

# Bab I

## Pendahuluan

### 1.1 Latar Belakang

Bahasa merupakan salah satu elemen yang sangat penting dalam kehidupan. Menurut Gorys Keraf (1997 : 1), bahasa adalah alat komunikasi antara anggota masyarakat berupa simbol bunyi yang dihasilkan oleh alat ucap manusia. Kemampuan setiap orang dalam mengucapkan bahasa dan memberikan intonasi terhadap kata yang diucapkan pastilah berbeda-beda. Penekanan dan pengucapan kata tersebut kemudian akan sangat berpengaruh terhadap sistem pengenalan ucapan kata. Pada sistem pengenalan ucapan, sinyal ucapan yang dikenali akan dibandingkan dengan model bahasa akustik dari suara pengucapan manusia.

Dalam banyak bahasa, sistem pengenalan ucapan dapat berupa sistem pengenalan ucapan berbasis silabel ataupun berbasis fonem [6]. Akan tetapi, dari sejumlah penelitian yang telah dilakukan, menunjukkan bahwa sistem Pengenalan Ucapan Kontinu Kosakata Besar (PUKKB) berbasis silabel memberikan performansi lebih baik dibandingkan dengan yang berbasis fonem [13]. Hal tersebut kemudian membuat penelitian terhadap silabel menjadi lebih difokuskan saat ini.

PUKKB berbasis silabel memiliki beberapa keunggulan diantaranya memiliki ketahanan yang lebih baik akan sinyal suara gangguan atau derau dan dapat mengatasi serta beradaptasi dengan berbagai dialek Bahasa yang ada [3]. Salah satu tantangan yang dihadapi dalam sistem pengenalan ucapan yaitu tingkat keakuratan pada proses silabifikasi. Ketidakakuratan proses silabifikasi akan menyebabkan hasil pengenalan kata atau kalimat menjadi tidak jelas dan akan menimbulkan ambiguitas yang tinggi. Silabifikasi merupakan proses untuk memisahkan kata kedalam bentuk silabel atau suku kata [1]. Setiap Bahasa dapat memiliki struktur silabelnya masing-masing. Untuk Bahasa Indonesia sendiri tergolong dalam bahasa yang memiliki kompleksitas silabel yang tinggi. Dimana dalam satu kata dapat mengandung lima atau lebih silabel. Misalnya pada kata ‘pertanggungjawaban’ yang memiliki enam silabel. Model silabel yang diaplikasikan pada bahasa Indonesia adalah silabel pada barisan grafem. Namun hal tersebut memberikan performansi yang relatif rendah dengan ting-

kat *Syllable Error Rate* (SER) yang cukup tinggi. Hal tersebut dikarenakan ambiguitas serta pengecualian pada barisan grafem yang kurang diperhatikan. Pengecualian silabifikasi dalam bahasa Indonesia dapat diartikan sebagai suatu kumpulan kata-kata yang memiliki barisan grafem yang hampir sama serta memiliki pemenggalan suku kata yang sama pula namun terdapat satu kata yang memiliki kemiripan dengan kata tersebut tetapi memiliki pemenggalan suku kata yang berbeda sendiri. Satu kata yang memiliki pemenggalan yang berbeda tersebutlah yang disebut dengan pengecualian. Kata yang termasuk pengecualian dalam bahasa Indonesia adalah kata ‘tepati’ yang disilabifikasi menjadi ‘te.pa.ti’, dimana kata lain yang mirip dengan kata tersebut seperti tepat, sempit, tempat memiliki silabifikasi yang sama yaitu ‘te.pat’, ‘sem.pat’, ‘tem.pat’ yaitu grafem  $\langle a \rangle$  dan  $\langle t \rangle$  selalu berada pada satu silabel yang sama. Pengecualian pada kata dalam bahasa Indonesia biasanya terjadi jika suatu kata diikuti oleh sebuah huruf vokal pada akhir kata.

Masalah lain dalam proses silabifikasi adalah proses silabifikasi dengan menggunakan kata berupa nama orang. Pengucapan nama seseorang memiliki tingkat ambiguitas yang cukup tinggi. Misalnya pengucapan untuk nama ‘Adhi’, yang dapat diucapkan menjadi ‘Adi’, ‘Adhi’, dimana pada kata ‘Adi’ dan ‘Adhi’ akan menghasilkan silabifikasi yang berbeda yaitu menjadi ‘A.di’ dan ‘Ad.hi’.

Untuk mengatasi masalah tersebut akan dibangun sebuah model sistem yang dapat digunakan dalam proses silabifikasi untuk meningkatkan keakuratan proses silabifikasi, dan mengurangi *error* serta ambiguitas yang ada. Model sistem tersebut akan dibangun menggunakan sebuah metode statistik yang bekerja secara lokal yaitu *Fuzzy K-Nearest Neighbor in Every Class* (FKNNC). FKNNC merupakan salah satu metode varian dari *K-Nearest Neighbor*. Dalam penerapannya, metode tersebut menggabungkan *K-Nearest Neighbor* (KNN) dan *Fuzzy System*. KNN yang diterapkan bekerja sebagai penentu kedekatan jarak antar data uji dan kelas data sementara *Fuzzy System* akan menangani masalah ambiguitas yang mungkin ditimbulkan oleh KNN. Penerapan kedua metode tersebut dianggap mampu untuk memberikan hasil klasifikasi dengan tingkat kesalahan yang rendah. Pemilihan metode berbasis statistik lokal juga dikarenakan masalah silabifikasi grafemis tidak bisa diselesaikan menggunakan metode statistik yang bersifat global seperti Naive Bayes karena pengambilan seluruh data untuk menentukan titik silabis akan memberikan hasil yang kurang optimal.

Selain itu, pemilihan metode berbasis statistik lokal dikarenakan pada pembangunan kata berupa nama orang, tidak memiliki suatu aturan tertentu atau aturan yang pasti. Oleh sebab itu tidak bisa diterapkan metode berbasis aturan dalam proses silabifikasi grafemis pada Tugas Akhir ini.

## 1.2 Perumusan Masalah

Sebagai bahasa yang kompleks, dengan banyak pengecualian dan ambiguitas masalah silabifikasi grafemis bahasa Indonesia sulit diselesaikan menggunakan metode berbasis statistik yang bersifat global.

## 1.3 Batasan Masalah

Model silabifikasi grafemis diterapkan pada kata-kata dalam Kamus Besar Bahasa Indonesia (KBBI) dan nama-nama orang yang diperoleh dari Forum Alumni Universitas Telkom (FAST).

## 1.4 Tujuan

Membangun model silabifikasi grafemis bahasa Indonesia menggunakan metode berbasis statistik yang bekerja secara lokal.

## 1.5 Rencana Kegiatan

### 1. Studi Literatur

Studi literatur pada Tugas Akhir ini dilakukan dengan cara mengumpulkan, mempelajari serta memahami materi terkait dengan silabifikasi dan grafem dalam bahasa Indonesia serta metode *Fuzzy K-Nearest Neighbor in every Class*. Pengumpulan literatur dilakukan dengan mencari dan mengumpulkan informasi terkait, yang berasal dari paper, buku dan situs web.

### 2. Pengumpulan Data

Pengumpulan data dilakukan dengan mengambil data yang berasal dari Kamus Besar Bahasa Indonesia (KBBI) Luar Jaringan edisi ketiga (Bahasa, 2008), serta data nama orang yang diperoleh dari Forum Alumni Universitas Telkom (FAST).

### 3. Perancangan dan Pembangunan Sistem

Pada penelitian terkait Tugas Akhir ini model yang akan dibangun berupa model silabifikasi grafemis bahasa Indonesia yang akan menentukan titik silabis yang tepat pada kata yang merupakan masukan sistem dengan proses klasifikasi.

### 4. Implementasi Sistem

Mengimplementasikan model sistem yang akan dibangun sesuai dengan rancangan yang telah dibuat.

### 5. Pengujian dan Analisis Sistem

Menguji model sistem yang telah dibangun untuk mengetahui apakah model tersebut telah sesuai dengan dengan rancangan yaitu apakah dapat mengatasi masalah ambiguitas pada kata dan menurunkan *Syllable Error Rate* (SER).

## **6. Pembuatan Laporan dan Dokumentasi**

Membuat laporan hasil dan dokumentasi dari hasil pembangunan model sistem pada Tugas Akhir ini.