PEMBUATAN APLIKASI PENGOLAHAN CITRA TEKS INGGRIS DAN PENERJEMAHANNYA KE DALAM BAHASA INDONESIA BERBASIS SISTEM OPERASI ANDROID

Irman Nurmulya Malik, Wendi Zarman
Jurusan Teknik Komputer Unikom, Bandung
irman.n.m@gmail.com, wendizar@gmail.com

ABSTRAK

Bahasa Inggris merupakan bahasa resmi di dunia Internasional yang sering ditemukan dalam kehidupan sehari-hari. Masyarakat di Indonesia saat ini dituntut untuk bisa berbahasa Inggris, karena Bahasa Inggris merupakan salah satu bahasa yang diakui di dunia Internasional. Namun ini menjadi kendala karena masyarakat Indonesia pada umumnya mengunakan Bahasa Indonesia dalam kehidupan sehari-hari. Kondisi ini menyebabkan sulitnya dalam mengetahui arti terjemahan dari teks berbahasa Inggris. Aplikasi ini dibuat untuk dapat memudahkan masyarakat Indonesia mengetahui arti terjemahan teks Bahasa Indonesia pada saat membaca koran, majalah, atau buku teks berbahasa Inggris, dengan cara memfoto teks yang ingin diketahui arti terjemahannya, kemudian aplikasi mengenali teks hasil foto tersebut dan penerjemahannya ke dalam teks Bahasa Indonesia.

Kata kunci: Bahasa Inggris, Aplikasi Android, Gambar teks, Kamus.

1. PENDAHULUAN

Kamus terjemahan merupakan sebuah buku yang berisi tentang terjemahan dari satu bahasa ke bahasa lainnya. Kenyataan dalam kehidupan sehari-hari kamus dirasa kurang efisien selain karena tebal dan berat, juga lama dalam mencari kata yang ingin diterjemahkan, sehingga ini menjadi salah satu kendala yang terjadi. Masalah lain muncul ketika masyarakat Indonesia dalam kehidupan sehari-hari selalu menggunakan Bahasa Indonesia, yang menyebabkan semakin sulit untuk terbiasa berbahasa asing khususnya Bahasa Inggris, sehingga dibutuhkan suatu aplikasi yang dapat membantu proses dalam menerjemahkan suatu bahasa ke bahasa yang lainnya dengan proses yang cukup efesien.

2. TEORI PENUNJANG

2.1 Sistem Operasi Android

Sistem Operasi Android merupakan salah satu sistem operasi *open source* berbasis linux yang digunakan untuk perangkat seluler *smartphone*. Android dikembangkan dalam bahasa pemrograman *java*, sehingga memungkinkan para *programmer* untuk membuat berbagai macam aplikasi, seperti aplikasi permainan, aplikasi edukasi, dan aplikasi lainnya.

2.1.1 Struktur Aplikasi Sistem Operasi Android

Struktur aplikasi pada sistem operasi android mengadopsi konsep *component reuse* yang memungkinkan aplikasi dapat berbagi aktifitas, layanan dan data dengan aplikasi lain, dengan akses kelola yang terbatas.

2.2 Pengolahan Citra

Suatu citra adalah fungsi intesitas 2 dimensi f(x,y), dimana x dan y adalah koordinat spasial dan f pada titik (x,y) merupakan tingkat kecerahan (*brightness*) suatu citra pada suatu titik. Citra digital ialah representasi dari sebuah citra 2 dimensi yang dimana kumpulan nilai digital disebut piksel atau elemen gambar. Piksel adalah satuan terkecil dari citra yang mengandung nilai yang mewakili kecerahan dari sebuah warna pada sebuah titik tertentu.

2.2.1 Jenis-jenis dari Pengolahan Citra

Suatu pixel memiliki nilai dalam nilai rentang tertentu, dari rentang nilai minimum hingga rentang nilai maksimum. Jangkauan rentang tiap piksel berbeda-beda tergantung dari jenis warnanya. Secara umum jangkauan dari nilai rentangnya adalah 0-255. Pengolahan citra dengan penggambaran seperti ini digolongkan ke dalam pengolahan citra

integer. Berikut adalah jenis-jenis pengolahan citra berdasarkan nilai pixelnya.

2.2.1.1 Pengolahan Citra Warna

RGB (*Red Green Blue*) adalah model warna yang terdiri atas warna merah, hijau, dan biru, yang digabungkan dalam membentuk suatu susunan warna yang luas. Citra warna sendiri memiliki 8 bit yang mewakili setiap piksel dengan jumlah warna maksimal berjumlah 256 warna.

2.2.1.2 Pengolahan Citra *Grayscale*

Pengolahan Citra *Grayscale* memiliki tingkatan warna hitam, abu-abu dan putih. Tingkatan warna abu-abu merupakan warna abu dengan berbagai tingkatan dari warna hitam yang mendekati hingga warna putih.

2.2.1.3 Pengolahan Citra Biner

Pengolahan citra biner adalah warna-warna piksel yang berada dalam rentang gradasi warna hitam dan putih.

2.2.2 OCR

Dalam pengolahan citra ada beberapa aplikasi yang dapat mengenali objek yaitu aplikasi pengenalan objek. Salah satu dari aplikasi pengenal objek adalah OCR (*Optical Character Recognition*). Aplikasi OCR ini sering digunakan dalam mengidentifikasi citra teks yang kemudian diubah ke dalam bentuk file tulisan.

2.2.3 Segmentasi Pengolahan Citra

Segmentasi citra merupakan bagian dari proses pengolahan citra. Proses segmentasi citra ini merupakan suatu proses sebelum pengolahan pada sistem pengenalan objek dalam citra. Segmentasi citra (*image segmentation*) mempunyai arti membagi suatu citra menjadi wilayah-wilayah yang homogen berdasarkan kriteria keserupaan yang tertentu antara tingkat keabuan suatu piksel dengan tingkat keabuan piksel – piksel tetangganya, kemudian hasil dari proses segmentasi ini akan digunakan untuk proses tingkat tinggi lebih lanjut yang dapat dilakukan terhadap suatu citra, misalnya proses klasifikasi citra dan proses identifikasi objek.

2.2.4 Pengambangan

Metode *thresholding* atau pengambangan adalah proses mengubah citra berderajat keabuan menjadi citra biner atau hitam putih sehingga dapat diketahui daerah mana yang termasuk objek dan *background* dari citra secara jelas. Citra hasil *thresholding* biasanya digunakan lebih lanjut untuk proses pengenalan obyek serta ekstraksi fitur.

2.2.5 Pengambangan dengan Metode Otsu

Tujuan dari *thresholding* dengan metode otsu adalah membagi histogram citra tingkat *grayscale* ke dalam dua daerah yang berbeda secara otomatis tanpa membutuhkan bantuan *user* untuk memasukkan nilai ambang. Pendekatan yang dilakukan dengan metode otsu adalah dengan melakukan analisis diskriminan yaitu menentukan suatu variabel yang dapat membedakan antara dua atau lebih kelompok yang muncul secara alami. Analisis Diskriminan akan memaksimumkan *variable* tersebut agar dapat membagi objek *foreground* dan *background*.

2.3 OCR SDK

OCR SDK (*Software Development Kit*) atau *optical character recognition* SDK adalah salah satu aplikasi yang digunakan untuk mengenali objek atau gambar teks, yang dimana setelah OCR mengenali gambar teks, selanjutnya OCR akan merubah objek teks menjadi sebuah teks atau tulisan.

2.4 Microsoft Translator

Microsoft translator merupakan layanan mesin penerjemah bahasa, menerjemahkan secara otomatis dari satu bahasa ke bahasa lainnya yang digunakan di aplikasi web. Cara kerja aplikasi ini sama halnya dengan cara kerja mesin penerjemah lainnya seperti Microsoft Translator. Aplikasi ini sendiri terdapat di Windows Azure Marketplace.

2.5 Pemrograman Berorientasi Objek

Pemrograman berorientasi objek atau *object-oriented programming* (OOP) adalah pemrograman yang hanya berorientasi kepada objek saja. Semua data dan fungsi dikategorikan dalam *object-object*. Setiap objek dapat saling berinteraksi dalam mengirim pesan atau data, memproses, menerima, menyimpan dan memanipulasi data atau pesan, yang dimana masing-masing objek harus memiliki informasi mengenai dirinya sendiri dan juga dapat saling berinterkasi dengan objek-objek yang lain.

2.6 Bahasa Inggris (*English*)

Bahasa Inggris adalah bahasa Jermanik yang pertama kali dgunakan di Inggris pada Abad Pertengahan Awal dan saat ini merupakan bahasa yang paling umum digunakan di seluruh dunia. Bahasa Inggris digunakan sebagai bahasa pertama oleh mayoritas penduduk di berbagai negara, termasuk Britania Raya, Irlandia, Amerika Serikat, Kanada, Australia, Selandia Baru, dan sejumlah negara-negara Karibia, serta menjadi bahasa resmi di 60 negara berdaulat. Bahasa Inggris adalah bahasa ketiga yang paling banyak dituturkan di seluruh dunia, setelah bahasa Mandarin dan bahasa Spanyol. Bahasa Inggris juga digunakan sebagai bahasa kedua dan bahasa resmi oleh Uni Eropa, Negara Persemakmuran, dan Perserikatan Bangsa-Bangsa, serta beragam organisasi lainnya.

2.6.1. Huruf Inggris dan Penulisannya

Bahasa Inggris mempunyai 26 fonem yaitu 21 huruf mati dan 5 huruf hidup. Seperti yang dapat dilihat pada gambar 1.

Huruf besar	Huruf kecil	IPA	Huruf besar	Huruf kecil	IPA
A	a	/eɪ/	N	n	/ɛn/
B	b	/biː/	O	o	/eʊ/
C	c	/siː/	P	p	/piː/
D	d	/diː/	Q	q	/kjuː/
E	e	/iː/	R	r	/ɑr/
F	f	/ɛf/	S	s	/ɛs/
G	g	/dʒiː/	T	t	/tiː/
H	h	/eɪtʃ/	U	u	/juː/
I	i	/aɪ/	V	v	/viː/
J	j	/dʒeɪ/	W	w	/ˈdʌbəljuː/
K	k	/keɪ/	X	x	/ɛks/
L	l	/ɛl/	Y	y	/waɪ/
M	m	/ɛm/	Z	z	/ziː/

Gambar 1. Huruf Inggris dan penulisannya.

2.7 Bahasa Indonesia (*Indonesian*)

Bahasa Indonesia adalah bahasa Melayu yang dijadikan sebagai bahasa resmi Republik Indonesia dan bahasa persatuan bangsa Indonesia. Bahasa Indonesia diresmikan penggunaannya setelah Proklamasi Kemerdekaan Indonesia, tepatnya sehari sesudahnya, bersamaan dengan mulai berlakunya konstitusi. Dari sudut pandang linguistik, bahasa Indonesia adalah salah satu dari banyak ragam bahasa Melayu. Dasar yang dipakai adalah bahasa Melayu Riau (wilayah Kepulauan Riau). Dalam perkembangannya Bahasa Indonesia mengalami perubahan akibat penggunaanya sebagai bahasa kerja di lingkungan administrasi kolonial dan berbagai proses pembakuan sejak awal abad ke-20. Penamaan Bahasa Indonesia diawali sejak dicanangkannya Sumpah Pemuda, 28 Oktober 1928, untuk menghindari kesan imperialisme bahasa apabila nama bahasa Melayu tetap digunakan. Proses ini menyebabkan berbedanya Bahasa Indonesia saat ini dari varian bahasa Melayu yang digunakan di Riau maupun Semenanjung Malaya. Hingga saat ini, Bahasa Indonesia merupakan bahasa yang hidup, yang terus menghasilkan kata-kata baru, baik melalui penciptaan maupun penyerapan dari bahasa daerah dan bahasa asing. Meskipun dipahami dan digunakan oleh lebih dari 90% warga Indonesia, Bahasa Indonesia bukanlah bahasa bagi kebanyakan penggunanya. Sebagian besar warga Indonesia menggunakan salah satu dari 748 bahasa yang ada di Indonesia sebagai bahasa sehari-hari. Penutur Bahasa Indonesia kerap kali menggunakan versi sehari-hari (kolokial) dan mencampuradukkan dengan dialek Melayu lainnya atau bahasa daerah lainnya. Meskipun demikian, Bahasa Indonesia digunakan sangat luas di perguruan-perguruan, di media massa, sastra, perangkat lunak, surat-menyurat resmi, dan berbagai forum publik lainnya, sehingga dapat dikatakan bahwa bahasa Indonesia digunakan oleh semua warga Indonesia.

2.7.1. Huruf Indonesia dan Penulisannya

Seperti halnya Bahasa Inggris, Bahasa Indonesia juga mempunyai 26 fonem yaitu 21 huruf mati dan 5 huruf hidup. Seperti yang dapat dilihat pada gambar 2.

Huruf besar	Huruf kecil	IPA	Huruf besar	Huruf kecil	IPA
A	a	/ɑ:/	N	n	/n/
B	b	/b/	O	o	/ɔ, o/
C	c	/tʃ/	P	p	/p/
D	d	/d/	Q	q	/q/
E	e	/e, ɛ, ə/	R	r	/r/
F	f	/f/	S	s	/s/
G	g	/g/	T	t	/t/
H	h	/h/	U	u	/u/
I	i	/i/	V	v	/v, ʊ/
J	j	/dʒ/	W	w	/w/
K	k	/k/	X	x	/ks/
L	l	/l/	Y	y	/j/
M	m	/m/	Z	z	/z/

Gambar 2. Huruf Indonesia dan penulisannya.

3. PERANCANGAN SISTEM

3.1 Analisis Masalah

Permasalahan yang diangkat dalam penelitian pembuatan aplikasi ini adalah menganalisis kondisi suatu gambar teks berbahasa Inggris yang diproses oleh program aplikasi sehingga dapat menghasilkan sebuah teks Inggris yang dikenali secara optimal dan dapat menghasilkan hasil terjemahan teks yang baik setelah diproses oleh mesin penerjemah.

3.2 Analisis Kebutuhan Non Fungsional

Analisis kebutuhan non fungsional meliputi dari perangkat keras, perangkat lunak, dan *user* (pengguna).

3.2.1 Analisis Kebutuhan Perangkat Keras

Kebutuhan perangkat keras yang diperlukan untuk membuat aplikasi ini adalah *smartphone* Android dengan spesifikasi sebagai berikut:

1. *Smartphone* Android dengan sistem operasi android minimal versi 4.0 (*Ice Cream Sandwich*), untuk aplikasi program di perangkat keras.
2. Kamera 5 MP dengan fitur autofokus, untuk pengambilan gambar teks secara langsung.
3. Android tipe layar sentuh, untuk pemotongan gambar teks secara manual.
4. Koneksi internet jaringan 3G, untuk koneksi internet yang menghubungkan antara aplikasi program dengan layanan penerjemah *microsoft translator*.

3.2.2 Analisis Kebutuhan Perangkat Lunak

Perangkat lunak yang perlu dipakai dalam pembuatan program aplikasi, diantaranya sebagai berikut:

1. Sistem operasi Windows 7 ultimate 32 bit, untuk menjalankan aplikasi Eclipse.
2. Eclipse Ide 3.7 sebagai perangkat lunak untuk membuat program aplikasi android.
3. Android SDK untuk perancangan, pembuatan, dan pengembangan aplikasi program.
4. Android *Development Tool* sebagai *plugin* untuk mengintegrasikan Android SDK dengan program eclipse.
5. Tesseract OCR untuk android sebagai aplikasi pengenalan gambar teks menjadi teks.
6. Microsoft translator sebagai mesin penerjemah teks berbahasa Inggris ke teks berbahasa Indonesia.

3.2.3 Analisis Pengguna

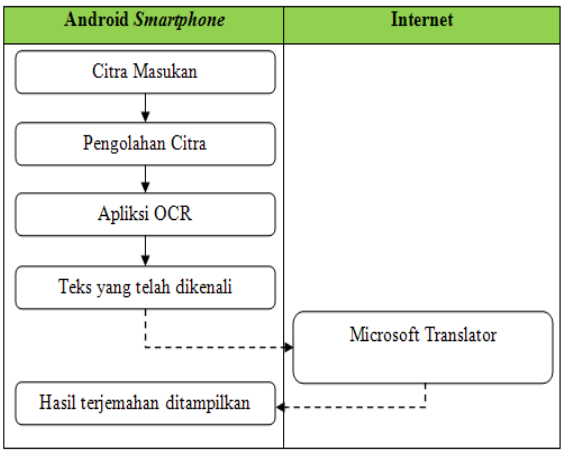
Aplikasi yang akan dibuat adalah aplikasi untuk para pengguna *smartphone* android yang ada di Indonesia, khususnya untuk para pembaca koran, majalah, atau buku teks berbahasa Inggris, karena aplikasi ini merupakan aplikasi yang berfungsi sebagai penerjemah gambar teks berbahasa Inggris yang diterjemahkan ke dalam Bahasa Indonesia.

3.3 Perancangan Sistem

Fungsi utama dari sistem aplikasi android yang akan dibuat ini adalah untuk melakukan proses pengolahan citra juga proses mengenali gambar teks Inggris menjadi hanya teks Inggris saja. Terjemahan teks Inggris sendiri memanfaatkan layanan penerjemah *microsoft translator*. Jadi tujuan utama dari aplikasi ini adalah mengolah citra teks berwarna menjadi citra teks biner dan

mengubah citra teks tersebut menjadi hanya teks Inggris saja yang selanjutnya teks tersebut diterjemahkan ke dalam bahasa Indonesia.

3.3.1 Perancangan Sistem Aplikasi Android



Gambar 3. Diagram blok perancangan sistem aplikasi android

Berikut ini penjelasan rancangan sistem aplikasi android:

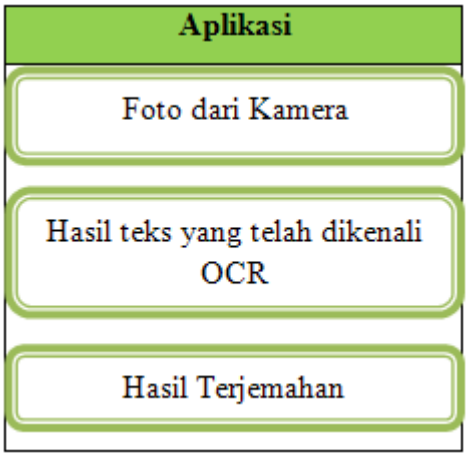
1. Citra masukkan merupakan citra teks Inggris yang akan diproses oleh aplikasi OCR. Citra ini dapat dari pengambilan foto secara langsung ataupun dapat dari menu di galeri *smartphone*.
2. Citra teks berwarna akan diproses menjadi citra biner dimana teks menjadi berwarna hitam dan latar belakang menjadi berwarna putih.
3. Citra biner lalu diproses di aplikasi OCR, yang akan mengubah citra teks menjadi teks saja.
4. Teks yang telah diolah di aplikasi OCR selanjutnya dikirmkan ke layanan penerjemah microsoft translator untuk diterjemahkan ke dalam teks Bahasa Indonesia.
5. Hasil terjemahan dikirim kembali ke aplikasi program yang selanjutnya menampilkan hasil terjemahan tersebut di aplikasi program.

3.3.2 Perancangan Struktur Pengolahan Citra

Rancangan struktur pengolahan citra menggunakan tiga teknik, Tujuan menggunakan tiga teknik ini untuk dapat menghasilkan citra teks yang dapat dikenali oleh aplikasi program OCR. Ketiga teknik yang akan digunakan, yaitu *cropping*, *grayscaling*, dan otsu *thresholding*.

3.3.3 Perancangan Antarmuka

Rancangan antarmuka untuk sistem aplikasi yang akan dibuat memiliki dua tampilan utama, yaitu menu utama dan menu tampilan gambar, yang di mana menu utama merupakan tampilan sumber foto yang akan diolah, sedangkanmenu tampilan gambar untuk menampilkan foto tersebut



Gambar 4. Antarmuka aplikasi Android

4. KESIMPULAN DAN SARAN

4.1 Kesimpulan

Dari hasil pengujian yang telah dilakukan, maka diperoleh kesimpulan bahwa program aplikasi yang dibuat telah berhasil dalam mengolah citra teks Inggris menjadi hanya teks Inggris saja yang kemudian teks Inggris diterjemahkan ke dalam teks Bahasa Indonesia. Dengan syarat dalam keadaan kondisi-kondisi sebagai berikut ini:

1. Gambar teks yang akan dikenali harus memiliki perbedaan warna yang jelas antara warna teks dengan warna latar belakang pada gambar teks.
2. Posisi pengambilan gambar teks dilakukan dengan pengambilan posisi dan kemiringan yang tegak lurus agar OCR dapat mengali teks dengan baik, jika tidak maka OCR tidak dapat mengenali gambar teks.
3. Cara pemotongan atau seleksi teks harus dilakukan dengan serapih mungkin agar teks dapat dikenali oleh OCR dengan optimal, sehingga menghasilkan teks yang jelas untuk diterjemahkan.
4. Teks yang akan diterjemahkan diusahakan memiliki ukuran *font* 12, dan *line spacing* minimal 1.5 spasi.
5. Penerjemahan teks hanya dapat menerjemahkan teks sebanyak satu kata, jika melebihi satu kata maka OCR tidak dapat mengenali teks dengan optimal.

4.2.1 Saran

Berdasarkan hasil pengujian dan kesimpulan , berikut ini beberapa saran untuk pengembangan aplikasi ini selanjutnya, diantaranya:

1. Membuat sistem penerjemahan teks sendiri yang sudah terintegrasi dengan program aplikasi.
2. Program aplikasi dapat menerjemahkan minimal 1 kalimat atau lebih.
3. Aplikasi OCR dapat mengenali teks lebih dari satu kata pada gambar teks yang diambil.

5. DAFTAR PUSTAKA

[1] Heriyanto. (2012). *Perancangan Aplikasi Pengolahan Citra Teks Arab dan Penerjemahannya ke Dalam Bahasa Indonesia Menggunakan Smartphone Android*. Universitas Komputer Indonesia, Bandung.

[2] Aplikasi OCR Tesseract untuk sistem operasi windows: http://code.google.com/p/tesseract-ocr/ diakses pada 13 Desember 2013. http://code.google.com/p/tesseract-ocr/downloads/detail?name=tesseract-ocr-3.02-win32-portable.zip&can=2&q

[3] Lee, Wei-Meng. (2012). *Beginning Android 4 Application Development*. Idianapolis: John Wiley & Sons, inc.

[4] Jackson, Wallace. (2013). *Learning Android App Development.* New York: Springer Science + Business Media.

[5] Mulyadi, ST. (2010). *Membuat Aplikasi Untuk ANDROID*, Yogyakarta: Multimedia Center Publishing.

[6] Gonzalez, R. C. dan Woods, R. E. (2002). *Digital Image Processing*. Upper Saddle River, NJ: Prentice Hall.