

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 sesederhana 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.

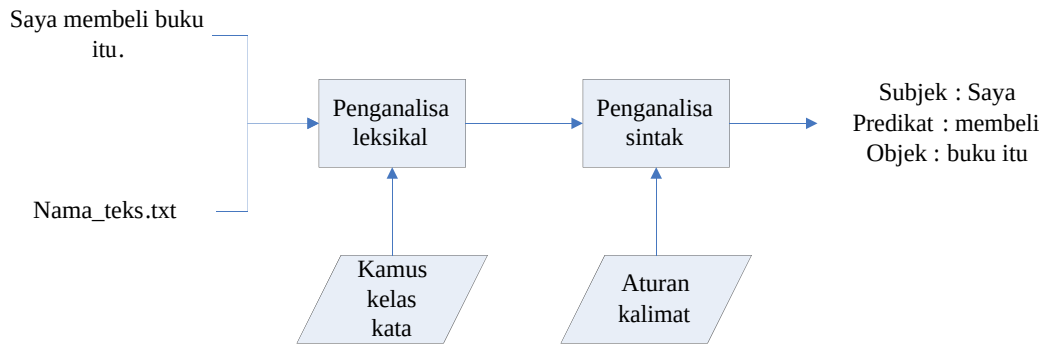
Untuk mempersempit cakupan materi penyusunan tugas akhir ini, masalah-masalah yang telah dirumuskan dibatasi pada hal-hal berikut ini :

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 ketatabahasaan.
 - 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.