

1. Pendahuluan

1.1 Latar belakang

Setiap bahasa alami memiliki kata yang dapat bermakna lebih dari satu, sesuai dengan konteks kalimat yang menyertainya. Kata bermakna lebih dari satu tersebut, dapat berpotensi menyebabkan keragu-raguan atau ambigu. Usaha untuk memilih makna dari kata tersebut berdasarkan konteks kalimat disebut *word sense disambiguation*. Bahasa Indonesia sebagai salah satu bahasa alami, memiliki kata yang bermakna ambigu dan memiliki struktur kalimat yang khas. Hal tersebut yang menarik minat penulis untuk membuat aplikasi yang dapat menghilangkan ambiguitas makna kata dalam kalimat berbahasa Indonesia.

Ada beberapa pendekatan untuk menghilangkan ambiguitas makna kata dalam kalimat berbahasa Inggris, yaitu *supervised learning*, dan *unsupervised*. Pendekatan *supervised learning* menggunakan data latih yang mengandung kumpulan besar contoh kalimat yang mengandung kata yang bermakna ambigu, setiap kata tersebut ditandai oleh manusia disertai makna dimana kata tersebut digunakan. Kemudian sekumpulan aturan secara otomatis belajar dari data latih tersebut. Sebagai contoh jika kata *dog* dan *bark* kedua-duanya ada dalam kalimat sedangkan kata *tree* tidak ada dalam kalimat tersebut, maka kata *bark* dapat diartikan sebagai lolongan anjing. Metode ini memiliki kelemahan yaitu tidak adanya sekumpulan aturan yang dapat melakukan penentuan makna untuk seluruh kata yang bermakna ambigu dan ketergantungan terhadap penandaan yang dilakukan oleh manusia kepada data latih, sehingga penentuan makna tidak dapat dilakukan terhadap kata yang tidak ditandai.

Pendekatan *unsupervised* menggunakan sumber informasi lain sebagai pengganti penandaan terhadap kata yang bermakna ambigu, yaitu kamus. Pendekatan ini diterapkan oleh algoritma Lesk. Algoritma ini berdasarkan intuisi bahwa kata yang bermakna ambigu yang terdapat bersamaan dalam kalimat, digunakan untuk merujuk topik yang sama dan makna yang berhubungan dengan topik tersebut didefinisikan di dalam kamus dengan menggunakan kata yang sama. Algoritma ini cocok untuk kalimat yang pendek, sedangkan untuk kalimat yang lebih panjang membutuhkan relasi antar kata dalam kalimat. Oleh karena itu maka algoritma Lesk diimplementasikan kepada semantik peristilahan yang terdapat dalam basis data atau disebut sebagai *WordNet*.

Dalam tugas akhir ini, homograf digunakan sebagai kata yang bermakna ambigu dan pendekatan yang digunakan adalah *unsupervised* karena dapat menghilangkan ambiguitas makna kata tanpa menggunakan penandaan kata yang bermakna ambigu, dilengkapi dengan alat bantu yaitu pengurai. Pengurai yang digunakan adalah hasil penelitian [10]. Pengurai tersebut menghasilkan pohon urai, kata dan kelas kata. Kelas kata dan kata hasil penguraian akan digunakan sebagai perbandingan terhadap kelas kata dari kata yang tersimpan dalam basis data. Kata yang bermakna ambigu biasanya memiliki kelas kata lebih dari satu. Jika kata yang bermakna ambigu memiliki kelas kata hanya satu buah, maka makna yang dipilih dari basis data adalah makna yang berdasarkan kelas kata yang dimilikinya. Jika kata yang bermakna ambigu memiliki kelas kata lebih dari satu, maka akan dilihat kesamaan dari kelas kata-nya. Misalnya kata yang

bermakna ambigu tersebut memiliki dua buah kelas kata, dan kedua-duanya adalah kata benda. Maka akan dilakukan proses penentuan makna dengan menggunakan algoritma Lesk yang diterapkan pada basis data semantik peristilahan atau *WordNet*. Dari penerapan algoritma Lesk akan diperoleh skor dari setiap makna yang menjadi kandidat. Kemudian dilakukan perbandingan terhadap skor – skor tersebut, dan skor terbesar yang akan dipilih sebagai acuan pemilihan makna dari makna kandidat. Jika misalnya kata yang bermakna ambigu tersebut memiliki dua buah kelas kata, yaitu kata benda dan kata sifat, maka makna yang dipilih dari basis data adalah makna yang berdasarkan kelas kata hasil penguraian.

1.2 Perumusan masalah

Titik berat tugas akhir adalah pembahasan tentang bagaimana cara menentukan makna yang harus dipilih untuk menghilangkan ambiguitas makna dari kata dalam kalimat berbahasa Indonesia. Adapun hal – hal yang mempengaruhi titik berat pembahasan, yaitu struktur kalimat bahasa Indonesia, jenis kelas kata yang akan digunakan, penguraian kalimat bahasa Indonesia, cara membaca hasil penguraian dan kamus yang akan digunakan.

Batasan masalah untuk tugas akhir ini adalah sebagai berikut :

1. Kalimat yang digunakan adalah kalimat yang sesuai dengan tata bahasa baku bahasa Indonesia
2. Kalimat yang digunakan berupa kalimat tertulis dan deklaratif.
3. Kalimat aktif dan pasif.
4. Tidak menangani frasa ambigu dan kata ambigu berimbuhan.
5. Aspek semantik peristilahan yang digunakan adalah homograf.
6. Asumsi basis data sudah ada dan lengkap.
7. Referensi kamus yang digunakan adalah Kamus Besar Bahasa Indonesia dan Kamus Lengkap Bahasa Indonesia. Penggunaan kedua kamus bertujuan untuk saling melengkapi data yang dimiliki oleh masing-masing kamus.

1.3 Tujuan

Penelitian ini bertujuan untuk membuat suatu aplikasi penghilangan ambiguitas makna kata dengan cara menggunakan pengurai, *WordNet* dan algoritma Lesk.

1.4 Metodologi penyelesaian masalah

Metode yang digunakan dalam penyelesaian tugas akhir ini adalah :

1. Mempelajari struktur kalimat dalam tata bahasa baku bahasa Indonesia, mempelajari teori pengurai, mempelajari jurnal-jurnal yang berkaitan dengan penghilangan ambiguitas makna kata dalam kalimat. Mempelajari cara kerja PC-PATR dan teori yang diimplementasikan oleh PC-PATR
2. Mengumpulkan kalimat – kalimat yang akan digunakan sebagai kasus uji.
3. Mengumpulkan kata - kata yang termasuk homograf beserta makna dan kelas kata dari homograf tersebut.
4. Mengumpulkan kata – kata yang biasanya berhubungan dengan homograf.

5. Mengumpulkan kata – kata yang biasanya berelasi dengan kata – kata yang berhubungan dengan homograf guna membangun *WordNet*.
6. Menganalisa algoritma Lesk dan menerapkannya ke dalam *WordNet*.
7. Merancang basis data, merancang antar muka.
8. Melakukan implementasi dari perancangan menggunakan PHP dan MySQL.
9. Melakukan pengujian hasil implementasi dan pengujian terhadap kalimat.
10. Evaluasi terhadap hasil pengujian