BUSINESS REQUIREMENT DOCUMENT

OCR Based Form Filling

Document Information

|  |  |
| --- | --- |
| Document Attributes | Description |
| Document ID | Doc001 |
| Author | Sandeep Taksande |
| Contributor | - |

Revision History

|  |  |  |
| --- | --- | --- |
| Version | Date | Changes |
| Draft | 01-October-2024 | New Document |
| V0.1 |  |  |
|  |  |  |

Revision History

|  |  |  |  |
| --- | --- | --- | --- |
| Reviewer/ Approver | Role | Comments | Date |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

Contents

[1. Introduction 4](#_Toc179189942)

[1.1 Purpose 4](#_Toc179189943)

[1.2 Scope 4](#_Toc179189944)

[2. Business Objectives 4](#_Toc179189945)

[3. As-Is Situation 4](#_Toc179189946)

[4. To-Be Situation 4](#_Toc179189947)

[5. Key Stakeholders 4](#_Toc179189948)

[6. Project Constraints 5](#_Toc179189949)

[7. Cost-Benefit Analysis 5](#_Toc179189950)

[7.1 Costs 5](#_Toc179189951)

[7.2 Benefits 5](#_Toc179189952)

[8. Functional Requirements 5](#_Toc179189953)

[8.1 PDF Document Upload 5](#_Toc179189954)

[8.2 OCR Processing 5](#_Toc179189955)

[8.3 Data Extraction and Mapping 6](#_Toc179189956)

[8.4 Form Filling 6](#_Toc179189957)

[8.5 Error Handling 6](#_Toc179189958)

[9. Non-Functional Requirements 6](#_Toc179189959)

[9.1 Performance 6](#_Toc179189960)

[9.2 Security 6](#_Toc179189961)

[9.3 User Interface 6](#_Toc179189962)

[9.4 Compatibility 6](#_Toc179189963)

[10. Success Criteria 7](#_Toc179189964)

# 1. Introduction

1.1 Purpose

This document outlines the business requirements for a web application that utilizes Optical Character Recognition (OCR) to extract data from PDF documents and automatically populate corresponding fields in an online form.

## 1.2 Scope

The scope of this project includes:

* Developing a web application capable of processing PDF documents.
* Implementing OCR functionality to extract text from the PDF documents.
* Mapping extracted data to relevant fields in the online form.
* Automating the filling of the form based on the extracted data.

# 2. Business Objectives

* Streamline the data entry process by eliminating manual input.
* Improve accuracy and reduce errors associated with manual data entry.
* Enhance efficiency and productivity.
* Provide a user-friendly interface for interacting with the application.

# 3. As-Is Situation

* **Current Process:** Manual data entry from PDF documents into an online form.
* **Challenges:**
  + Time-consuming and error-prone
  + Low efficiency and productivity
  + Potential for inconsistencies and inaccuracies

# 4. To-Be Situation

* **Desired Outcome:** Automated data entry from PDF documents into an online form.
* **Benefits:**
  + Significant reduction in manual data entry time and effort.
  + Improved data accuracy and consistency.
  + Enhanced efficiency and productivity.
  + Streamlined workflow and reduced operational costs.

# 5. Key Stakeholders

* **Department Heads:** Finance, Human Resources, Operations
* **Data Entry Staff:** Individuals responsible for manual data entry
* **IT Department:** Responsible for application development and maintenance
* **End-Users:** Individuals who will benefit from the automated data entry process

# 6. Project Constraints

* **Budget:** Allocated budget for project development and implementation.
* **Timeline:** Deadline for project completion and deployment.
* **Technical Constraints:** Existing IT infrastructure and compatibility requirements.
* **Regulatory Compliance:** Adherence to data privacy and security regulations.

# 7. Cost-Benefit Analysis

## 7.1 Costs

* Development costs
* Hardware and software requirements
* Training and support
* Ongoing maintenance and updates

## 7.2 Benefits

* Time savings from reduced manual data entry
* Improved data accuracy and reduced errors
* Increased efficiency and productivity
* Reduced operational costs
* Enhanced customer satisfaction

# 8. Functional Requirements

## 8.1 PDF Document Upload

* **The application should allow users to upload PDF documents from their local devices.**
* **The application should support various PDF file formats and versions.**

## 8.2 OCR Processing

* **The application should utilize a reliable OCR engine to accurately extract text from the PDF documents.**
* **The OCR engine should be capable of handling complex layouts, fonts, and images.**
* **The application should provide options for adjusting OCR parameters to optimize accuracy for different document types.**

## 8.3 Data Extraction and Mapping

* **The application should identify and extract relevant data fields from the PDF documents based on predefined templates or rules.**
* **The extracted data should be mapped to corresponding fields in the online form.**
* **The application should handle variations in data formats and layouts within the PDF documents.**

## 8.4 Form Filling

* **The application should automatically populate the online form fields with the extracted data.**
* **The application should validate the extracted data to ensure accuracy and consistency.**
* **The application should provide options for manual review and editing of the extracted data before filling the form.**

## 8.5 Error Handling

* **The application should handle potential errors during OCR processing, data extraction, or form filling.**
* **The application should provide informative error messages to guide users in resolving issues.**

# 9. Non-Functional Requirements

## 9.1 Performance

* **The application should process PDF documents efficiently, minimizing processing time.**
* **The application should be scalable to handle increasing workloads.**

## 9.2 Security

* **The application should implement appropriate security measures to protect user data and prevent unauthorized access.**
* **The application should comply with relevant data privacy regulations.**

## 9.3 User Interface

* **The application should provide a user-friendly and intuitive interface.**
* **The application should be accessible to users with disabilities.**

## 9.4 Compatibility

* **The application should be compatible with various web browsers and operating systems.**

# 10. Success Criteria

* The application accurately extracts data from PDF documents and populates the online form fields.
* The application significantly reduces manual data entry time and effort.
* The application improves data accuracy and consistency.
* The application provides a positive user experience.
* The application meets the project's budget and timeline constraints.
* The application complies with relevant regulatory requirements.