

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Android merupakan salah satu *mobile Operating System* atau sistem operasi untuk perangkat *mobile* yang berbasis *linux*, berupa *software platform open source* yang dapat mengontrol sistem [14]. Semakin banyaknya pengguna *smartphone* berbasis android maka dapat dimanfaatkan untuk memperkenalkan salah satu budaya yang ada di indonesia yaitu budaya sunda dengan cara mengembangkan aplikasi android yang memiliki fitur kamus bahasa sunda

Bahasa merupakan alat komunikasi manusia yang digunakan untuk mengadakan hubungan dalam pergaulan sehari-hari, selain itu juga berfungsi sebagai sarana untuk menyampaikan informasi [9]. manusia hidup dalam sebuah dunia bahasa. Kegiatan-kegiatan manusia seperti bekerja, belajar, berbicara, atau aktivitas serupa, hampir seluruhnya menggunakan komunikasi bahasa. Ketika bertemu kata-kata asing, maka umumnya kamus digunakan sebagai alat bantu untuk mengerti kata-kata tersebut.

Penggunaan kamus cetak untuk mencari arti kata cenderung tidak efisien karena pengguna harus mencari kata satu demi satu, membolak-balik halaman kamus, dan mencari dengan teliti. Hal inilah yang mendorong dibuatnya kamus digital. Kamus digital membuat pencarian kata menjadi lebih cepat dan mudah. Namun, masalah akan muncul ketika pengguna tidak memahami sistem penulisan kata pada bahasa yang ingin diterjemahkan. Sebagai contoh, ketika seseorang

ingin mengetik karakter sunda pada kamus digital namun tidak memahami cara penulisannya.

Bahasa Sunda adalah sebuah bahasa dari cabang Melayu-Polinesia dalam rumpun bahasa Austronesia. Bahasa ini dituturkan oleh setidaknya 42 juta orang dan merupakan bahasa Ibu dengan penutur terbanyak kedua di Indonesia setelah Bahasa Jawa [10]. Bahasa sunda merupakan salah satu bahasa yang sulit untuk dipelajari. Bahasa Sunda dipelajari karena penulisannya yang relatif rumit dan Bahasa Sunda merupakan bahasa yang unik dengan tingkatan-tingkatan berbahasa, atau lebih dikenal dengan istilah undak-usuk yang nyaris tidak dimiliki oleh bahasa lain. Saat ini, bahasa Sunda ditulis dengan abjad latin. Bahasa Sunda memiliki lima huruf vokal murni, yaitu a, i, u, é, o, serta dua vokal netral yaitu, e pepet dan eu. Konsonan bahasa Sunda terdiri dari 18 fonem, yaitu p, b, t, d, k, g, c, j, h, ng, ny, m, n, s, w, l, r, dan y. Kesulitan para pendatang berbicara bahasa Sunda adalah ketika melafalkan suara eu. Misalnya, untuk mengucapkan "Cicaheum", para pendatang sering mengucapkan eu dengan e pepet sehingga bunyinya menjadi "Cicahem", tanpa "u" [10].

Dalam melakukan pencocokan kata pada kamus tersebut, penulis menggunakan algoritma boyer more sebagai algoritma pencocokannya karena dianggap paling *efisien* dari algoritma lainnya seperti bruteforce dan Knuth Morris Pratt. Karena algoritma boyer moore mulai melakukan pencocokan pada karakter paling kanan *pattern*, dan jika tidak terjadi kecocokan maka pencocokan langsung melakukan lompatan karakter sebanyak *pattern* sehingga membuatnya lebih *efisien* dalam loncatan karakter saat pencocokan tersebut [19].

Untuk mengatasi masalah pengetikan karakter, maka penulis memanfaatkan teknologi *Optical Character Recognition* (OCR). OCR adalah sebuah sistem komputer yang dapat membaca huruf. OCR dapat diimplementasikan pada *platform Android* tentunya ini memanfaatkan sumber dari Google yang telah menyediakan banyak sumber referensi dan salah satunya adalah menggunakan *Tesseract Engine* sebagai metode untuk menerapkan teknologi OCR pada *platform Android*. Pemanfaatan sumber yang sudah ada akan dibuat sebuah aplikasi yang dapat mengubah citra teks yang bisa diimplementasikan pada perangkat *Android*. Aplikasi ini nantinya akan menerjemahkan suatu objek berupa karakter teks/tulisan dalam bahasa Sunda yang tertangkap oleh kamera ke dalam bahasa Indonesia. Pengimplementasian OCR pada perangkat *Android* akan dimanfaatkan sebagai penerjemah bahasa Sunda– Indonesia secara online yang terhubung dengan *library* Microsoft Translator dan Google Translator [18]., Dengan adanya sistem OCR maka pengguna dapat lebih leluasa memasukkan kata karena pengguna tidak harus memakai papan ketik.

Berdasarkan latar belakang tersebut, terbentuklah tugas akhir yang berjudul ***“Implementasi Algoritma Boyer-Moore Pada Aplikasi Kamus Bahasa Sunda-Indonesia Menggunakan Tesseract Untuk Fitur Optical Character Recognition (OCR)”***

1.2 Rumusan Masalah

Berdasarkan permasalahan yang telah dijabarkan sebelumnya, maka diambil perumusan masalah yaitu sebagai berikut :

1. Bagaimana membangun aplikasi kamus bahasa sunda-indonesia dengan mengimplementasikan Algoritma *Boyer-Moore* ?
2. Bagaimana membuat sebuah aplikasi kamus penerjemah berbasis *mobile* dimana kata yang akan diterjemahkan tidak lagi diinput secara manual melainkan dicapture dan diproses dengan menggunakan fitur OCR (*Optical Character Recognition*) dengan menggunakan *Tesseract* yang dikembangkan oleh Google ?

1.2 Tujuan Penelitian

Adapun tujuan dari penelitian ini yaitu sebagai berikut :

- a. Merancang dan membangun aplikasi kamus bahasa sunda-indonesia dengan mengimplementasikan Algoritma *Boyer-Moore*.
- b. Menerapkan penggunaan OCR (*Optical Character Recognition*) dalam aplikasi penerjemah untuk mendapatkan *input* karakter melalui kamera.

1.3 Batasan Masalah

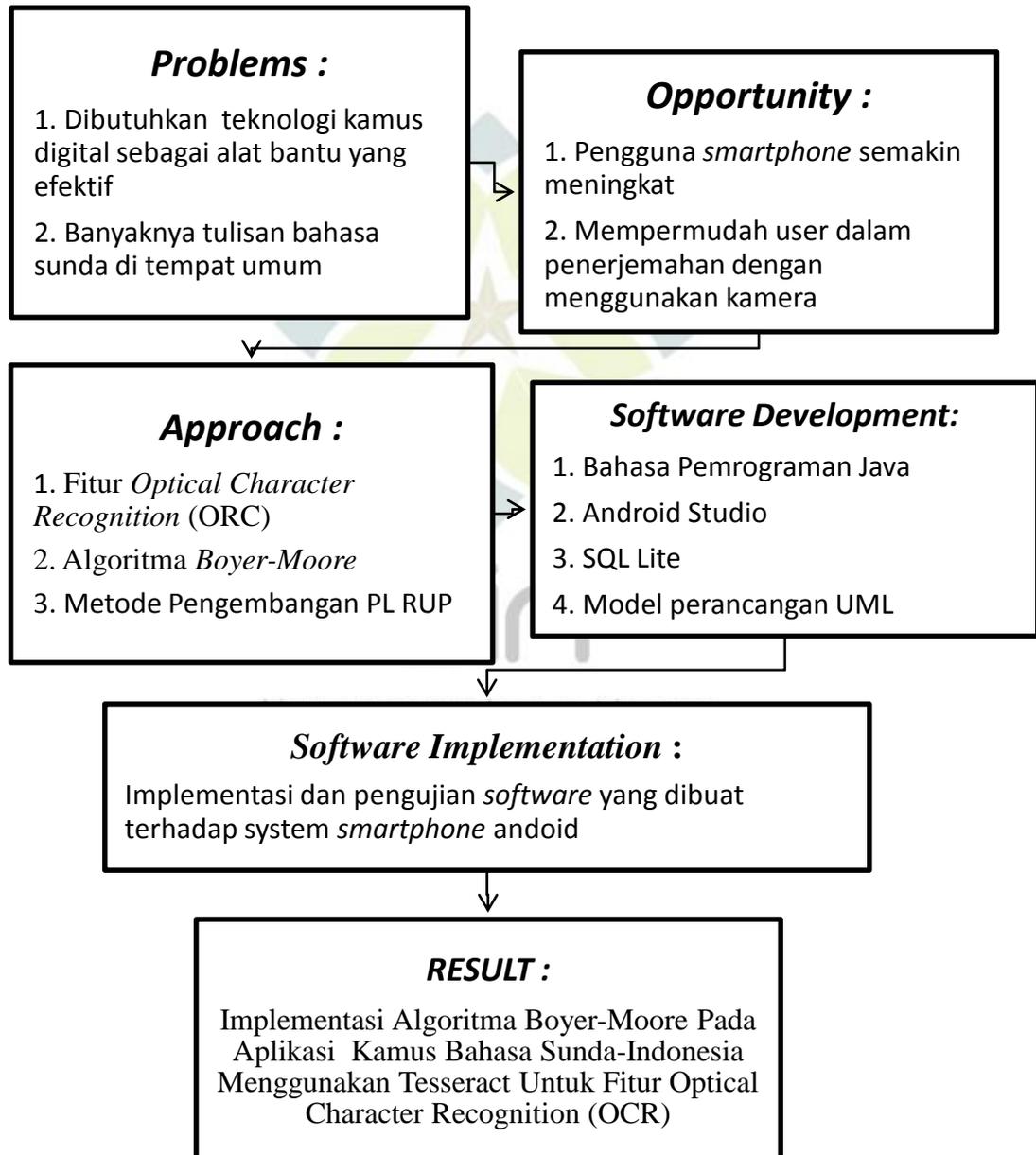
Adapun batasan masalah dalam penelitian ini yaitu sebagai berikut :

- a. Aplikasi kamus Bahasa ini dapat melakukan pencarian kata melalui bahasa Sunda lalu diartikan kedalam bahasa Indonesia dengan mengimplementasikan algoritma *Boyer-Moore*.
- b. Aplikasi ini akan menerjemahkan teks dari gambar yang ditangkap oleh kamera *smartphone*, baik dari buku tulis ataupun papan tulis.
- c. Jumlah data yang ada didalam kamus bahasa sunda ini yaitu kecap sunda yang banyaknya 8805 kecap dan babasan 1330 kata.
- d. Pencarian arti hanya dibatasi pada pencarian arti per-kata atau frase.

- e. Aplikasi ini menggunakan *Tesseract* sebagai *library* yang digunakan untuk melakukan konversi *image to text*.

1.4 Kerangka Pemikiran

Adapun kerangka pemikiran dalam penelitian ini yaitu seperti pada Gambar 1.1 sebagai berikut.



Gambar 1.1 Kerangka Pemikiran

1.5 Metode Penelitian

1.5.1 Tahap Pengumpulan Data

Untuk mengumpulkan berbagai data yang diperlukan oleh kelancaran pembuatan aplikasi ini ada beberapa tahap yang akan dilakukan sehingga hasil yang didapatkan menjadi maksimal.

1. Studi Literatur

Studi literatur merupakan tahap mencari beberapa materi serta keterkaitan data ataupun perbandingan jurnal yang telah membahas kajian yang sama dengan tema penelitian ini. Biasanya ke pustakaan disini lebih ke sisi pengembangan aplikasi yang akan dibuat serta *user interface* yang cocok dengan bahan kajian. Sehingga pada saat mengalami kesulitan dalam mengembangkan masalah, dapat terpecahkan dengan beberapa studi pustaka diatas.

2. Studi Observasi

Studi observasi yaitu pengumpulan data dengan cara melakukan pendataan langsung dengan mempelajari dan meneliti data-data yang sudah ada sebelumnya.

1.5.2 Metodologi Pengembangan Perangkat Lunak

Metodologi *Rational Unified Process* (RUP). Metode *RUP* merupakan metode pengembangan kegiatan yang berorientasi pada proses [12]. Metodologi RUP memiliki empat tahap pengembangan perangkat lunak yaitu:

1. *Inception* (Permulaan)

Tahap ini lebih pada memodelkan proses bisnis yang dibutuhkan (*business modeling*) dan mendefinisikan kebutuhan akan sistem yang akan dibuat (*requirement*).

2. *Elaboration* (Perluasan/Perencanaan)

Tahap ini difokuskan pada perencanaan arsitektur sistem. Tahap ini juga dapat mendeteksi apakah arsitektur sistem yang diinginkan dapat dibuat atau tidak. Tahap ini lebih pada analisis dan desain sistem serta implementasi sistem yang fokus pada *prototype*.

3. *Construction* (Konstruksi)

Tahap ini berfokus pada pengembangan komponen dan fitur-fitur sistem. Tahap ini lebih pada implementasi dan pengujian sistem yang berfokus pada implementasi perangkat lunak pada kode program.

4. *Transition* (Transisi)

Tahap ini lebih pada *deployment* atau instalasi sistem agar dapat dimengerti oleh *user*. Tahap ini menghasilkan produk perangkat lunak dimana menjadi syarat *Initial Operational Capability Milestone* atau batas/tonggak kemampuan operasional awal. Aktifitas pada tahap ini termasuk pada pelatihan *user*, pemeliharaan dan pengujian sistem apakah sudah memenuhi harapan *user*.

1.6 Sistematika Penulisan

Sistematika pembuatan perangkat lunak ini dibagi menjadi 5 (lima) bab yang masing-masing bab telah dirancang dengan suatu tujuan tertentu. Berikut penjelasan tentang masing-masing bab :

BAB I : PENDAHULUAN

Bab I menguraikan latar belakang masalah, perumusan masalah yang meumuskan berbagai masalah yang diteliti secara lebih jelas, batasan masalah untuk memberikan batasan yang tegas dan jelas serta sistematika penulisan yang menguraikan urutan penyajian yang digunakan dalam penyusunan skripsi ini.

BAB II : LANDASAN TEORI

Bab II membahas tentang landasan teori dari topik penulisan skripsi secara mendalam beserta dengan referensinya.

BAB III: ANALISIS DAN PERANCANGAN

Bab III akan menguraikan hasil analisis dan perancangan aplikasi yang akan dibangun

BAB IV : IMPLEMENTASI DAN PENGUJIAN

Bab IV akan menguraikan implementasi aplikasi yang telah dianalisa dan dirancang sebelumnya kemudian dilakukan proses pengujian terhadap aplikasi yang telah dibangun.

BAB V : PENUTUP

Bab V berisi uraian tentang kesimpulan dan saran terhadap software yang hendak dibangun dan bila akan dikembangkan lebih lanjut.