

Keep your documents safe,
organized, and easily retrievable

Malaf Archive

User manual

SteadyPoint

CONTENTS

Document Information.....	2
Document Purpose.....	2
Malaf Overview	2
Getting Started with malaf archive	2
Logging in.....	2
Add a new document library.....	3
Connect to document library	3
Content type	4
Settings.....	5
General settings	5
Scanner settings	9
Sync settings.....	10
OCR templates.....	12
Validation rules.....	14
Batch profiles	15

DOCUMENT INFORMATION

DOCUMENT PURPOSE

This training manual is focused on providing information and hands-on steps to assist end-users in using "**Malaf Archive**" Application.

MALAF OVERVIEW

MALAF; which means file in Arabic, is a solution for capturing documents to the Office 365 SharePoint Online. It is a desktop application, which is integrated with office 365, WIA, and TWAIN scanners

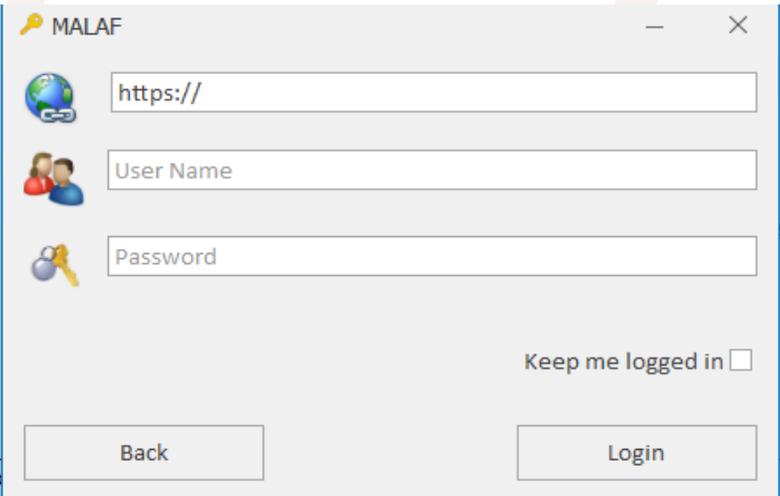
The process of archiving documents goes through five stages:

1. **Capture**
Capture images from scanner or hard disk
2. **Extract**
Extract data from images using OCR and Barcode technologies
3. **Validate**
Validate extracted data against predefined rules
4. **Separate**
Consolidate images to output files depending on predefined rules
5. **Classify**
Do auto indexing and classification for output files
6. **Migrate**
Migrate output files to SharePoint online with corresponding meta data

GETTING STARTED WITH MALAF ARCHIVE

LOGGING IN

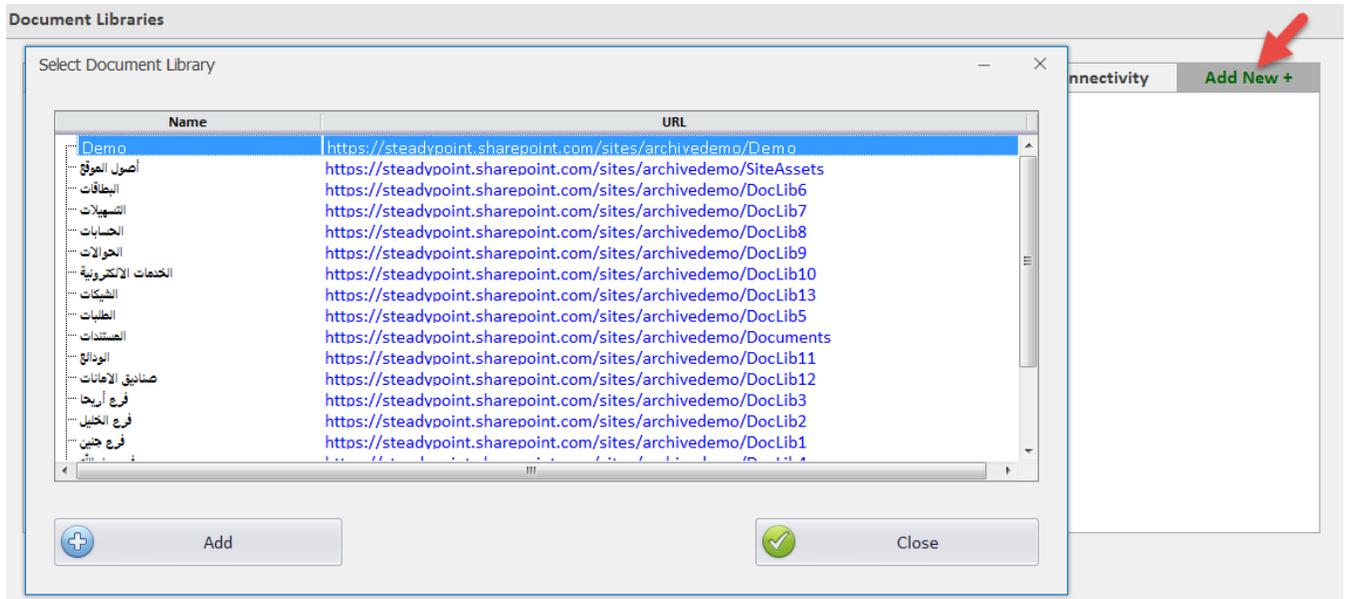
The first step you should do after installing the application is to log in by entering the link of the site collection that contains the document library you want to use for archiving, and logging in using your organization email and password.



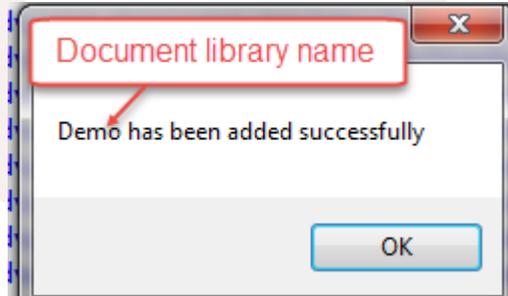
The screenshot shows a login window titled "MALAF". It contains three input fields: "https://" with a globe icon, "User Name" with a person icon, and "Password" with a key icon. There is a "Keep me logged in" checkbox and two buttons: "Back" and "Login".

ADD A NEW DOCUMENT LIBRARY

Click on **Document libraries** tab, click on **Add New +**, list of all document libraries in entered site collection will appear in tree view, choose document library you want and click **Add**.



Message will appear defining that adding the document library has been done successfully



After adding the document library click **Close**.

CONNECT TO DOCUMENT LIBRARY

After adding the document library, you need to connect to it by clicking on **Connect** beside document library name in Document libraries tab.

MALAF Settings ? DemoUser

No document library connected [Connect Now](#)

- Settings
- Document Libraries**
- Content Types
- OCR Templates
- Validation Rules
- Batch Profiles

Document Libraries

Name	URL	Status	Connectivity	Add New +
Demo	https://SteadyPoint.sharepoint.com/sites/ArchiveDemo/Demo	Not Connected	Connect	-

Note: status before connecting will be **Not Connected**, after connecting status will be changed to **Connected**.

Document Libraries

Name	URL	Status	Connectivity	Add New +
Demo	https://SteadyPoint.sharepoint.com/sites/ArchiveDemo/Demo	Connected	Disconnect	-

✔ **Notification**
 Now you can upload files
 Please click File Explorer button

After connecting to document library, notification will appear modifying that you can upload files.

To delete document library, click on (-).

CONTENT TYPE

Content types for selected document library will appear in **content types** tab, you can enter default values for each column in the content type.

To see the columns of the content type, click on (+) beside the content type name.

Content Types

	Content Type	Add New +
▶ (+)	مستند	-
▶ (+)	سجل	-

To delete content type, click on (-).

SETTINGS

You can modify the application settings using settings tab which includes three types of settings: **General settings**, **Scanner settings**, **Sync settings**.

GENERAL SETTINGS

The screenshot shows the 'Settings' dialog box with the 'General Settings' tab selected. The dialog has three sub-tabs: 'General Settings', 'Scanner Settings', and 'Sync Settings'. The 'General Settings' tab contains the following options:

- Use Separation:** A checkbox that is currently unchecked. Below it are three radio buttons: 'Separator Sheet' (selected), 'Metadata Separation', and 'Per Number of Pages'. To the right of these radio buttons are two buttons: 'Print Separator' and 'Separation Rules'. Below the 'Separation Rules' button is a small numeric input field containing the number '1'.
- Output Format:** A dropdown menu currently set to 'PDF'.
- Correct Skew:** An unchecked checkbox.
- Correct Binarization:** An unchecked checkbox.
- Enable OCR:** An unchecked checkbox.
- Default OCR Template:** A dropdown menu.
- Default Content Type:** A dropdown menu currently set to 'مستند' (Document).
- Document Name Prefix:** A text input field containing 'Demo'.
- Show parameters Dialog:** A checked checkbox.

USE SEPARATION

In General Settings you can enable use separation, and choose what type of separation you want. MALAF supports multiple methods to separate multiple individual papers and combine them into one or more documents. Use of the appropriate method depends on the following factors:

- Structure of scanned templates.
- Number of scanned templates in one batch.
- Number of pages per template.
- Quality of text in scanned images.

Below are the separation methods supported by Malaf:

- Separator sheet
- Barcode separation
- Metadata separation
- Per number of pages

Separator sheet



- Separator Sheet is a predefined page, which contains a recognizable barcode by **Malaf** engine. Recognizing this sheet would inform MALAF that this is the start of a new document.
- This method can be used in case of handwritten or very low-quality text that cannot be recognized by OCR.
- This method contains options to save predefined data in QR code, with mapping matrix to file metadata

Choosing separator sheet method will enable print separator icon to print, or preview the separator sheet with or without QR code.

Print Separator
— X

Use QR Code

Available Templates New Template

Template Name Demo_14032018153124_سجل

Document Library https://steadypoint.sh...

Content Type سجل

Column Name	Data
العنوان	
نوع الطلب	
رقم الحساب	
رقم الفرع	
الموظف المسؤول	
التعميل	

Print

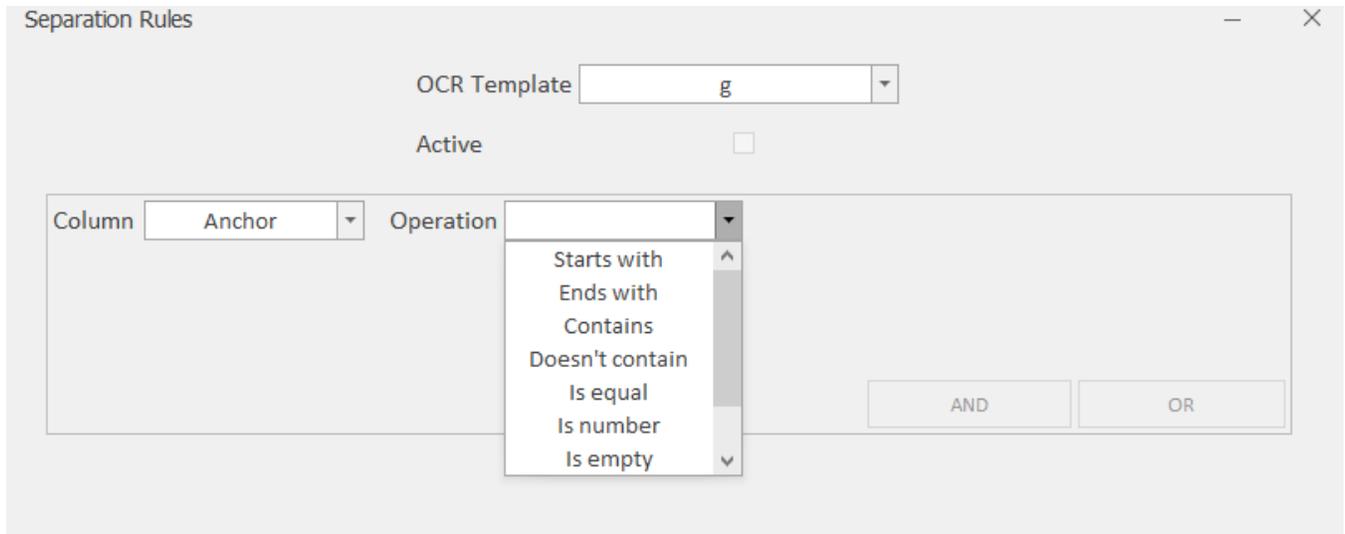
Preview

Save

Metadata separation

Malaf engine can separate documents depending on metadata recognized by OCR or barcodes printed on the scanned documents. You can also define the separation rules, based on metadata with mixed conditions; such as *account field*. *Account field* has a specific number of characters that ranges from 5 to 7. Accordingly, if this rule is applicable, the engine would proceed with the separation.

To define a separation rule, click on Separation rules button beside metadata separation, window like below will appear.



You should select the OCR template then select the column that you will use for separation, we prefer to use Anchor value for separation.

You can combine two rules in the same column by adding AND or OR. Click Save to close the window.

Rule must be active to be used for separation, so make sure to check the Active checkbox

If you want to not use the rule just uncheck the box.

Per number of pages

This method separates the documents following fixed number of pages. For instance, if we separate every **3** pages, it means that MALAF will create a document every 3 pages. This is a good option for archiving documents with the same number of pages.

OUTPUT FORMAT

Output format is the format of output documents, for now we only support PDF Format.

CORRECT SKEW

One of the most common issues that affect document quality is **Skew**. Images might contain skewed content or might get skewed during the scanning process. To extract content, the skew must be corrected. Enabling **correct skew** will correct both horizontal and vertical skews, by checking skew angel and rotate the image accordingly.

CORRET BINARIZATION

Binarization is the process when removing noise from the captured texts. The high quality binarized image can give more accuracy in character recognition; as compared to the original image, since noise is present in the original image. Enabling **correct binarization** removes image noise by applying binarization process before feeding the images to OCR.

DEFAULT CONTENT TYPE

Choose default content type to use for scanned documents from list of content types in connected document library.

DOCUMENT NAME PREFIX

Type in the document name prefix which will appear in document name after scanning.

SHOW PARAMETER DIALOG

This checkbox will enable the appearance of parameters dialog one more time when scanning

SCANNER SETTINGS

Settings

General Settings **Scanner Settings** Sync Settings

Use Scanner Machine

Scanner HP ScanJet Pro 3000 s3 [Refresh](#)

Protocol Twain

Source

Auto Rotate

Remove Blank Pages

Use Duplex

Scanning Resolution

Output Resolution

Output Type Colored

Page Size A4

Use a specific Folder [Browse](#)

You have two options to scan from: Use scanner machine, or use a specific folder.

USE SCANNER MACHINE

choose the name of the scanner you want to scan images from, click on **Refresh** to get the name of defined scanners. After clicking on the name of the scanner the supported features (auto rotate, remove blank pages, use duplex, scanning resolution and output resolution which must be equal or larger than scanning resolution) will start appearing According to the specifications of the scanner.

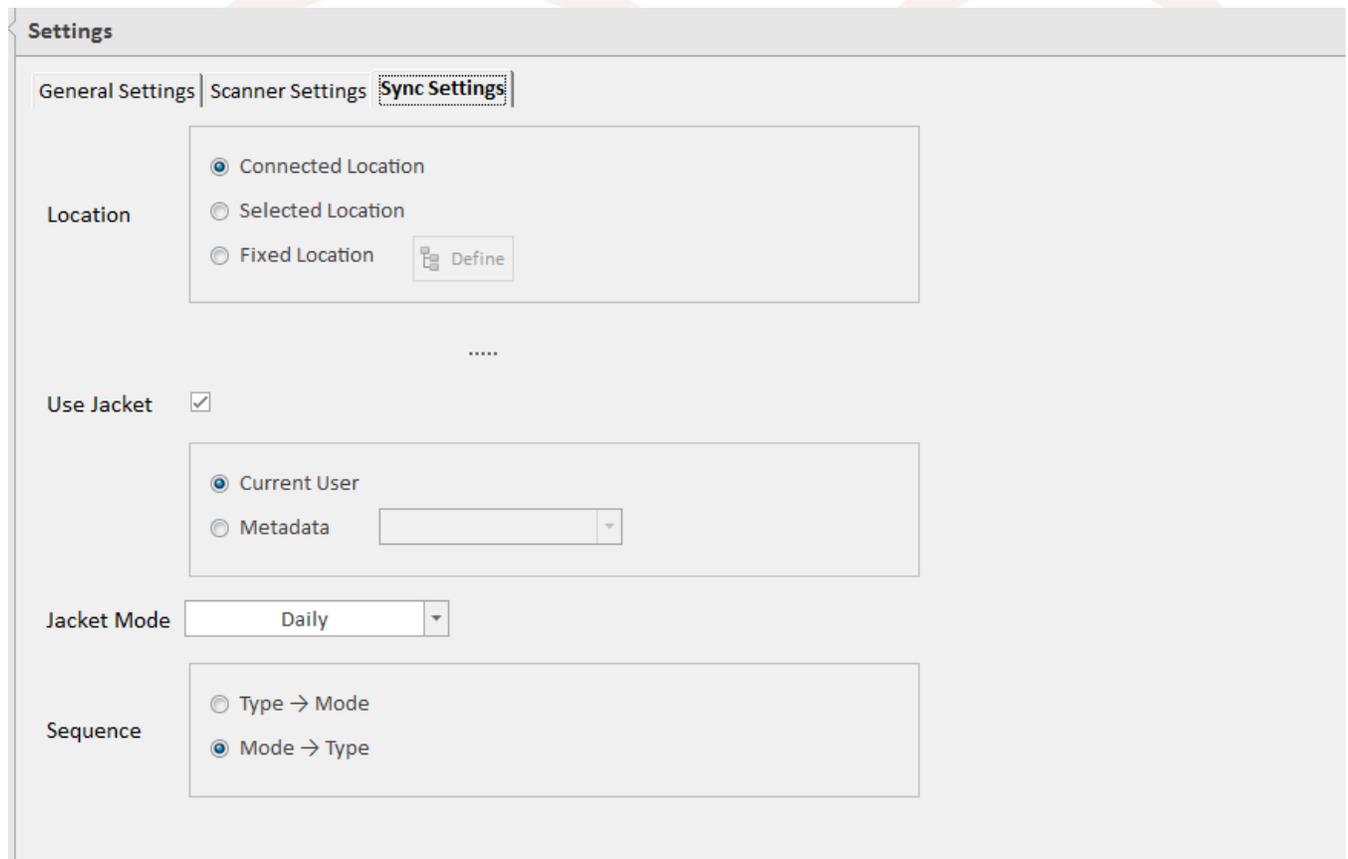
Output type of scanned documents could be colored or black and white.

Page size the page size of scanned document, for now only A4 supported.

USE SPECIFIC FOLDER

If you have a soft copy of the documents you want to scan, you can put them in a folder and browse it to use it instead of scanning them. Same setting will be applied to soft copy documents as hard copy.

SYNC SETTINGS



The screenshot shows a 'Settings' window with three tabs: 'General Settings', 'Scanner Settings', and 'Sync Settings'. The 'Sync Settings' tab is active. It contains the following sections:

- Location:** Three radio buttons: 'Connected Location' (selected), 'Selected Location', and 'Fixed Location'. A 'Define' button is next to 'Fixed Location'.
- Use Jacket:** A checked checkbox.
- Current User:** A radio button (selected) and a dropdown menu labeled 'Metadata'.
- Jacket Mode:** A dropdown menu showing 'Daily'.
- Sequence:** Two radio buttons: 'Type → Mode' and 'Mode → Type' (selected).

Syncing/Migrating to Office 365 has different settings might be modified according to your needs.

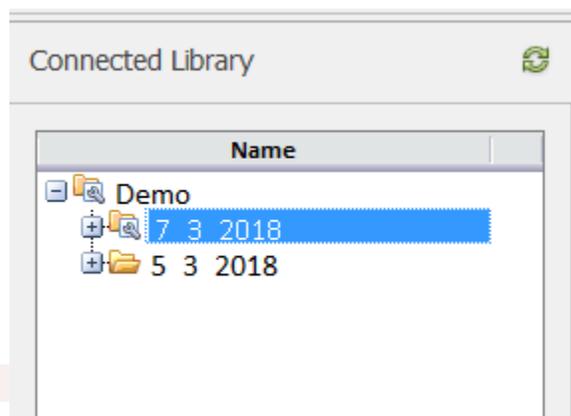
LOCATION

There are three locations you can choose from

Connected location: which means documents will be synced directly to connected document library

Selected location: which means documents will be synced to selected folder from the tree of folders appears in File explorer.

Fixed location: you can choose a specific fixed location will always be used for syncing documents.



USE JACKET

You can use Jacket to keep archived documents aggregated and organized, distinguish data on geographical and managerial basis per area/employee/department/branch ...etc. every document data will be well organized for farther approval, auditing, and archiving activities on different time basis (daily, monthly, yearly).

We support two types of jackets:

Current user

If you use this type, folder with the name of the current user will be created in chosen location.

Metadata

If you use this type, you need to define the column you want to use in order to create the folder according to its value, for example if you choose the account number, folder will be created for each account number in chosen location.

The jacket modes we support are four modes:

One Time

If you choose this mode the folder will be created one time only.

Daily

If you choose this mode, the folder will be named as current date, i.e. DD-MM-YYYY.

Monthly

If you choose this mode, the folder will be named as current Month, i.e. MM-YYYY.

Yearly

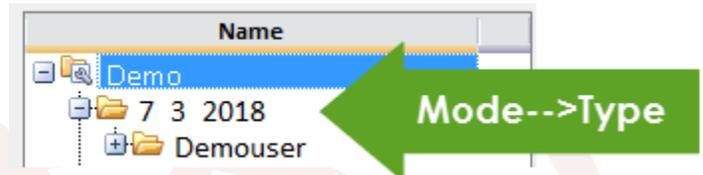
If you choose this mode, the folder will be named as current Year, i.e. YYYY.

Sequence

Choosing the sequence is important to organize the sequence of creating the jacket folders,

for example, if you want the documents to be uploaded in user name first then the date choose **type→ mode**

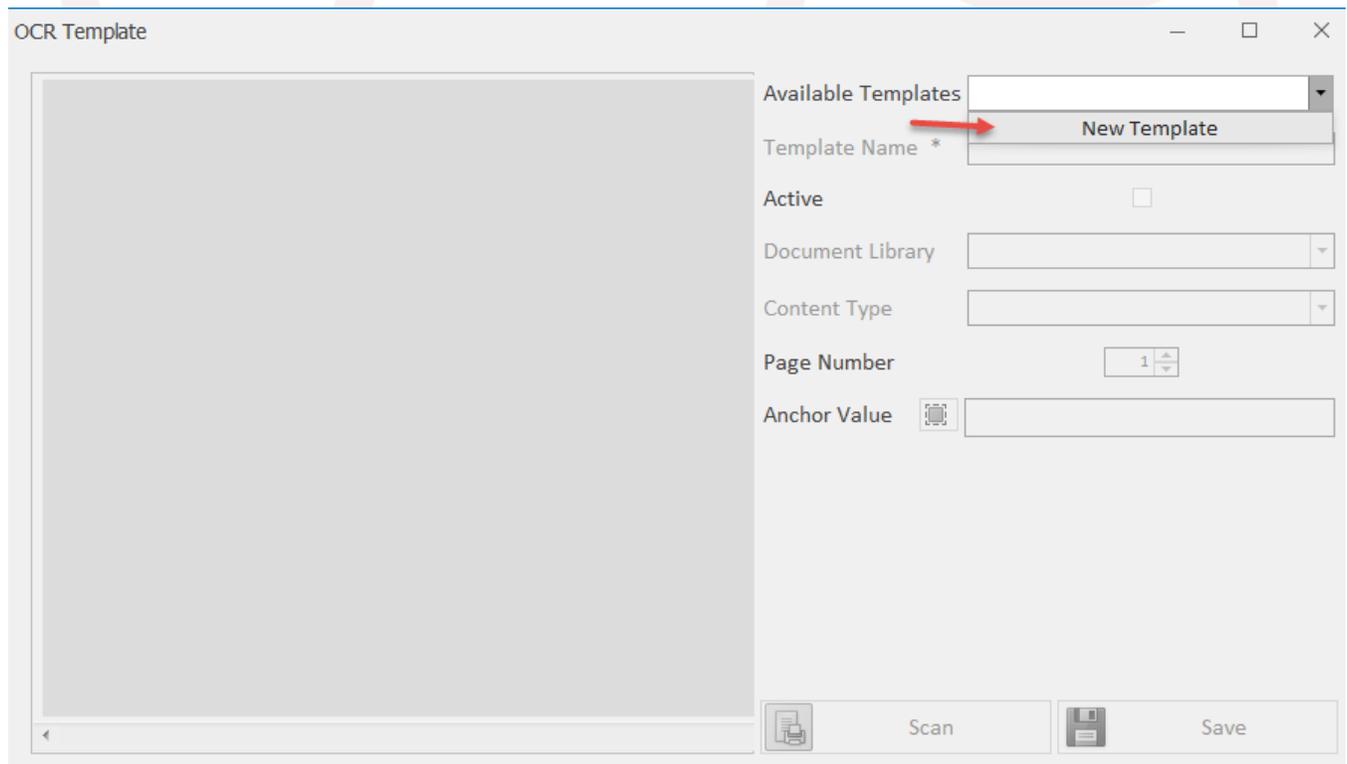
But if you want the folder of the date to be created first then the jacket type (user or metadata) then choose **mode→type**.



OCR TEMPLATES

MALAF uses OCR engine to extract texts from scanned images. It supports Arabic, English, and mixed languages

Here where you can define OCR templates, in order to recognize texts of created zones inside images to be used for data extraction.



When you Click on **New Template**, the **Scan** button start blinking to let you know that now you can scan the template you want to create.

After scanning, image will appear inside the grey area.

Sample Employment Contract

نوع الطلب

Document is entered into between Mohammad Khaled Abu Gh
nd ATCS Company (hereinafter
under the terms and conditions of employment be
ffective from

Column Name	Data
العنوان	
نوع الطلب	
رقم الحساب	
رقم الفرع	
الموظف العضو	
التسجيل	

Enter the **template name**, then click on **Active** checkbox which means this OCR template is active to let OCR engine use this template when extracting data.

After that choose the **document library** and **content type** related to this OCR template so you can create the zones for needed columns.

To create a new zone, click on the Square icon  for standard zone such as text or number, or on combo box icon  for values in one row table.



If you click on the combo box icon, a dialog will appear (see below image) to choose the language and the number of digits you want to read, for example account number is 7 digits, choose 7 and click OK.

Language English

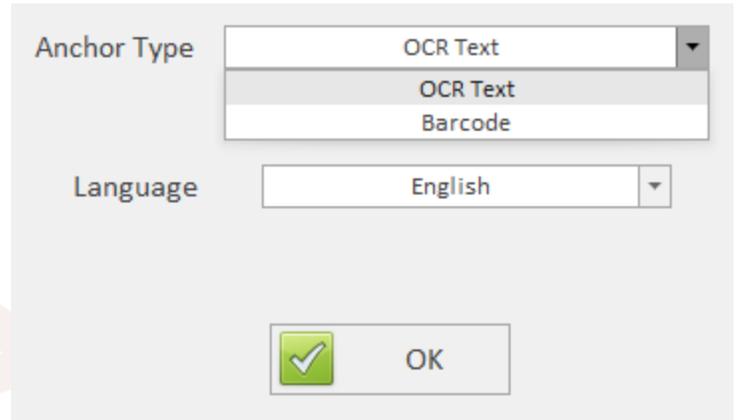
of digits 1

OK

Put the zone between the numbers and make sure the numbers are centralized between the edges of the squares. After you put the zone you can click on the Eye icon  to see the value the OCR engine has read which appears beside the column name.

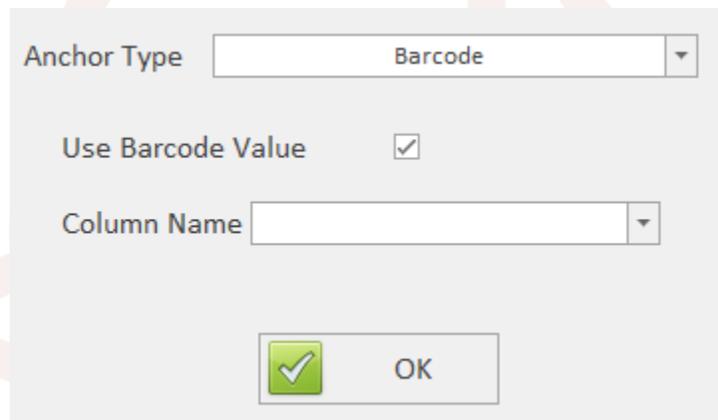
The **Page number** means in which page to use this OCR template, for example if you want to use this template in the second page of each document choose number 2.

Anchor value is the most important zone you should define, because all other zones will be related to anchor to get the coordinates for more accuracy.



Anchor Type: OCR Text
Language: English
OK

Anchor should be a unique value so you can use it for separation, it could be an OCR text or Barcode, if you choose the OCR text you need to select the language. If you choose the Barcode you have an option to use Barcode Value and then define the column you want to put the value on .



Anchor Type: Barcode
Use Barcode Value:
Column Name:
OK

After creating the template and defining the zones, click **Save**, Window will not close so you can create another template at the same time.

VALIDATION RULES

MALAF contains a validation module to validate the quality of extracted data and manually filled data using two techniques.



Column Name	Column Type
-------------	-------------

Add New +

To set up a validation rule click on **Add New +**, Dialog will appear to define the document library and content type so you can select the column you want to set a validation rule

Validation Rules

Document Library

Content Type

Column

Active

After choosing the column, select the operation and the operand that meet your needs, for example if the account number can be from 5 to 7 digits you can set a validation rule : Account number # of characters from 5 to 7 , see below image.

Column Operation From To

You can combine two rules in the same column by adding AND or OR, click Save to close the window.

To edit the rule, click on  icon, to delete it click on .

Rule must be active to be validated, so make sure to check the Active checkbox

BATCH PROFILES

If you have more than one user, and each user archive different documents, you can create batch profile for each user and add the OCR templates for each of them, also you can change the general settings and sync settings for each batch profile.

To add new batch profile, click on **Add New +**, window will open. Select the document library and the content type you want to use for this batch profile, then enter the batch profile name.

Batch Profiles

Batch Profile Name	<input type="button" value="Add New +"/>

In each batch profile you will see four tabs:

1. General settings

Here where you can change the general settings of the batch profile which is exactly the same as general settings in the main page.

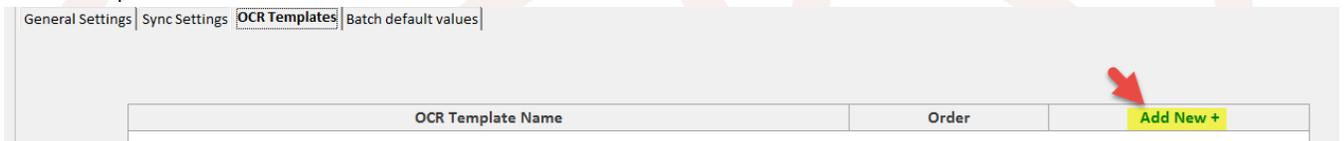
You can use separation method and enable OCR templates if you want to extract data, note that you must define the OCR templates first so you can use it inside the batch profiles.

2. Sync settings

Sync settings are the same as the sync settings in the main page. Each batch profile can sync in different location, mode and type.

3. OCR Templates

Here where you can add a new OCR templates to be used in this batch profile. To add a new OCR template, click on **Add New +**, list of all OCR templates will appear, select the OCR template name and click Select.



Then you should give an order for each OCR template so that searching for separator will be easier and faster.

Click on  to delete the OCR template.

4. Batch Default values

Clicking on the batch default values tab will show the columns in selected content type.

You have an option to enter default values especially for the batch profile, for example you can put the name of the employee as a default value for Employee name column.

----- End of Document -----