

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Sistem Pengenalan Ucapan Kontinu Kosakata Besar (PUKKB) mempunyai fungsi mengubah suara menjadi teks. Sistem tersebut banyak yang didesain dengan menggunakan fonem sebagai basisnya. Namun, Ganapathiraju[6] dalam penelitiannya mendapatkan hasil yang lebih bagus dengan penggunaan silabel. Penggunaan tanda baca dalam sebuah kalimat dapat mempengaruhi makna, yang sering diekspresikan dengan intonasi yang berbeda, dari kalimat tersebut[2]. "Aku." dan "Aku?" mempunyai makna yang berbeda walau memiliki satu kata pembangun "aku" yang sama. Pembicara sering menggunakan intonasi untuk memperjelas maksud dari apa yang dibicarakan. "Aku." adalah sebuah pernyataan yang identik dengan nada datar, sedangkan "Aku?" akan memiliki nada lebih tinggi pada silabel 'ku'. Kedua hal di atas menjadi dasar pertimbangan dibutuhkan sebuah sistem PUKKB dengan berbasis silabel dan mengakomodir penggunaan tanda baca. Untuk membangun sistem tersebut, dibutuhkan sebuah korpus teks yang didesain dengan mempertimbangkan kemunculan silabel dan tanda baca. Korpus teks yang digunakan adalah himpunan atau sekumpulan kalimat unik. Algoritma *LTM Greedy* dapat menghasilkan himpunan kalimat dengan jumlah kalimat paling sedikit dengan waktu eksekusi yang lebih cepat [1]

### 1.2 Rumusan Masalah

Korpus teks Bahasa Indonesia saat ini masih kurang seimbang secara silabel dan intonasi suara sehingga belum dapat digunakan untuk membangun korpus suara yang berbasis silabel dan mengakomodasi intonasi suara yang berbeda-beda.

### 1.3 Tujuan

Membangun korpus teks yang lebih seimbang secara silabel dan intonasi suara dengan memperhatikan tanda baca.

### 1.4 Batasan Masalah

1. Himpunan kalimat induk yang berisi 10 juta kalimat unik didapatkan dari hasil sejumlah website media massa.
2. Himpunan kalimat induk yang digunakan hanya berbahasa Indonesia.

### 1.5 Metodologi Penyelesaian Masalah

1. Study Literatur.  
Pencarian referensi yang berkaitan dengan Algoritma *Greedy Search*, *speech corpus*, silabel dan penggunaan tanda baca.
2. Konsultasi dengan Pembimbing Tugas Akhir
3. Pengumpulan data kalimat menggunakan cara *web crawling*.
4. Perancangan Sistem  
Perancangan sistem dimulai dari merancang ekstraksi silabel, kemudian algoritma greedy.
5. Implementasi Rancangan sistem yang telah dibuat diimplementasikan dengan menggunakan IDE Netbeans dan bahasa pemrograman Java.
6. Pengujian dan Analisis Sistem kemudian dijalankan dengan data yang sudah ada. Hasil kemudian dianalisis.
7. Penulisan Laporan Tugas Akhir

### 1.6 Sistematika Penulisan

1. **BAB I PENDAHULUAN:** Pada bab ini berisi latar belakang, perumusan masalah, tujuan, metode penyelesaian masalah, serta sistematika penulisan.
2. **BAB II DASAR TEORI:** Pada bab ini berisi penjelasan singkat mengenai landasan-landasan teori yang berhubungan dengan masalah yang dibahas.

3. **BAB III PERANCANGAN SISTEM:** Bab ini berisi mengenai perancangan sistem, yaitu bagaimana sistem akan dibuat mulai dari perancangan processing, dan algoritma *greedy*.
4. **BAB IV PENGUJIAN DAN ANALISIS:** Bab ini berisi mengenai hasil pengujian sistem yang telah dibuat sesuai dengan rancangan yang dibuat di bab sebelumnya. Bab ini juga berisi hasil analisis terhadap algoritma *greedy* yang digunakan.
5. **BAB V KESIMPULAN DAN SARAN:** Pada bab ini dikemukakan kesimpulan yang dapat diambil dari hasil analisis sistem dan saran untuk pengembangan penelitian selanjutnya.