| **Business Requirement**  **Document** |
| --- |

| u **Client:**  REQUORDIT Inc.  u **Delivered on:** 12th Apr, 2023  u **Submitted by:** Anish Ahmed  u **Contact Info:** anish.ahmed@amygb.ai /91 903 646 1276 |
| --- |

**Contents**

1. Introduction

2. Purpose of the Document

3. Scope of Work

4. Reports and Accuracy Logics

5. Billing Logic

6. Technical Infrastructure Requirement Details [PROD & UAT]

**1.** **Introduction**

The main purpose of this project is to provide an automated invoice processing system “**IDP (Intelligent Document Processing)”** to ensure smooth, quick and efﬁcient services to REQUORDIT by AmyGB.ai. As compared to other mechanism of invoice processing, IDP will ensure that the requests, which falls under the scope can be addressed within the agreed and definite TAT committed to the customer. This will help the REQUORDIT team to reduce the TAT for document processing and have a higher productivity in their day to day operations.

| **Particulars** | **Details** |
| --- | --- |
| **Initial requirement carried out on-** | **14th Sep 2022** |
| **Requirement carried out by-** | **Akanksha Singh** |
| **REQUORDIT SPOC Name-** | **Lennin Mendoza** |
| **SPOC Contact Details-** | **lennin.mendoza@requordit.com** |
| **Industry / Domain-** | **Document Management/OCR** |

**2.** **Purpose of the Document**

The purpose of the document is to list down and document each and every necessary steps like Current Flow, Proposed Flow, Scope of the Project, Technical Infra Details and Report Mechanism.

This Document is prepared on the basis of the discussions carried out by various discussions between REQUORDIT and AmyGB Team.

**3.** **Requirement Scope**

An application to be developed by AmyGB to Automate the invoices wherein After the invoice is uploaded via API, Application will carry out the OCR, extract the defined key values and get the review done by the indexers and post review, make the file available to download in the PDF/tiff format.

An avg. of 8000 documents will get processed/day except month end wherein this may reach up to 20000 (last couple of days of the month). Each doc will contain an average of 5 pages. 80% of the docs for the day are uploaded between 6AM -8AM CDT which to be processed by noon and output will be downloaded by 6 PM CDT. Processing shall happen on AWS - on cloud. Extracted data is deleted (OCR- text + image) post 23 Hrs(From the time file is downloaded) and retain the meta data (which include the kvp, line items, doc class etc.)

**Below are the discussed scoped items to be delivered in the phase wise manner-**

1. Classification/Re-Classification

2. Extraction & Review

3. Supervisor QC model

4. User & Team Management

5. PDF & TIFF Download options

6. Rotation of a file

7. File merging and Re-ordering

8. File lock and release

9. Tabular Content Extraction + Draw Table Bounds only for tabular failures

10. Tabular OCR Review with merging/splitting of rows on UI

11. Real time generation of reports of various doc lifecycle

12. Feedback Module

**The following is the High Level process flow-**

**4.** **Scope of Work**

The Scope of work for the project will be divided into three parts:

I. Phase 1 Scope

II. Phase 2 Scope

III. Phase 3 Scope

**I.** **Phase 1 Scope:**

Features to be considered for Phase 1 are as follows based on the discussions with the REQUORDIT team.

1. Classification & Extraction Review

2. Supervisor QC model

3. User & Team Management

4. PDF & TIFF Download options

5. Rotation of a file

6. File merging and Re-ordering

7. File lock and release

8. Real time generation of reports

**Phase 1** will consist scenarios related to the above categories described as below

l **Classification/Re-Classification & Extraction Review-**

A document should go through the OCR, and get classified as per the standard Invoice Or Supporting Document. Post Classification, if we realize there was error in classification, it should have option to re-classify and submit classification. Once classification is done, 16 defined key values with mandatory and optional fields to be extracted and reviewed by the indexers.

l **Supervisor QC model-**

Also review can be done by the supervisor and admin as part of Supervisor QC review process wherein set % of the particular customer id based on the team it belongs to, files will flow to mapped supervisor for the QC

l **User & Team Management**-

This will have 3 roles namely as 1) Indexer 2) Supervisor 3) Admin

Sup and Admin role will be allowed the addition, update and mapping of the users to customer IDs and supervisors

l **PDF & TIFF Download options-**

Once the review is completed and submitted from both indexers and supervisors (in 100% QC model), file should be available to be downloaded as PDF/Tiff for the end customer

l **Rotation of a file-**

If a document is skewed as an input, while performing the classification, it should rotate the image to 90 Degree or more and same should get downloaded as an output post completion of the Classification and Extraction

l **File merging and Re-ordering-**

If an invoice is available in file1 out of multiple files in the same batch and Supporting Document of the respective invoice is available in file2, we should have option to merge the SD from File2 with Invoice in file1 and vice versa. This should have option of re-ordering of the page count with respect to the changes

l **File lock and release-**

This feature will function as, once a file is opened by any indexer, it will be locked for 30 minutes and no one else can work on that file. Also, if an indexer opens multiple tabs or try to work on multiple files together, it will release the older files and allow to work on the latest file only to avoid the ambiguity

l **Real time generation of reports of various doc lifecycle-**

There should be an option available which can be used to generate the reports on the real time for the document processing stats from Admin panel

Note- After the Go Live, the timeline for the Phase 1 hyper-care stage is 75 days where the AmyGB team will analyze the accuracy on the area of opportunities and fix the bugs comes on the way

**II.** **Phase 2 Scope:**

Features to be considered for Phase 2 are as follows based on the discussions with the REQUORDIT team.

1. Table extraction spread across multiple pages within same/different files

2. Draw table bounds if table missed or incomplete table identified

3. Addition of rows at end of the table (new rows or missing row)

4. Merge/split/editing of rows of the table

5. QC of the extracted tabular data/line items with option to pick values via snippet

**Phase 2** will consist scenarios related to the above categories described as below

l **Table extraction spread across multiple pages within same/different files-**

This feature should allow the data to be extracted from the tabular invoices across multiple pages within same/different files

l **Draw table bounds if table is missed or incomplete -**

Draw table bounds if table is missed or incomplete table identification for reprocessing via OCR

l **Addition of rows-**

This feature will allow to add of rows at end of the table (new rows or missing row)

l **Merge/split/editing of rows of the table-**

This feature should allow to merge/split/Edit the table rows where it is needed

l **QC of the extracted tabular data/line items -**

This feature should allow to do the QC of the extracted tabular data/line items with option to pick values via snippet

Note- After the Go Live, the timeline for the Phase 2 to start is after 75 days, the requirements for the Phase 2 will commence.

**III.** **Phase 3 Scope:**

Features to be considered for Phase 3 are as follows based on the discussions with the REQUORDIT team.

1. Feedback Module

**Phase 3** will consist scenarios related to the above categories described as below

l **Feedback Module-**

This feature should allow the application to accept the feedback bases the inputs from the users. This will allow Key - Value Feedback - Instead of manual corrections on all key value extraction give feedback on selected failed files. The USP of this feature is that it also retrains the doc extraction module so that similar errors are not repeated in future batches

**5.** **Reports and Accuracy Logics**

This section will give a clarity on the Mechanism of Reports for Uploaded and Processed Data.

After we go live with the solution in place, we shall generate the reports for document life cycle which should have visibility on page level details of the Invoices and supporting documents for billing purpose.

l **Monthly Billing Reports:** Monthly billing reports will be created for billing purpose. This report will contain all the invoices along with page detail processed by the IDP for a specific month. This report will be used by AmyGB for all the invoicing purpose

**6.** **Billing Logic**

The billing will be based on the total number of pages of the Invoices processed by the application. Commercials are closed over other email thread.

**7.** **Technical Infrastructure Requirement Details. [PROD & UAT]**

This section will list down all the technical level infrastructural details for Production Server that will be necessary for this project. The scope of the project will require technical infrastructure at the REQUORDIT UAT server and also at the Production server.

| **Production Server Resources and OS Details** | | | |
| --- | --- | --- | --- |
|  | **Platform** | **AI** | **OCR** |
| Server counts | 1 | 2 | 2 |
| Purpose of the Server | Production | Classification + Extraction + Image/PDF Processing | OCR (Original Text + Coordinates) |
| ( Ex:- Production – Preprod – UAT – SIT – VDI - Others ....) |  |  |  |
| Role of the Server | Web + Application + DB | Classification + Extraction + Image/PDF Processing | OCR (Original Text + Coordinates) |
| ( Ex:- Web - Application – Database – Scheduler – VDI – Others... ) |  |  |  |
| Server Operating system | Ubuntu | Ubuntu | Ubuntu |
| ( Ex:- Windows – Linux – Ubundu – AmazonLinux - Others ....) |  |  |  |
| Server Operating system version | 22.04 | Ubuntu 22.04.1 LTS (GNU/Linux 5.19.0-1022-aws x86\_64) | Ubuntu 22.04.1 LTS (GNU/Linux 5.19.0-1022-aws x86\_64) |
| ( Ex:- WIN2016, WIN2019, RHEL7.9, RHEL8.1, AML2, Others ) |  |  |  |
| CPU Count in VCPU | 4 | 4 | 4 |
| Memory in GB | 8 | 16g RAM + 100G Disk Space | 8g RAM + 100G Disk Space |
| Server Usage ( Ex:- 24/7, 24/5, 18/5, 16/5 .....) | 24/7 | 24/7 | 24/7 |
| Server Live Timings ( Ex:- MON - FRI 07:00 AM to 11:00 PM ) if not 24/7 |  |  |  |

**Infrastructure for UAT/Staging Server for UAT Testing:**

After the training of the Application, the codes will be deployed on the UAT server of the REQUORDIT. It is important for the UAT server infrastructure to be similar to PROD server, as all the initial level testing will get carried out on the UAT server. UAT server should also have the similar API’s as the PROD server for the AmyGB to carry out the UAT level testing for the same.

Note- Any changes, bug fix or new requirements to be tested on Staging first before deploying it to Prod to have the no impact on the Prod environment

| **UAT/ Staging Server Resources and OS Details** | |
| --- | --- |
| Server counts | 1 |
| Purpose of the Server | UAT Server |
| ( Ex:- Production – Pre-Prod – UAT – SIT – VDI - Others ....) |  |
| Role of the Server | Hosting Platform WebApp, AI WebApp, MongoDB |
| ( Ex:- Web - Application – Database – Scheduler – VDI – Others... ) |  |
| Server Operating system | Ubuntu - 18+ |
| ( Ex:- Windows – Linux – Ubundu – AmazonLinux - Others ....) |  |
| Server Operating system version | same as above |
| ( Ex:- WIN2016, WIN2019, RHEL7.9, RHEL8.1, AML2, Others ) |  |
| CPU Count in VCPU | 4 |
| Memory in GB | 8 |
| Server Usage ( Ex:- 24/7, 24/5, 18/5, 16/5 .....) | 24 hours a day x 365 days a year |
| Server Live Timings ( Ex:- MON - FRI 07:00 AM to 11:00 PM ) if not 24/7 |  |

THANK YOU