

SISTEM PENGURAIAN KALIMAT BERDASARKAN ATURAN SINTAKS BAHASA INDONESIA DENGAN PENDEKATAN LINGUISTIC STRING ANALYSIS

Wahyu Muhammad Iqbal¹, Dhinta Darmantoro², Dyas Puspandari Ss.mpd.³

¹Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

Abstrak

Bahasa merupakan unsur terpenting dalam berkomunikasi dengan sesama manusia. Dalam bahasa terkandung ilmu pengetahuan dari satu generasi ke generasi lainnya yang berbentuk tulisan. Oleh karena itu, munculah ide untuk membuat perangkat lunak yang dapat membantu dalam menguraikan struktur kalimat.

Pengembangan perangkat lunak pada tugas akhir ini menggunakan pendekatan Linguistic String Analysis (LSA) dengan menggunakan aturan produksi BNF dan Visual Basic .NET sebagai bahasa pemrogramannya.

Dari hasil pengujian, tingkat keberhasilan dalam menguraikan kalimat dengan benar, cukup baik. Akan tetapi masih terdapat kalimat yang tidak bisa diuraikan. Hal ini disebabkan oleh, aturan kalimat yang belum sempurna dan kamus kata yang masih kurang lengkap.

Kata Kunci : bahasa, menguraikan, BNF, Linguistic String Analysis

Abstract

Language is important element in communicating with people. In language consist of several knowledges from one generation to other generation which in form of article. Because of that, the idea is to make software that can help elaborating sentence structure.

The application is developed using Linguistic String Analysis (LSA) with BNF rule and Visual Basic .NET as the programming language.

From result of examination, mount efficacy in elaborating sentence with true condition, is good enough. However there still are sentence which cannot be elaborated. This matter because of, sentence rule are not perfect and word dictionary is still less completed.

Keywords : language, elaborating, BNF, Linguistic String Analysis

Telkom
University

0. Pendahuluan

1.1 Latar belakang

Bahasa merupakan unsur terpenting dalam berkomunikasi dengan sesama manusia. Dalam bahasa terkandung ilmu pengetahuan dari satu generasi ke generasi lainnya yang berbentuk tulisan. Bahasa juga berperan dalam mengarahkan tingkah laku manusia sehari-hari dalam hubungannya dengan sesamanya.

Motivasi dalam melaksanakan penelitian ini adalah bahwa kemampuan pemrosesan bahasa alami akan mengubah cara penggunaan komputer.[2] Kebanyakan informasi tersimpan dalam bentuk bahasa sehingga komputer yang dapat mengerti bahasa alami dapat mengakses informasi ini. Komputer yang memiliki antar muka yang kompleks dalam bahasa alami akan dapat digunakan oleh setiap orang.

Penelitian dalam bidang pemrosesan bahasa alami sudah banyak dilakukan. Akan tetapi, penelitian untuk bahasa Indonesia sendiri masih jarang dilakukan. Oleh karena itu, diperlukan suatu penelitian terhadap pemrosesan bahasa Indonesia terutama kalimat.

Kalimat adalah satuan bahasa terkecil, dalam wujud lisan dan tulisan, yang mengungkapkan pikiran yang utuh. Kalimat umumnya berwujud rentetan kata yang disusun sesuai dengan kaidah yang berlaku. Kalimat dimulai dengan huruf kapital dan diakhiri dengan tanda titik, tanda tanya, atau tanda seru; sementara itu, didalamnya disertakan pula berbagai tanda baca seperti koma, titik dua, tanda pisah, dan spasi.[5]

Pemrosesan bahasa alami salah satunya adalah pengurai (*parser*) struktur kalimat. Pengurai ini memberikan pengetahuan bagaimana setiap kata dalam kalimat berhubungan. Struktur ini juga mengenali kata-kata yang menjadi inti dari suatu kalimat, kata-kata mana yang saling mempengaruhi satu sama lain, dan bagaimana kata-kata bersatu membentuk klausa atau frasa.

Proses menguraikan kalimat menjadi sesuatu yang dapat dimengerti oleh komputer sama seperti proses menguraikan tata bahasa pemrograman dalam dunia komputer. Perbedaan yang mendasar pada keduanya ada pada konteks bahasanya. Pada komputer tata bahasa yang digunakan bebas konteks (*context free grammar*), sedangkan pada bahasa alami adalah terikat konteks (*context sensitive*). Untuk dapat menjabarkan ke dalam bahasa pemrograman dibutuhkan pemisahan analisa konteks dari suatu kalimat (semantik) dengan analisa sintaks. Walaupun demikian, pengolahan struktur kalimat berdasarkan analisa sintaks tidak sederhana yang dibayangkan.

Linguistic String Analysis (LSA) adalah suatu pendekatan yang menyatakan bahwa kalimat dibangun dari satu kalimat inti dan ditambah string sisipan pada bagian tertentu dari kalimat inti tersebut. Pendekatan ini digunakan untuk menentukan mana kalimat inti dari suatu kalimat dan mana kata-kata sisipan pada kalimat inti tersebut. Pada bahasa Indonesia, *linguistic string* merupakan rangkaian dari satu atau lebih kata.

Aturan sintaks yang dibuat menggunakan bentuk *backus naur form* (BNF) yang cocok digunakan untuk algoritma pengurai yang memiliki sifat *context*

free.[10] String tata bahasa yang didefinisikan oleh BNF adalah kelas-kelas string yang menggambarkan kategori dari *string analysis*. [10]

Dalam tugas akhir ini akan dikembangkan aplikasi penguraian kalimat berdasarkan aturan-aturan sintaks bahasa Indonesia dengan pendekatan *Linguistic String Analysis (LSA)*.

1.2 Perumusan masalah

Perumusan masalah yang dijadikan bahan penelitian pada penyusunan tugas akhir ini adalah :

1. Bagaimana mendefinisikan kamus kata yang akan digunakan
2. Bagaimana mendefinisikan aturan sintaks yang digunakan
3. Bagaimana merepresentasikan kalimat ke dalam bahasa komputer.
4. Bagaimana menentukan pendekatan yang akan digunakan.

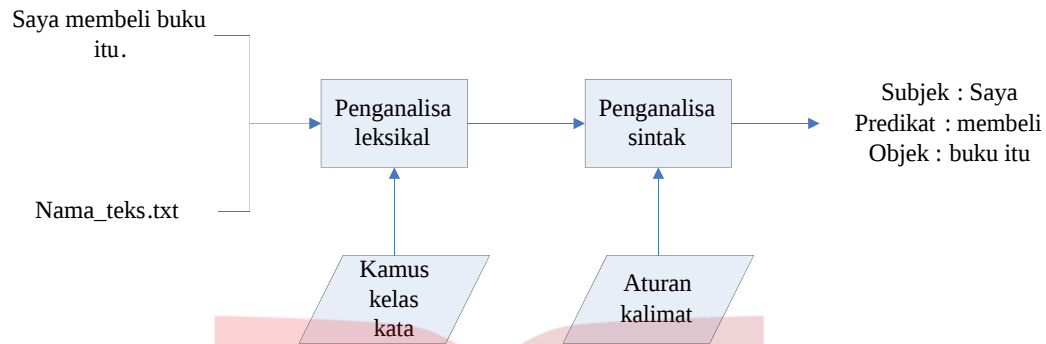
Untuk mempersempit cakupan materi penyusunan tugas akhir ini, masalah-masalah yang telah dirumuskan dibatasi pada hal-hal berikut ini :

1. Bentuk kalimat yang digunakan adalah kalimat berita (*kalimat deklaratif*).
2. Kata-kata yang diperiksa hanyalah kata-kata yang sudah didefinisikan terlebih dahulu.
3. Tidak memeriksa konteks bahasa (*semantik*).
4. *Ambiguitas* (outputan banyak jawaban) kalimat ditangani dengan memilih salah satu jawaban yang benar.
5. Kalimat yang diperiksa adalah kalimat dasar dan kalimat majemuk setara.
6. Contoh kalimat diambil dari halaman utama salah satu surat kabar di Indonesia
7. Pendekatan yang digunakan adalah *Linguistic String Analysis (LSA)*.
8. Menggunakan bahasa pemrograman Visual Basic .NET 2003 dalam pengerjaannya.
9. Tidak memeriksa keterhubungan antar kalimat.

1.3 Tujuan

Tujuan pembahasan yang ingin dicapai dari penelitian tugas akhir ini adalah :

1. Membuat aplikasi yang dapat menguraikan struktur kalimat bahasa Indonesia sesuai dengan aturan sintaks kalimat yang benar.
2. Metode *Linguistic String Analysis (LSA)* dengan menggunakan BNF sebagai aturan produksinya dapat diterapkan dalam sistem penguraian kalimat, mulai dari kalimat dasar sampai dengan kalimat majemuk setara.
3. Tingkat keberhasilan penguraian kalimat yang ingin dicapai adalah di atas 80% dari seluruh kalimat yang masuk ke dalam sistem.



Gambar 1-1: Diagram proses pengolahan kalimat

1.4 Metodologi penyelesaian masalah

Metodologi penyelesaian masalah yang digunakan pada penelitian tugas akhir ini adalah sebagai berikut :

1. Studi literatur
 - a. Konsep dan karakteristik kalimat berdasarkan aturan sintaks ketatabahasa.
 - b. Konsep dari LSA dalam melihat kalimat sebagai string inti dan string sisipan.
 - c. Implementasi aturan produksi BNF dalam penjabaran pola-pola kalimat.
 - d. Visual Basic .NET sebagai platform pengembangan aplikasi.
2. Analisis kebutuhan
 - a. Jenis-jenis kata dan kelas kata untuk dipelajari dan disesuaikan dengan sistem yang ingin dibuat.
 - b. Analisis pola-pola kalimat yang akan digunakan sesuai dengan metode LSA.
 - c. Komponen-komponen yang saling terkait pada Visual Basic .NET untuk pengembangan sistem.
 - d. Kalimat atau teks sebagai masukan aplikasi.
3. Desain sistem
 - a. Penentuan kelas-kelas kata yang digunakan sistem.
 - b. Penentuan aturan kalimat dengan menggunakan aturan produksi BNF.
 - c. Pengujian data kalimat atau teks berdasarkan dari aturan produksi yang telah dibuat.
4. Implementasi sistem dalam bentuk aplikasi yang menggunakan bahasa pemrograman Visual Basic .NET.
5. Pengujian
 - a. Pengujian dilakukan dengan memasukkan inputan berupa halaman utama surat kabar sebanyak 30 buah (dalam bentuk teks).
 - b. Perbandingan hasil outputan aplikasi dengan hasil analisis menggunakan teori bahasa Indonesia.
 - c. Evaluasi tingkat keberhasilan penguraian struktur kalimat bahasa Indonesia.
6. Kesimpulan dan penyusunan laporan penelitian tugas akhir.

0. Kesimpulan Dan Saran

1.13 Kesimpulan

1. Dari 610 kalimat, hanya 65 kalimat (10,82%) yang sesuai dengan lingkup masalah tugas akhir ini. Kebanyakan kalimat yang digunakan adalah kalimat majemuk bertingkat.
2. Dari 65 kalimat, sistem berhasil menguraikan kalimat dengan struktur kalimat benar sebanyak 56 kalimat (86,15%). Sistem berhasil menguraikan kalimat tapi struktur kalimat salah sebanyak 2 kalimat (3,08%). Sistem tidak berhasil menguraikan tapi kalimat masukan harusnya bisa diuraikan sebanyak 1 kalimat (1,54%). Sistem tidak berhasil menguraikan dan kalimat masukan salah sebanyak 6 kalimat (9,23%).
3. Pendekatan Linguistic String Analisis (LSA) dengan menggunakan aturan produksi BNF dapat menguraikan struktur kalimat berdasarkan aturan sintaks dengan tingkat keberhasilan 86,15% dari 65 kalimat.
4. Tingkat keberhasilan penguraian sesuai dengan tingkat keberhasilan yang diinginkan, yaitu di atas 80% (sebesar 86,15%).
5. Hasil implementasi penguraian yang memuat struktur kalimat tunggal dan kalimat majemuk setara bahasa Indonesia dengan menggunakan definisi BNF cukup baik. Dari 65 kalimat, *parser* berhasil menguraikan 86,15% struktur kalimat masukan dengan benar.
6. Pola dasar kalimat yang paling sering digunakan adalah pola dasar tipe 4 (SPPel) sebanyak 30,76%.
7. Dari 56 kalimat, sebanyak 13 kalimat (23,22%) memiliki lebih dari satu hasil keluaran. Hal ini diakibatkan karena ada kata-kata yang menggunakan lebih dari satu kelas kata.
8. Sistem salah menguraikan sebanyak 2 kalimat. Hal ini disebabkan sistem tidak menganggap kata-kata diantara tanda ‘ adalah frasa nomina.

1.14 Saran

1. Aturan kalimat yang digunakan sebaiknya diperbanyak untuk dapat menangani permasalahan seperti letak bilangan yang berada di belakang pada fungsi keterangan.
2. Untuk menangani hasil yang keluaran lebih dari satu dapat digunakan metode berdasarkan aturan semantik dari kelas kata sehingga keterkaitan antar kata dapat diperiksa dan menghasilkan solusi yang lebih sedikit.
3. Jumlah kata dalam kamus kata perlu ditambah lagi lebih banyak agar tidak terjadi kekurangan kamus kata.

Daftar Pustaka

- [1] **Alwi, Hasan; Dardjowidjojo, Soenjono; Lapoliwa, Hans; M. Moeliono, Anton.** *Tata Bahasa Baku Bahasa Indonesia*, Jakarta : Balai Pustaka, 2003.
- [2] **Allen, J.** *Natural Language Understanding*, The Benjamin/Cumming Publishing Company, Inc., Redwood City, CA, 1995.
- [3] **Badudu, Dr. J. S.** *Inilah Bahasa Indonesia Yang Benar*.
- [4] **Deitel, H. M; Deitel, P. J.; Nieto, T. R.** *Visual Basic .NET How To Program*, Prentice Hall, New Jersey, 2000.
- [5] **Iskak, Hendrawan.** *Pengurai Sintaks Kalimat untuk Bahasa Indonesia dengan Metode Linguistic String Analysis*, Fakultas Ilmu Komputer Universitas Indonesia, Depok, 1999.
- [6] **Laboratorium Rekayasa Perangkat Lunak STT Telkom.** *Modul Praktikum Rekayasa Perangkat Lunak edisi 3*; 2004
- [7] **M. Ramlan, Prof. Drs.** *Ilmu Bahasa Indonesia "Sintaksis"*.
- [8] **M. Ramlan, Prof. Drs.** *Tata Bahasa Indonesia : Penggolongan Kata*, Yogyakarta : Andi Offset, 1985.
- [9] **Puspandari, Dyas.** *Proses Sintaksis Dalam Bahasa Indonesia*, Fakultas Sastra, Universitas Gajah Mada, Yogyakarta 1997.
- [10] **Sager, N.** *Natural Language Information Processing: A Computer Grammar of English and Its Applications*, Addison-Wesley Publishing Company, Massachusetts, 1981.
- [11] **Sugono, D.** *Berbahasa Indonesia dengan Benar*, Penerbit Puspa Swara, Jakarta 1997.