

**MODEL PENGENALAN TEKS LEMBARAN BERITA ACARA SERAH  
TERIMA INVENTARIS PADA SEKRETARIAT DPRD PROVINSI  
KEPULAUAN BANGKA BELITUNG**

**SKRIPSI**



**OLEH :**

**NUR AZIZAH**

**1511500081**

**PROGRAM STUDI TEKNIK INFORMATIKA  
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER**

**ATMALUHUR  
PANGKALPINANG**

**2019**

**MODEL PENGENALAN TEKS LEMBARAN BERITA ACARA SERAH  
TERIMA INVENTARIS PADA SEKRETARIAT DPRD PROVINSI  
KEPULAUAN BANGKA BELITUNG**

**SKRIPSI**

**Diajukan Untuk Melengkapi Salah Satu Syarat  
Memperoleh Gelar Sarjana Komputer**

**OLEH :**

**NUR AZIZAH**

**1511500081**



**PROGRAM STUDI TEKNIK INFORMATIKA  
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER  
ATMALUHUR  
PANGKALPINANG**

**2019**

## LEMBAR PERNYATAAN

Yang bertanda tangan dibawah ini:

NIM : 1511500081

Nama : Nur Azizah

Judul Skripsi : MODEL PENGENALAN TEKS LEMBARAN BERITA  
ACARA SERAH TERIMA INVENTARIS PADA  
SEKRETARIAT DPRD PROVINSI KEPULAUAN  
BANGKA BELITUNG

Menyatakan bahwa Laporan Tugas Akhir saya adalah hasil karya sendiri dan bukan plagiat. Apabila ternyata ditemukan didalam laporan Tugas Akhir saya terdapat unsur plagiat, maka saya siap untuk mendapatkan sanksi akademik yang terkait dengan hal tersebut.

Pangkalpinang, 3 Juli 2019



Nur Azizah

**LEMBAR PENGESAHAN SKRIPSI**

**MODEL PENGENALAN TEKS LEMBARAN BERITA ACARA SERAH  
TERIMA INVENTARIS PADA SEKRETARIAT DPRD PROVINSI  
BANGKA BELITUNG**

Yang dipersiapkan dan disusun oleh

**NUR AZIZAH  
1511500081**

Telah dipertahankan di depan Dewan Penguji  
Pada Tanggal 03 Juli 2019

**Anggota Penguji**



**Tri Sugihartono, M.Kom**

**NIDN: 0224129301**

**Kaprodi Teknik Informatika**



**R. Burham Isnanto F., S.Si, M.Kom**

**NIDN: 0224048003**

**Dosen Pembimbing**



**Yohanes Setiawan, M.Kom**

**NIDN: 0219068501**

**Ketua Penguji I**



**Fransiskus Panca Juniawan, M.Kom**

**NIDN: 0201069102**

Skripsi ini telah diterima dan sebagai salah satu persyaratan  
untuk memperoleh gelar sarjana komputer  
Tanggal 10 Juli 2019

**KETUA STMIK ATMA LUHUR PANGKALPINANG**



**Dr Husni Teja Sukmana, S.T., M.Sc**

**NIP: 197710302001121003**

## KATA PENGANTAR

Puji syukur Alhamdulillah kehadiran Allah SWT yang telah melimpahkan segala rahmat dan karuniaNya, sehingga penulis dapat menyelesaikan laporan skripsi yang merupakan salah satu persyaratan untuk menyelesaikan jenjang strata satu (S1) pada program Studi Teknik Informatika STMIK Atma Luhur.

Penulis menyadari bahwa laporan skripsi ini masih jauh dari sempurna. Karena itu, kritik dan saran akan senantiasa penulis terima dengan senang hati.

Dengan segala keterbatasan, penulis menyadari pula bahwa laporan skripsi takkan terwujud tanpa bantuan, bimbingan, dan dorongan dari berbagai pihak. Untuk itu, dengan segala kerendahan hati, penulis menyampaikan ucapan terima kasih kepada:

1. Allah SWT yang telah menciptakan dan memberikan kehidupan didunia.
2. Bapak dan Ibu tercinta yang telah mendukung penulis baik spirit maupun materi dan mendidik, mendukung penulis sepenuh hati.
3. Bapak Drs. Djaetun Hs yang telah mendirikan Atma Luhur.
4. Bapak Dr. Husni Teja Sukmana, S.T., M.Sc. selaku Ketua STMIK Atma Luhur.
5. Bapak R.Burham Isnanto Farid, S.Si., M.Kom. selaku Kaprodi Teknik Informatika.
6. Bapak Yohanes Setiawan, M.Kom. selaku dosen pembimbing.
7. Saudara dan sahabat-sahabatku terutama kawan-kawan Angkatan 2015 yang telah memberikan dukungan moral untuk terus menyelesaikan skripsi ini.

Semoga Tuhan Yang Maha Esa membalas kebaikan dan selalu mencurakan hidayah serta taufikNya, Amin.

Pangkalpinang, Juni 2019

Penulis



## **ABSTRACT**

*AI is a part of computer science that makes machines able to do work like humans. One of the artificial intelligence intelligence is object recognition. Object recognition is one application that is currently being developed and is an important application in image recognition. The most common application is OCR (Optical Character Recognition) which can read characters in the image. At the Secretariat of the Provincial Parliament of the Bangka Belitung Islands, many minutes of records are stored in the form of digital text and searches are returned for a long time. In dealing with these problems, a system is needed that can read and translate the contents of the minutes into digital form, namely the application of OCR. OCR writers consider to be able to support the smooth running of archiving activities at the Provincial Secretariat of the Bangka Belitung Islands DPRD office, especially regarding archiving the handover minutes of inventory. In making the OCR system a python-tesseract library is used. This study uses prototype models, object-oriented development methods, and UML as a system development tool. Based on the results of the research conducted, it can be concluded that the OCR system was successfully created. From the results of the OCR test on the report sheet, the handover of inventory concludes 98.80% - 99.52% of the text in the document can be identified.*

*Keyword: OCR, Arsip Digital, Tesseract*



## ABSTRAK

AI merupakan bagian dari ilmu komputer yang membuat mesin dapat melakukan pekerjaan seperti manusia. Salah satu kemampuan kecerdasan buatan adalah pengenalan objek (*object recognition*). Pengenalan objek merupakan salah satu aplikasi yang saat ini banyak dikembangkan dan merupakan aplikasi yang penting pada pengenalan citra. Aplikasi yang paling banyak dijumpai adalah OCR (*Optical Character Recognition*) yang dapat membaca karakter pada sebuah citra. Pada Sekretariat DPRD Provinsi Kepulauan Bangka Belitung banyak arsip berita acara yang mesti disimpan ke dalam bentuk teks digital dan pencarian kembali berita acara fisik membutuhkan waktu yang lama. Dalam mengatasi masalah tersebut diperlukan adanya sistem yang dapat membaca dan menetik secara otomatis isi dari berita acara tersebut ke dalam bentuk digital, yaitu penerapan OCR. OCR penulis anggap dapat menunjang kelancaran kegiatan pengarsipan pada kantor Sekretariat DPRD Provinsi Kepulauan Bangka Belitung terutama mengenai pengarsipan lembaran berita acara serah terima inventaris. Dalam pembuatan sistem OCR digunakan pustaka *python-tesseract*. Penelitian ini menggunakan model prototipe, metode pengembangan berorientasi obyek, dan UML sebagai alat bantu pengembangan sistem. Berdasarkan hasil penelitian yang dilakukan dapat disimpulkan bahwa sistem OCR berhasil dibuat. Dari hasil pengujian OCR pada lembaran berita acara serah terima inventaris menghasilkan kesimpulan 98,80% - 99,52% teks pada dokumen dapat dikenali.

Kata kunci: *OCR, Arsip Digital, Tesseract*



## DAFTAR ISI

	Halaman
<b>LEMBAR PERNYATAAN .....</b>	<b>i</b>
<b>LEMBAR PENGESAHAN .....</b>	<b>ii</b>
<b>KATA PENGANTAR.....</b>	<b>iii</b>
<b>ABSTRACT .....</b>	<b>iv</b>
<b>ABSTRAK .....</b>	<b>v</b>
<b>DAFTAR ISI .....</b>	<b>vi</b>
<b>DAFTAR GAMBAR .....</b>	<b>ix</b>
<b>DAFTAR TABEL .....</b>	<b>xi</b>
<b>DAFTAR SIMBOL .....</b>	<b>xii</b>
<b>BAB I PENDAHULUAN</b>	
1.1. Latar Belakang .....	1
1.2 Rumusan Masalah .....	5
1.3. Batasan Masalah .....	5
1.4. Tujuan dan Manfaat Penelitian .....	5
1.5. Sistematika Penulisan .....	6
<b>BAB II LANDASAN TEORI</b>	
2.1. Model Prototipe .....	7
2.1.1. Kelebihan dan Kekurangan Model Prototipe .....	7
2.1.2. Tahapan – Tahapan Prototipe .....	8
2.2. Metode Pemrograman Berorientasi Obyek .....	9
2.2.1. Konsep Dasar Metode Pemrograman Berorientasi Obyek .....	10
2.3. Alat Bantu Pengembangan Perangkat Lunak.....	11
2.3.1 UML (Unified Modelling Language) .....	11
2.3. Jenis – Jenis Diagram UML.....	12
2.4. Teori Pendukung .....	16
2.4.1 Artificial Intelligence .....	16



2.4.2. Pengenalan pola .....	17
2.4.3. Pengenalan Citra .....	18
2.4.4. Konversi Citra ke Teks.....	19
2.4.5. Pengubahan Citra Menjadi Karakter .....	19
2.4.6. OCR (Optical Character Recognition) .....	20
2.4.7. Tesseract .....	21
2.4.8. Python .....	22
2.4.9. OpenCV .....	23
2.4.10. Notepad ++ .....	23
2.4.11. Microsoft Visual Basic 6.0.....	23
2.4.12. Black Box Testing.....	24
2.5. Tinjauan Pustaka Terdahulu.....	25

### **BAB III METODOLOGI PENELITIAN**

3.1. Model Prototype .....	33
3.2. Metode OOP (Object Oriented Programming) .....	34
3.3. Alat Bantu Pengembangan Sistem.....	35

### **BAB IV HASIL DAN PEMBAHASAN**

4.1. Struktur organisasi .....	36
4.2. Tugas dan Wewenang .....	36
4.3. Analisis Masalah Sistem Yang Berjalan .....	37
4.4. Analisis Kebutuhan .....	38
4.4.1 Kebutuhan Fungsional .....	39
4.4.2 Kebutuhan Non Fungsional .....	39
4.5. Analisis Sistem .....	41
4.6. Analisis Kebutuhan Sistem Usulan .....	41
4.7. Perancangan Sistem .....	41
4.7.1 Activity Diagram Aplikasi Pengenalan Teks .....	42
4.7.2. Use Case Diagram Aplikasi Pengenalan Teks .....	43
4.7.3. Perancangan Antar Muka.....	44

4.7.4. Sequence Diagram Aplikasi Pengenalan Teks.....	45
4.8. Implementasi .....	46
4.8.1 Implementasi Perangkat Lunak.....	47
4.8.2. Implementasi Perangkat Keras.....	47
4.8.3. Tampilan Layar .....	47
4.9. Tesseract OCR Library.....	51
4.10. Arsitektur Tesseract .....	65
4.11. Pengujian.....	69
4.11.1. Pengujian Black Box.....	69
4.11.2. Pengujian Dalam Bentuk html .....	70
4.11.3. Pengujian Dalam Bentuk txt .....	74
4.11.4. Pengujian Dalam Bentuk docx .....	74
4.11.5. Pengujian Dalam Kesamaan Dokumen .....	75
4.11.6. Pengujian Kemiringan Dokumen .....	77
<b>BAB V PENUTUP</b>	
5.1. Kesimpulan .....	81
5.2. Saran .....	81
<b>DAFTAR PUSTAKA</b> .....	82

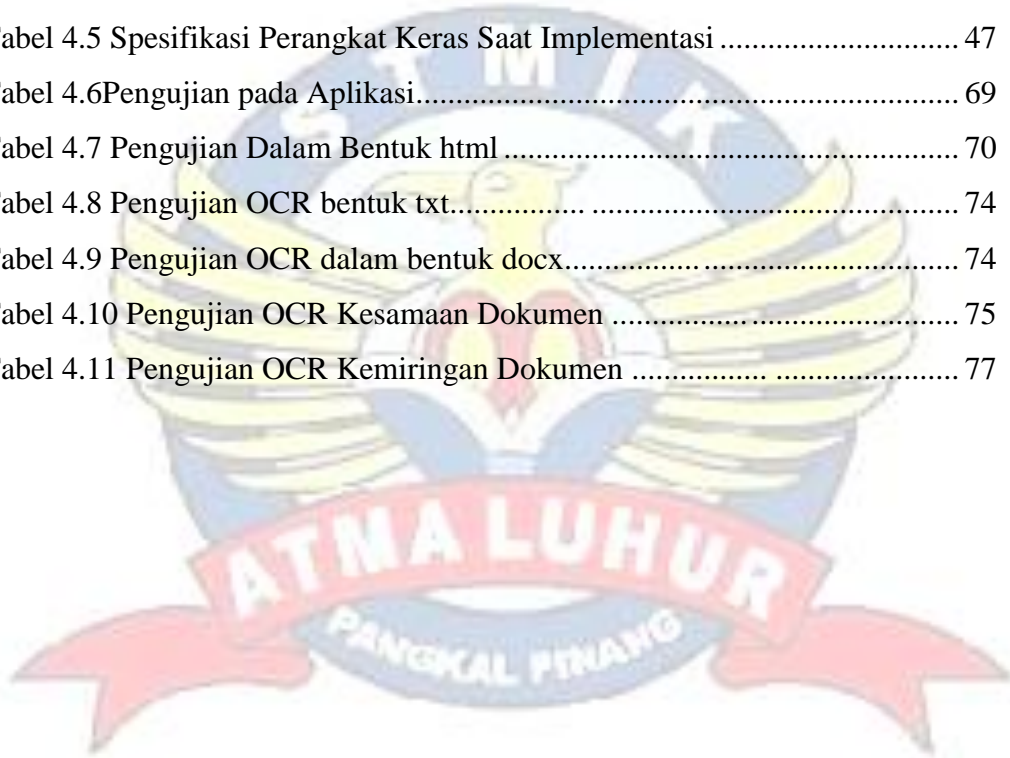
## DAFTAR GAMBAR

	Halaman
Gambar 2.1 Model Prototype .....	9
Gambar 2.2 Use Case Diagram.....	12
Gambar 2.3 Sequence Diagram.....	15
Gambar 2.4 Activity Diagram.....	16
Gambar 2.5 Metode Analisa Karakter .....	19
Gambar 2.6 Tahapan Perubahan Citra Menjadi Karakter.....	20
Gambar 3.1 Model Prototype.....	33
Gambar 4.1 Struktur Organisasi Sekretariat DPRD.....	36
Gambar 4.2 Sistem Berjalan.....	38
Gambar 4.3 Activity Diagram Form Utama .....	43
Gambar 4.4 Use Case Diagram Pengenalan teks .....	44
Gambar 4.5 Rancangan Antar Muka Form Utama .....	44
Gambar 4.6 Sequence Diagram Pengenalan Teks .....	46
Gambar 4.7 Tampilan Layar Form Utama.....	48
Gambar 4.8 Tampilan Open Image.....	49
Gambar 4.9 Tampilan Convert image .....	49
Gambar 4.10 Tampilan Path image dan Preview Image.....	50
Gambar 4.11 Tampilan proses Ocr .....	50
Gambar 4.12 Tampilan Hasil Keluaran Ocr .....	51
Gambar 4.13 Install Python.....	52
Gambar 4.14 Optional Python .....	53
Gambar 4.15 Advanced Options.....	54
Gambar 4.16 Setup Progress .....	54
Gambar 4.17 Setup Was Successful .....	55
Gambar 4.18 Command Prompt .....	56
Gambar 4.19 Pip Install Pipenv .....	56
Gambar 4.20 Upgrade Pip.....	57

Gambar 4.21 Menjalankan Perintah <code>mkdir ocr_server&amp;&amp;cd ocr_server&amp;&amp; pipenv install-th</code> .....	57
Gambar 4.22 Pipenv shell.....	58
Gambar 4.23 Install Pytesseract pillow.....	59
Gambar 4.24 Pip Install Pillow.....	59
Gambar 4.25 Install Pytesseract.....	60
Gambar 4.26 Install Opencv – contrib - Python.....	60
Gambar 4.27 Proses Awal Install.....	61
Gambar 4.28 License Agreement.....	62
Gambar 4.29 Choose Users.....	62
Gambar 4.30 Choose Install Locatin.....	63
Gambar 4.31 Choose Component.....	63
Gambar 4.32 Choose Start Menu Folder.....	64
Gambar 4.33 Installing.....	64
Gambar 4.34 Install Selesai.....	65
Gambar 4.35 Arsitektur Tesseract.....	65
Gambar 4.36 Baseline Fitting.....	67
Gambar 4.37 Chopping Joined Detection .....	68
Gambar 4.38 Asosiasi Karakter Penuh .....	68

## DAFTAR TABEL

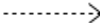

	Halaman
Tabel 2.1 Penelitian Terkait .....	29
Tabel 4.1 Kebutuhan Fungsional Form Utama .....	39
Tabel 4.2 Kebutuhan Perangkat Keras Pengembang .....	40
Tabel 4.3 Kebutuhan Perangkat Lunak Pengembang .....	41
Tabel 4.4 Spesifikasi Perangkat Lunak Saat Implementasi .....	47
Tabel 4.5 Spesifikasi Perangkat Keras Saat Implementasi .....	47
Tabel 4.6 Pengujian pada Aplikasi.....	69
Tabel 4.7 Pengujian Dalam Bentuk html .....	70
Tabel 4.8 Pengujian OCR bentuk txt.....	74
Tabel 4.9 Pengujian OCR dalam bentuk docx.....	74
Tabel 4.10 Pengujian OCR Kesamaan Dokumen .....	75
Tabel 4.11 Pengujian OCR Kemiringan Dokumen .....	77





## DAFTAR SIMBOL

### 1. Daftar Simbol Use Case Diagram

Gambar	Nama	Keterangan
	<i>Actor</i>	Menspesifikasikan himpunan peran yang pengguna mainkan ketika berinteraksi dengan <i>use case</i> .
	<i>Dependency</i>	Hubungan dimana perubahan yang terjadi pada suatu elemen mandiri ( <i>independent</i> ) akan mempengaruhi elemen yang bergantung padanya elemen yang tidak mandiri ( <i>independent</i> ).
	<i>Generalization</i>	Hubungan dimana objek anak ( <i>descendent</i> ) berbagi perilaku dan struktur data dari objek yang ada di atasnya objek induk ( <i>ancestor</i> ).
	<i>Include</i>	Menspesifikasikan bahwa <i>use case</i> sumber secara <i>eksplisit</i> .
	<i>Extend</i>	Menspesifikasikan bahwa <i>use case</i> target memperluas perilaku dari <i>use case</i> sumber pada suatu titik yang diberikan.
	<i>Association</i>	Apa yang menghubungkan antara objek satu dengan objek lainnya.



*System*

Menspesifikasikan paket yang menampilkan sistem secara terbatas.



*Use Case*

Deskripsi dari urutan aksi-aksi yang ditampilkan sistem yang menghasilkan suatu hasil yang terukur bagi suatu aktor.



*Collaboration*

Interaksi aturan-aturan dan elemen lain yang bekerja sama untuk menyediakan perilaku yang lebih besar dari jumlah dan elemen-elemennya (sinergi).

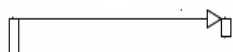


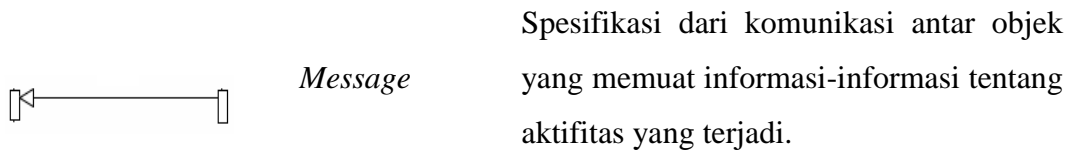
*Note*

Elemen fisik yang eksis saat aplikasi dijalankan dan mencerminkan suatu sumber daya komputasi.

## 2. Daftar Simbol Sequence Diagram

Gambar	Nama	Keterangan
	<i>LifeLine</i>	Objek <i>entity</i> , antarmuka yang saling berinteraksi.
	<i>Message</i>	Spesifikasi dari komunikasi antar objek yang memuat informasi-informasi tentang aktifitas yang terjadi.





### 3. Daftar Simbol Activity Diagram

Gambar	Nama	Keterangan
	<i>Activity</i>	Memperlihatkan bagaimana masing-masing kelas antarmuka saling berinteraksi satu sama lain.
	<i>Action</i>	<i>State</i> dari sistem yang mencerminkan eksekusi dari suatu aksi.
	<i>Initial Node</i>	Bagaimana objek dibentuk atau diawali.
	<i>Final Node</i>	Bagaimana objek dibentuk dan diselesaikan.
	<i>Fork Node</i>	Satu aliran yang pada tahap tertentu berubah menjadi beberapa aliran.