Software Requirements Specification

for

Form Recognition Engine (FRE)

Version 1.1

Prepared by Sohaib Altaf

Syslab Technologies Sdn Bhd

December 23, 2020

Revision History

Version	Primary Author(s)	Description of Changes	Date Completed
1.0	Sohaib Altaf Raja	SRS for FRE (New Artifacts)	December 23, 2020
1.1	Sohaib Altaf Raja	SRS for FRE (Minor Updates)	December 31, 2020

Introduction

FCE is a web-based application to recognize and capture information from different types of forms such as purchase orders and invoices automatically with minimal human intervention using the OCR (Optical Character Recognition) and the Machine Learning techniques. This system is designed for extracting data from raw files like images or pdf. Using OCR, it extracts the data from the file, and by applying machine learning techniques, it recognizes different information present in the files and process it. It then compiles the data and sends it back as a response to the user who uploaded it.

Purpose

This SRS document describes the user requirements for the development of the "Form Recognition Software''. The document will assist the audience further their understanding regarding the user requirements and business processes involved. It will explain the purpose and features of the FRE, the interfaces of the FRE, what the FRE will do, the constraints under which it must operate and how the FRE will integrate with new and existing applications. This document is intended for both the stakeholders and the developers of the FRE. Finally, this document will serve as a guideline during the design and development stage of the FRE.

Intended Audience

The intended users of this artifact will be the customer and software team of SYSLAB Technologies Sdn Bhd

Definitions, Acronyms and Abbreviations

Term	Description
SRS	Software Requirements Specification
FRE	Form Recognition Engine
OCR	Optical Character Reader
SMTP	Simple Mail Transfer Protocol

References

1. AI-OCR requirements.pdf
2. Templates for Purchase Orders and Invoices

System Interfaces Specification

External Interface Requirements

1. Send in Blue will be used for sending emails
2. 3rd Party System will access FRE through an API call

Internal Interface Requirements

Interface	Users
Admin	Admin will have access to all modules of the system.

System Functional Requirements

Form Recognition Engine

Upload Form to Process (via our Internal System Interface)

Brief Description	Admin wants to extract the processed data from FORM by uploading it (via our Internal System Interface). FORM includes purchase order and invoice.
List of Actors	System
Preconditions	1. Admin is logged-in into the system. 2. Sample template FORM for invoices and purchase orders are already created 3. 20 different types of FORM templates for invoices and purchase order are already trained in order to fulfill the training requirements as per the assumptions 4. All fields are mapped with correct labels
Post Condition	1. Raw data is extracted and converted into processed/refined

Form Recognition Engine (FRE)

	data
	2. Processed data is saved
	3. New logs added
Basic Flow	1. Admin will upload the FORM using the PDF/PNG format only
	2. System will validate the format and proceed if it is correct
	3. System will extract the raw data from the FORM and will convert it into processed/refined form.
	4. System will save the converted data and will sent the refined data in JSON format, in case of success
	5. Email is sent to admin along with the error in case of failure
	6. In case of either FORMS' label is not recognized due to limitations of training data or if threshold level is below 60 % then an alert email is sent to admin along with the error.
	7. System will add the logs

Upload Form to Process (via 3rd Party System)

Brief Description	Admin wants to extract the processed data from FORM by uploading it (via 3rd Party System using an API call). FORM includes purchase order and invoice.
List of Actors	3rd Party System
Pre-conditions	1. The 3rd Party system has valid API credentials.
	2. Sample template FORM for invoices and purchase orders are already created
	3. 20 different types of FORM templates for invoices and purchase order are already trained in order to fulfill the training requirements as per the assumptions
	4. All fields are mapped with correct labels
Post Condition	1. Raw data is extracted and converted into processed/refined data
	2. Processed data is saved & sent in response in JSON format.
	3. Email is sent to admin along with the error in case of failure
	4. New logs added
Basic Flow	1. 3rd Party system will call API by providing API credentials, form to process and type of the form (invoice or purchase order)
	2. System will authenticate the API call token. In case it is a valid token, then system will proceed else return message of Invalid API Call'
	3. System will check form format, if it is one of the given formats e.g. pdf or png then system will proceed to process

	the request. 4. OCR will extract the raw data from the FORM and will apply machine learning algorithm to improve data accuracy. 5. System will then send back the refined data in JSON format 6. In case of either FORMS' label is not recognized due to limitations of training data or if threshold level is below the required accuracy (set in system) then an alert email is sent to admin along with the error. 7. System will add the logs

Template management

Create/Setup New Template for FORMS

Brief Description	Admin want to setup a new template for FORMS (Model Templates) to process in order to extract raw data
List of Actors	Admin
Preconditions	1. Admin is logged-in into the system. 2. FORM categories for invoices and purchase orders are defined in the system.
Post Condition	1. A new template is added to the system 2. Sample FORMS are added and associated with a new template 3. Labels are saved and assigned to sample FORMS (mapped with fields of sample FORMS) 4. System is trained with sample FORMS 5. New logs added
Basic Flow	1. Admin will navigate to the template management section in order to create/setup a new template (Model Template) for Invoices/Purchase Orders 2. Admin will add the title of template which is a mandatory information 3. For training requirements, admin will add/associate sample FORMS with a new template and will add/define labels (order number, customer name) for example, 2 FORMS for purchase orders of PEPSI, 4 FORMS for purchase order of Mineral Water etc. For sufficient training purpose, admin should associate multiple sample FORMS because they may have variations 4. In order to proceed for training, admin will select any sample FORM and will add/assign the labels one by one to each

field of the sample FORM, for example, admin will map label (order number) with field (order no./order ref./order serial) of purchase order FORM (PEPSI)

5. Admin has to provide the location of the specific field to system so that system can the read/get its coordinates

6. Therefore, admin will drag the pointer/index towards the location of the particular field having text/word, so that system can read the coordinates of the selected text/word

7. Then system will prompt for user input, so that user can add specific label for selected text/word/line

8. Admin can also add one or more keyword against that label for example, inv no, invoice-no, invoice key to INVOICE NUMBER

9. System will extract all the fields of sample FORM and will save the labels against each specific field

10. System will also save the user provided keyword against the newly defined label, so all labels will be mapped with all fields of sample FORMS.

11. System will show the list of extracted fields/information along with the assigned labels so that admin can validate or make necessary correction. Admin can edit at this stage if any field/information is not mapped correctly

12. For sufficient training purpose, admin will add multiple sample FORMS, once added he will start the training session so that machine can learn

13. System will prompt for user intervention in case confidence level is low.

14. Admin can again start the training session by adding more samples. This accuracy threshold will be configured from the system settings

15. In case, system already contains the repository of sample FORMS for whom labels are already mapped/assigned, then system will show the list of these labels so that user can select among these.

16. Once validated, system will save the repository of selected sample FORMS, otherwise Admin will reassign/update the labels and resubmit the form

17. System will add the logs

Example of FORM Fields (to map labels with these fields)

FORM: Invoice

Fields for header: Invoice Date, Invoice Number, PO Reference number, Customer Code, Customer Name,

Fields for body: Item code, Item Description, Qty, Unit Price, Amount \ Net Amount, Total Amount (checksum again total of line Amount\Net Amount)

Update Template

Brief Description	SU wants to update an existing template.
List of Actors	Admin.
Preconditions	1. Admin is logged-in into the system. 2. Categories for invoices and purchase orders are defined in the system. 3. Labels & keywords (i.e. purchase order no, invoices no etc.) are already defined under the existing templates
Post Condition	1. Existing templates are updated along with its name, labels, keywords 2. New sample FORMS are added and mapped with existing template 1. New labels are assigned to sample FORM 2. New keyword is assigned to labels. 3. System is trained with new sample FORMS 4. New logs added
Basic Flow	1. Admin will select specific template in order to update the existing template 2. Admin will follow flow of previous use case/feature Create/Setup New Template for FORMS" from point no. 4 3. System will add the logs

Delete Template

Brief Description	SU wants to delete an existing template.
List of Actors	Admin.
Preconditions	1. Admin is logged-in into the system.
Post Condition	1. Existing template is removed from system 2. New logs added
Basic Flow	1. Admin will select specific template in order to delete it 2. System will add the logs

API Access management

Generate Token

Brief Description	Admin wants to generate API tokens, so that 3rd Party System can use them while calling FRE APIs.
List of Actors	Admin
Preconditions	1.Admin is logged-in into the system.
Post Condition	1. New API token is added 2. New logs added
Basic Flow	1. Admin will navigate to API token generation section from the settings 2. Admin will select option of new key generation 3. System will generate a unique token key and show it in a message box 4. Admin can copy it in order to send to any recipient 5. System will add the logs

Update Token

Brief Description	Admin wants to update the token access
List of Actors	Admin
Preconditions	1.Admin is logged-in into the system.
Post Condition	1. API token is updated 2. New logs added
Basic Flow	1. Admin will navigate to API token generation section from the settings 2. Admin will select specific API token and update its token key 3. System will generate a unique token key 4. System will add the logs

Block Token

Brief Description	Admin wants to block the token access
List of Actors	Admin
Preconditions	1.Admin is logged-in into the system.
Post Condition	1. API token access is blocked 2. New logs added
Basic Flow	1. Admin will navigate to API token generation section from the settings 2. Admin will select specific API token access and block it 3. System will block access to the API for that Token. 4. System will add the logs

System User & Settings

Add System User

Brief Description	Admin wants to add a new system user.
List of Actors	Admin
Preconditions	1. Admin is logged-in into the system.
Post Condition	1. A new system user is added in the system. 2. New logs added
Basic Flow	1. Admin will navigate to 'Add System User' option from the settings 2. Admin will add user credentials and system will save the information 3. System will add the logs

Delete System User

Brief Description	Admin wants to delete a particular system user.
List of Actors	Admin
Preconditions	1. Admin is logged-in into the system.

Post Condition	1. Existing system user is deleted from the system 2. New logs added
Basic Flow	1. Admin will navigate to 'Delete System User' option from the settings 2. System will show list of system users where admin will select specific system user for delete 3. System will add the logs

Configuration

Brief Description	Admin wants to add and update various parameters for system configuration
List of Actors	Admin
Preconditions	1. Admin is logged-in into the system.
Post Condition	1. Configuration information is update 2. New logs added
Basic Flow	1. Admin will navigate configurations from the settings 2. Admin will define threshold for accuracy level, so that system will send an alert email if accuracy level is below the threshold level 3. Admin will add one or more email accounts 4. System will add the logs

Reports

Brief Description	Admin want to see various reports of unsuccessful attempts so that he can intervene and take prompt action
List of Actors	Admin
Preconditions	1. Admin is logged-in into the system.
Post Condition	NA
Basic Flow	1. Admin will navigate to reporting section from Settings

	2. Admin can view various reports accuracy level of which are below the defined threshold

View System Logs

Brief Description	Admin wants view system logs
List of Actors	Admin
Preconditions	1. Admin is logged-in into the system.
Post Condition	NA
Basic Flow	1. Admin will navigate to the system logs section from Settings 2. System will show list of all logs ordered by latest date which contains all system user operations and transactions 3. Admin can search any specific information from search bar 4. Admin can open any specific log 5. System will show user name, date/time, IP address, information before and after action such as delete, update etc., name of file/table that is affected

Constraints

- Document to be uploaded must be in English, system should reject any other language e.g. Document/Text in Chinese language will be rejected.
- Processing speed will be 10 pages per second.
- System will process forms in different batches.
- Only JOSN will used for data exchange format

Assumptions

- Enough sample should be available for training requirements
- Template (Model Form) is correctly labeled by the admin
- System will used for two categories of FORMS only i.e. Invoices and Purchase orders

Form Recognition Engine (FRE)

- Admin intervention will be required in case accuracy is below the threshold level
- System will recognize the fields of any new FORM based on the trained labels and associated keywords
- Information should be clear in FORMS and readable
- Enough sample are provided for trainings requirements

Milestones

First Milestone ()

In the first milestone, we will give an option to upload forms (known & unknown both). System will process it and:

- Processing accuracy for all those forms which are new will be 75%.
- Processing accuracy for all those forms which are already in the system will be 90%

Second Milestone ()

In the second milestone, we will give an option to upload forms (known & unknown both) via the admin panel. System will process it and:

- Processing accuracy for all those forms which are new will be 80%.
- Processing accuracy for all those forms which are already in the system will be 95%

The Admin Panel with everything excluding template management will also be ready by this time.

Third Milestone ()

- Processing accuracy for all those forms which are new will be 90%.
- Processing accuracy for all those forms which are already in the system will be 100%
- Admin Panel 100% completed

Form Recognition Engine (FRE)