma CYK. Kedua algoritma ini dipelajari dan dibandingkan kemampuan masing-masing algoritma dalam memeriksa pola kalimat standar bahasa Indonesia. Array dua dimensi dengan jumlah baris dan kolom $n \times n$ digunakan sebagai proses parsing dengan algoritma CYK. Nilai n ditentukan oleh jumlah kata dalam sebuah kalimat. Hasil pengujian uji kalimat standar diperoleh tingkat akurasi algoritma CYK (Cocke-Younger-Kasami) sebesar 65% dengan tingkat kesalahan 0,35 dan algoritma LCP 60% dengan tingkat kesalahan 0,40.

Ada artikel lain yang berjudul "Translasi Context-Free Grammar Menjadi Parsing Tree Berbasis Algoritma Cocke-Younger-Kasami", yang juga membahas tentang CYK. Artikel ini menggunakan pohon untuk mewakili hasil proses penguraian, yang dikenal sebagai pohon parsing. "Untuk memudahkan pemahaman tentang pohon parsing, itu akan membutuhkan sebuah aplikasi, yang secara otomatis dapat menghasilkan pohon parsing dengan proses derivasi, berdasarkan string dan aturan. Aplikasi ini menggunakan aturan dalam algoritma *Context-Free Grammar*, berbasiskan pada Cocke-Younger-Kasami (CYK), dan menghasilkan gambar pohon parsing dengan proses derivasi pada derivasi paling kiri. "Berdasarkan hasil pengujian beberapa Kasus uji kompleksitas aturan dan string yang berbeda dalam artikel ini, dapat disimpulkan bahwa aplikasi CYK Parser dapat menangani berbagai peraturan dan string dari yang sederhana hingga yang kompleks. Ketika proses parsing gagal, aplikasi CYK Parser dapat menghasilkan kesalahan pohon parsing, sehingga pengguna dapat mengetahui kesalahan dalam proses parsing.

Perbedaan antara artikel berjudul Analisis Algoritma LCP dan Algoritma CYK Untuk Memeriksa Pola Kalimat Baku Bahasa Indonesia dan proyek ini sesuai peraturan. Pada artikel ini, CYK memiliki beberapa peraturan produksi. Dalam proyek ini tidak ada peraturan produksi. Proses input didasarkan pada kalimat yang dimasukkan. Antara artikel bertajuk Translasi Context-Free Grammar Menjadi Parsing Tree Based Algoritma Cocke-Younger-Kasami dan proyek ini berbeda berdasarkan metode yang digunakan. Pada artikel ini gunakan pohon parsing untuk melakukan metode parsing. Proyek ini menggunakan tabel matriks untuk melakukan metode parsing.

Metodologi Penelitian

Proyek ini menggunakan pendekatan inkremental. Inilah langkah-langkah dari proyek ini:

1. Mencari dan mengumpulkan referensi tentang algoritma CYK dan struktur data array yang digunakan dalam proyek ini.

Proyek ini mendapatkan beberapa artikel berjudul "Analisis Terjemahan Algoritma LCP dan Algoritma CYK Untuk Memeriksa Pola Kalimat Baku Bahasa Indonesia" dan "Translasi Konteks-Free Grammar Menjadi Parsing Tree Based Algoritma Cocke-Younger-Kasami". Artikel tersebut membantu proyek ini untuk mendapatkan referensi tentang algoritma CYK.

2. Menganalisis masalah dan membuat rencana untuk mengerjakan proyek.

Setelah mempelajari referensi di atas, langkah analisis dimulai. Analisisnya harus mencakup beberapa masalah seperti bagaimana memisahkan kalimat menjadi kata-kata, bagaimana mengenali apakah ada subjek, predikat, objek, atau keterangan, dan bagaimana mendapatkannya kembali dan menampilkannya di bagian akhir. Untuk memisahkan kalimat, menggunakan spasi sebagai bantuannya. Untuk mengenali unsur-unsurnya, karakteristik elemen dibandingkan dari setiap kata. Indeks dari array akan disimpan setiap kali menemukan elemen.

3. Melaksanakan struktur data dan algoritma.

Langkah implementasi dimulai dengan memisahkan kalimat menjadi kata-kata. Kata-kata tersebut akan dimasukkan ke dalam tabel matriks CYK untuk dianalisis. Proses analisis adalah untuk menemukan karakteristik sebuah kata. Indeks kata yang cocok akan disimpan untuk dipanggil nanti.

Hasil dan Pembahasan

Program ini dimulai dengan menganalisis kata yang dimasukkan, lalu menentukan apakah termasuk kalimat tunggal atau kalimat majemuk. Setiap kata dimasukkan ke dalam tabel matriks CYK dengan memasukkan kata yang dimasukkan dari kata terakhir sampai kata pertama. Berikut ini adalah contoh tabel matriks n CYK jika kalimat yang dimasukkan terdiri dari 6 kata.

Setelah memasukkan ke dalam tabel, program akan membaca dari kanan bawah ke kiri atas. Unsur pertama yang dicari adalah keterangan. Keterangan ditunjukkan oleh preposisi, seperti di, ke, karena, kapan, dan lain-lain (Berbahasa Indonesia dengan Benar, Dendy Sugono). Predikat kalimat dapat berupa kata *adalah* atau *ialah* (Berbahasa Indonesia dengan Benar, Dendy Sugono). Untuk menentukan predikat juga memiliki metode lain. Metode pertama adalah mengetahui kata dasar dari sebuah kata berimbuhan dengan bantuan dari sastrawi dan membandingkannya dengan karakteristik kata berimbuhan. Dalam bahasa Indonesia, kata kerja transitif memiliki enam kemungkinan bentuk, yaitu meN-, meN-kan, meN-i, memper, memper-kan, memper-i (Bahasa Indonesia yang Salah dan Benar). Setelah predikat, biasanya ada objek dan pelengkap. Tapi mungkin saja hanya ada objek, atau hanya pelengkap, atau bahkan tidak ada keduanya. Objek dapat ditentukan jika predikat memiliki imbuhan me- atau di-, maka kata selanjutnya akan menjadi objek. Karena subjek selalu di awal, maka saat

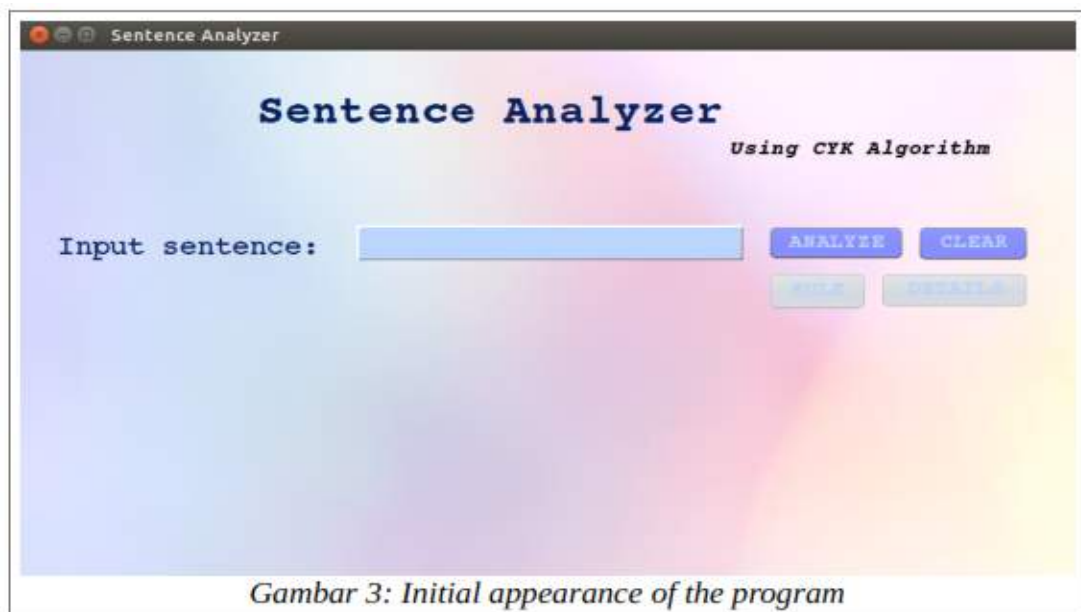
Tabel 1: CYK matrix

		TO					
		1	2	3	4	5	6
FROM	0	0-1	0-2	0-3	0-4	0-5	0-6
	1		1-2	1-3	1-4	1-5	1-6
	2			2-3	2-4	2-5	2-6
	3				3-4	3-5	3-6
	4					4-5	4-6
	5						5-6

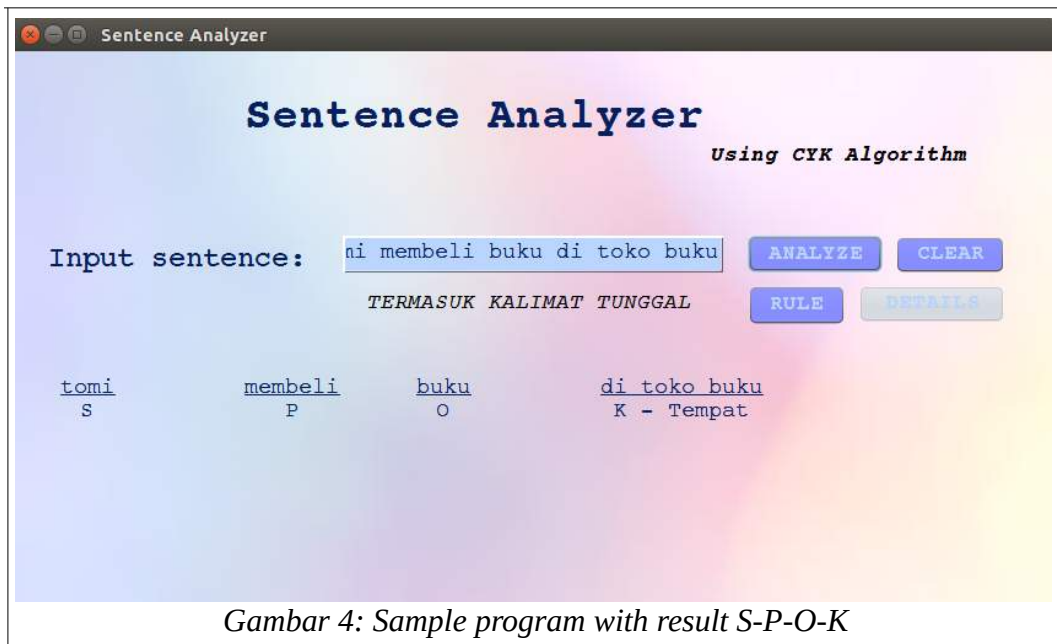
program menemukan predikat, kata pertama sampai kata sebelum predikat ditentukan sebagai subjek.

Program bisa langsung menentukan unsur jika hanya dua kata, maka akan ditetapkan sebagai keterangan jika lolos dari pengecekan keterangan. Jika tidak, kata pertama akan menjadi subjek dan yang kedua adalah predikat. Yang berikutnya adalah jika kalimat yang dimasukkan adalah 3 kata, itu memiliki beberapa kemungkinan. Yang pertama adalah ketika kata kedua adalah karakteristik predikat, jadi kata pertama adalah subjek dan sisanya adalah predikat. Kedua, adalah ketika kata kedua adalah *itu*, maka kata pertama dan kedua adalah subjek dan yang ketiga adalah predikat. Ketiga, adalah ketika kata kedua adalah karakteristik keterangan, jadi kata pertama adalah predikat dan kata kedua dan ketiga adalah keterangan. Keempat, adalah ketika bantuan dari sastra menemukan kata kedua sebagai predikat, maka yang pertama adalah subjek, yang kedua adalah predikat, dan yang ketiga ditentukan oleh karakteristik objek atau pelengkap.

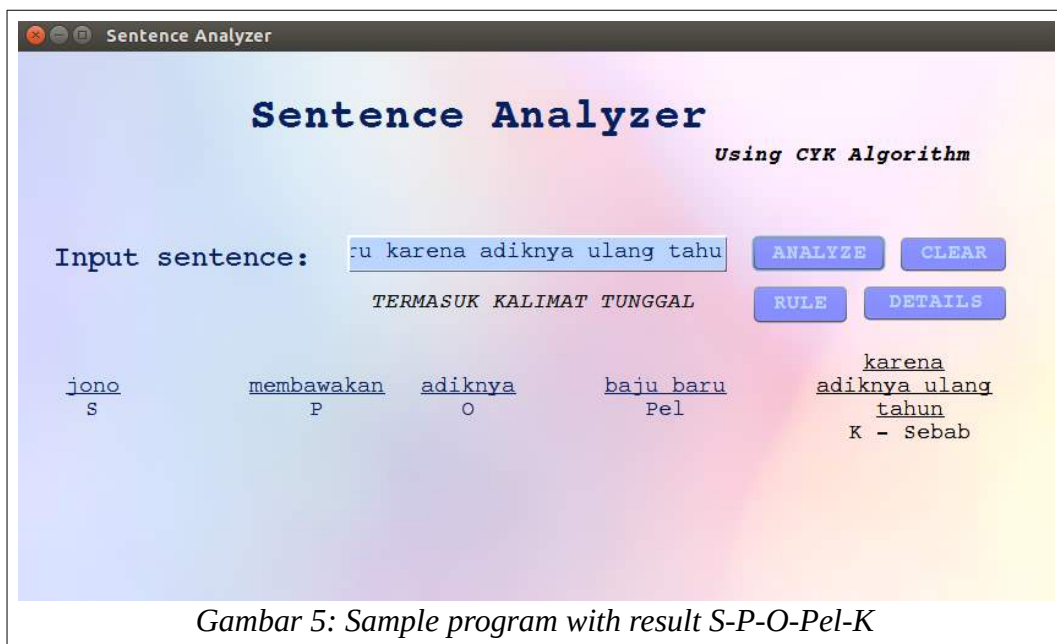
Ini adalah tampilan dan contoh dari program.



Gambar 3: Initial appearance of the program



Gambar 4: Sample program with result S-P-O-K



Gambar 5: Sample program with result S-P-O-Pel-K

Kesimpulan

Program ini dapat mengidentifikasi sebuah kalimat dengan menggunakan algoritma CYK dan menunjukkan masing-masing unsur penyusunnya. Tabel matriks CYK dapat digunakan untuk bahasa Indonesia, meskipun memiliki peraturan dalam bahasa Inggris. Program ini juga bisa menunjukkan rinciannya jika ada informasi lebih lanjut.

Daftar Pustaka

- [1] Alwi, H. et, al, “Tata Bahasa Baku Bahasa Indonesia”, Balai Pustaka, 1998
- [2] Arifin, E, Zaenal, “Dasar-Dasar Penulisan Karangan Ilmiah”, Grasindo, 2004
- [3] Ramlan , M. Dkk, “Bahasa Indonesia yang Salah dan yang Benar”, Andi Offset, 1992
- [4] Sugono, D., “Mahir Berbahasa Indonesia dengan Benar. Indonesia”, Gramedia Pustaka Utama, 2009
- [5] Anggara P. A., “Translasi Context-Free Grammar Menjadi Parsing Tree Berbasis Algoritma Cocke-Younger-Kasami”, Ilmu Komputer, Universitas Dian Nuswantoro Semarang, 2015
- [6] Susanti S., “Analisis Perbandingan Algoritma LCP (Left-Corner- Parsing) dan Algoritma CYK (Cocke-Younger-Kasami) Untuk Memeriksa Pola Kalimat Baku Bahasa Indonesia”, Ilmu Komputer, Universitas Komputer Indonesia, 2016