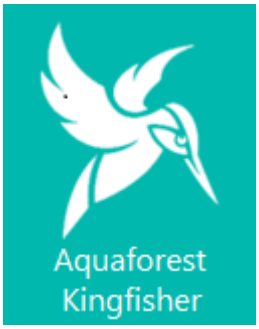
Kingfisher Reference Guide



Version 2.1
June 2019

© Aquaforest Limited 2001-2019
Web: www.aquaforest.com
E-mail: info@aquaforest.com

Aquaforest

CONTENTS

1 INTRODUCTION 3

1.1 INSTALLATION 3
1.2 SYSTEM REQUIREMENTS 3
1.3 LICENCING AND PURCHASING 3
1.4 VERSION BREAKDOWN 4

2 KINGFISHER WINDOWS INTERFACE 5

2.1 DASHBOARD 5
2.2 JOB SETTINGS 6
2.2.1 Job Definition 7
2.2.2 Location Settings 8
2.2.3 Select Variables 12
2.2.4 Scheduler 17
2.2.5 Alerts 18
2.2.6 File Naming 19
2.2.7 Advanced Settings 21
2.2.8 Running Job 25
2.3 OPTIONS 26
2.3.1 License Keys 26
2.3.2 Email 26
2.3.3 Theme 27
2.3.4 Advanced Settings 27
2.4 HELP 28

3 KINGFISHER OPERATIONS 29

3.1 RENAME PDF FILE 29
3.1.1 Rename Based on Barcode 30
3.1.2 Rename Based on Text Match 30
3.1.3 Rename Based on Barcode and Text Match 30
3.2 EXTRACT PAGES FROM PDF FILE 31
3.2.1 Extract Pages from File based on Text Match 31
3.2.2 Extract Pages from File Name based on Barcode 32
3.3 SPLIT PDF FILE 33
3.3.1 Split File based on Text Match 33
3.3.2 Split File based on Text Change 34
3.3.3 Split File based on Barcode 35
3.3.4 Split File based on Page Ranges 36
3.4 EXTRACT CONTENT FROM PDF FILE 37
3.4.1 Extract Text from PDF File 37
3.4.2 PDF to CSV/XLSX 37
3.4.3 Advanced Export to CSV/XLSX 37

4 CREATING AND RUNNING JOBS WITH THE GUI 38

4.1 CREATING A NEW JOB 38
4.2 EDITING A JOB 47

5 EXTRACT TABULAR DATA FROM PDF 48

5.1 THE DOCUMENT VIEW 49
5.2 TABLE VIEW 51
5.2.1 Tables 51
5.2.2 Grouped Tables 52
5.2.3 Data grid 55

6 ADVANCED EXPORT TO CSV/XLSX 56

7 CUSTOM SCRIPTS 60

7.1 CUSTOM SCRIPT ARGUMENTS 60
7.2 ENABLING CUSTOM SCRIPTS 61
7.3 CUSTOM SCRIPTS SAMPLES 61
7.3.1 archiveinputs.wsf 61
7.3.2 stamp.wsf 62
7.3.3 converttoxml.wsf 62

8 XML JOB DEFINITION FILES 63

9 USING KINGFISHER FROM THE COMMAND LINE 66

9.1 USING JOB DEFINITION FILE 66

10 AUTOBAHN DX INTEGRATION 66

11 KINGFISHER FILES AND DIRECTORIES 66

12 SUPPORT 67

13 ACKNOWLEDGEMENTS 67

1 INTRODUCTION

Kingfisher provides a set of operations for processing PDF files, this includes splitting, extracting pages, extracting text and renaming of the files. Kingfisher also provides the ability to set the security, metadata and other attributes of a PDF document.

What makes kingfisher standout is the ability to perform split, extract and rename operations based on text or barcode found in zones or coordinates in a PDF page. Kingfisher can use OCR (Recognise text in image PDF files) to extract text data from Image PDF files or pages.

1.1 Installation

The installation MSI will guide you through the process of setting up Kingfisher.

1.2 System Requirements

Supported Operating Systems	Windows 8.1 (64bit) Windows 10 (64bit) Windows Server 2008 (64bit) Windows Server 2012 Windows Server 2016
Supported SharePoint Versions	SharePoint Online (office 365) SharePoint Server 2016 SharePoint Server 2013 SharePoint Server 2010
Disk Space	260 Mb
.Net Framework	4.6.1
Visual C++ Redistributables	Visual C++ 2010 Visual C++ 2013 (x64) Visual C++ 2013 (x86)
Additional tools	SharePoint Server Client Components SDK (x64)

1.3 Licencing and Purchasing

Trial licenses are fully functional, but all generated PDF files will have the Aquaforest trial stamps on them.

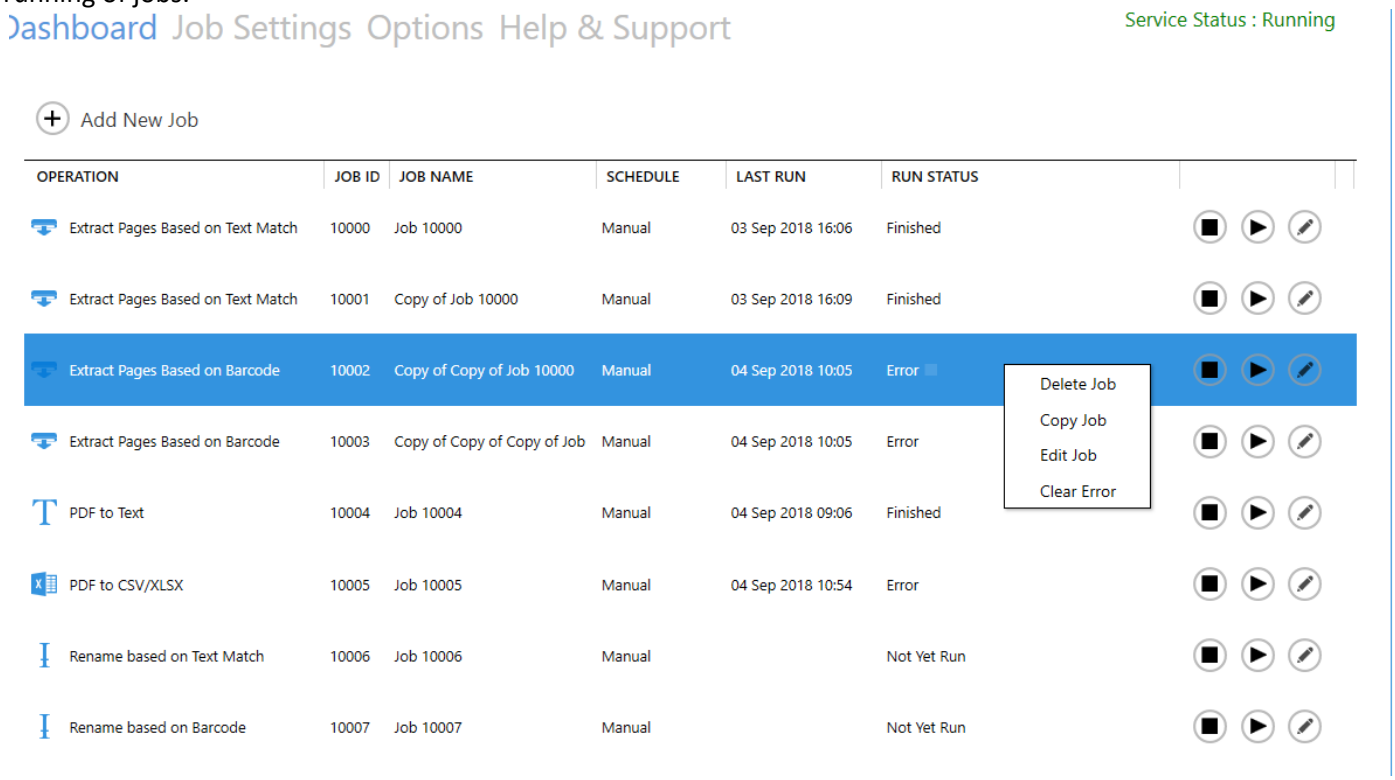
When a license is purchased (see http://www.aquaforest.com/en/tif_pdf_buy.asp#collapseEleven) a permanent license key will be issued for use with the product which should be entered using the Options tab of the GUI. There is no need to download another version of the software.

1.4 Version Breakdown

	Standard *Smart PDF Splitting and Renaming.*	**Professional** *Smart PDF Data Extraction, Splitting and Renaming.*	**Server** *Automated Smart PDF Data Extraction, Splitting and Renaming.*
Split PDF by Content, Barcode or Page Ranges	☑	☑	☑
Rename PDF by Content or Barcode	☑	☑	☑
Extract PDF Pages by Content or Barcode	☑	☑	☑
Includes Optical Recognition for processing Image PDF files	☑	☑	☑
Graphical User Interface	☑	☑	☑
Process Files in SharePoint	☑	☑	☑
Extract PDF Text and Data to CSV, Excel or Text File.		☑	☑
Command Line Interface			☑
Automated Scheduling			☑
Number of Cores Supported	1	1	4

2 KINGFISHER WINDOWS INTERFACE

When Kingfisher is launched, the main window is displayed as shown below. This allows the definition, saving and running of jobs.



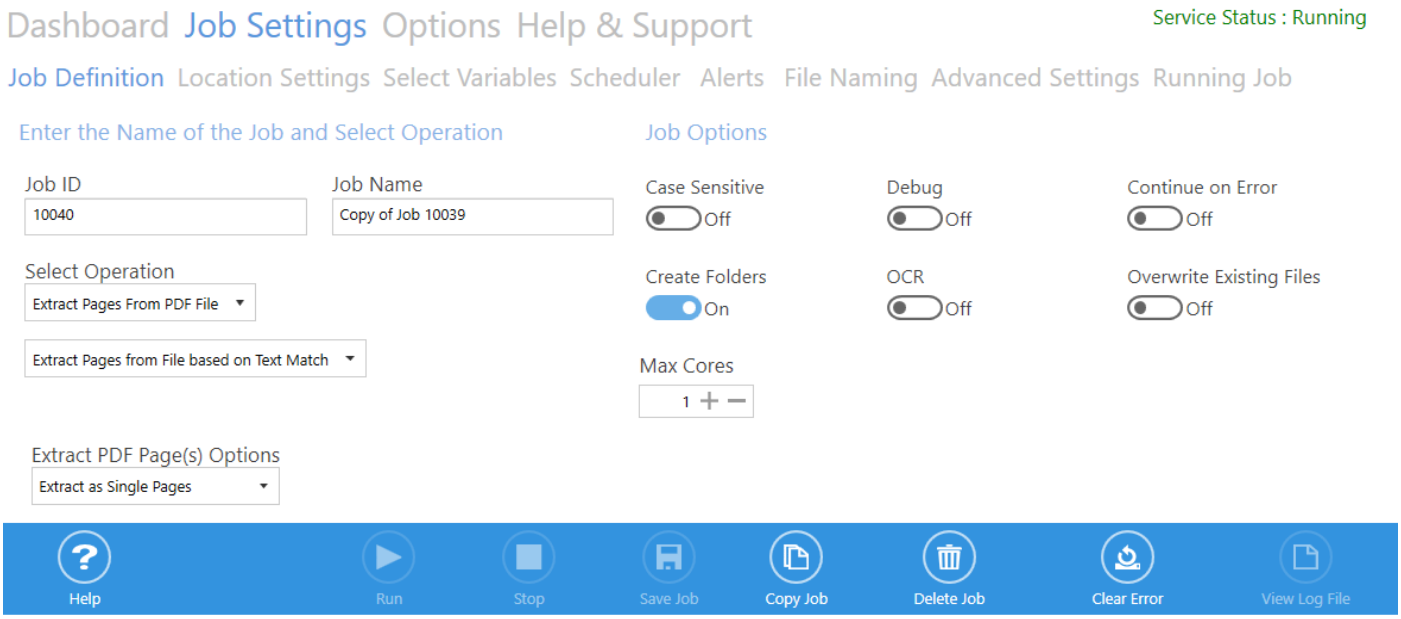
This section will explain all the components of the Kingfisher Graphical User Interface

2.1 Dashboard

The kingfisher Dashboard contains a list of all the jobs defined in kingfisher, you can start/stop jobs by clicking the "Play" and "stop" button by the side of the job name, right clicking on a Job also gives some options as shown on the image above.

Screen Field / Button	Description
Add New Job	This button launches a wizard to create a new kingfisher job.
Stop Job	Stops an already running job.
Run Job	It is represented by the play button, clicking on it runs the job associated with the button.
Edit Job	This can be accessed by either right clicking the job or clicking the button with the pen icon, use this option to Edit a job, or just double click on the job.
Copy Job	Use the copy job in the context menu to make a duplicate of the currently selected job.
Delete Job	Use the delete job in the context menu to delete the currently selected job.
Clear Error	Use this option to clear errors of jobs in error before running them.

2.2 Job Settings



Screen Field / Button	Description
View Log File	Click this button to display the log file of the selected Job.
Save	Click this button to save all the changes in the Job.
Run	Click this button to execute the rule on the selected files.
Stop	Click this button to halt the processing of the current batch of files at the earliest convenience.
Delete Job	Click this button to delete the currently selected job.
Copy Job	Click this button to make a copy of the selected job.
Clear Error	Use this option to clear errors of jobs in error before running them.
Explore Destination	Click this button to open the target folder in windows explorer, if the folder does not exist, it will display an error message.

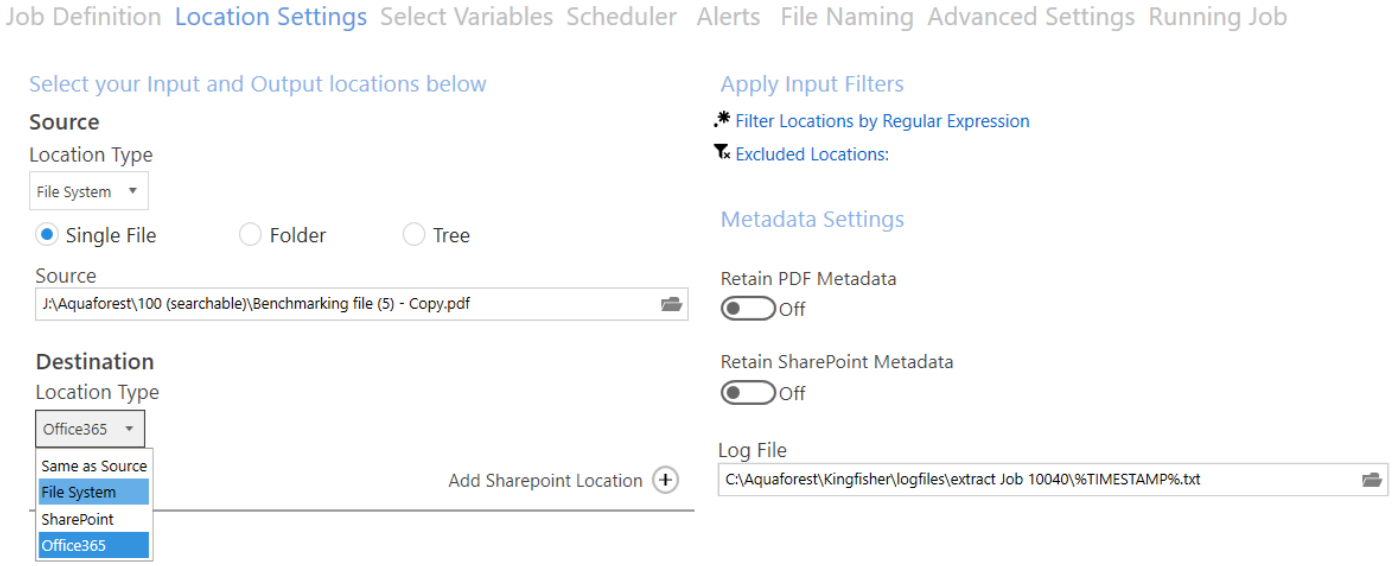
2.2.1 Job Definition

Job definition contains the common job properties that need to be set for the job to be executed.

Screen Field / Button	Description
Job Name	This is used to identify a job by the user, a default name will be generated when you create a new job, it is always a good practice to give the jobs sensible names when you create them.
Select Operation	Use this combe box to set the operation you want to carry out, there are nine options to choose form. Check section 4 for more details. • Rename PDF File • Split PDF Files • Extract Pages from PDF Files • Extract Content from PDF Files
Create Folders	If this is checked, Kingfisher will create the target directories if they do not exist.
Overwrite Existing Files	If checked, generated PDF files will overwrite existing files of the same name (but not source files).
Continue on Error	When checked, folder and tree processing jobs will continue when individual files fail.
Debug	If switched on the job log will be more detailed
Max Cores	Sets the number files that should be processed in parallel.
Case Sensitive	Setting this to true will make sure all the searches, text operations and pattern matching will be performed in a case sensitive fashion.
OCR	If switched, Kingfisher will use OCR to locate the text on the PDF document, this is good for image PDFs and PDF file with Encoding.

2.2.2 Location Settings

Kingfisher allows users to process files in SharePoint Online, SharePoint On-Premise and on the Windows file system. This section will show you the details of the "Location Settings" tab.



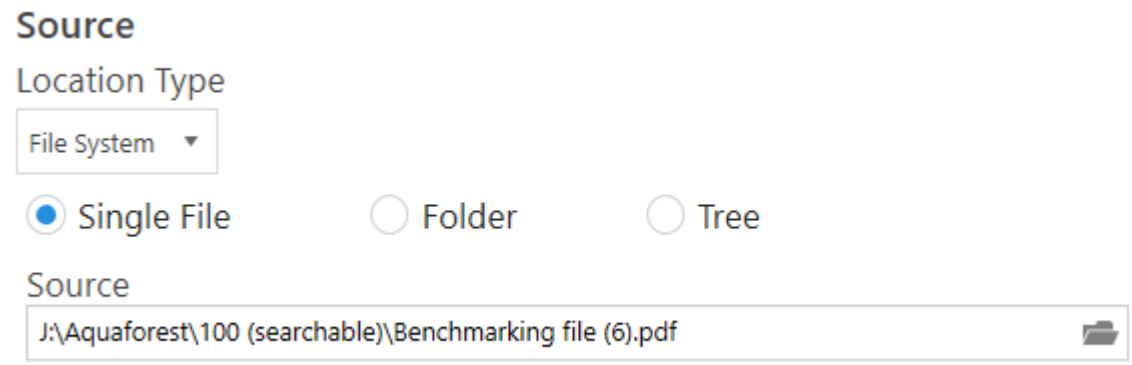
Screen Field / Button	Description
Source	Use this Combo Box to select the Source location type, the options are given below. • File System • SharePoint • Office365
Destination	Use this Combo Box to select the Source location type, the options are given below. • Same as Source: This option puts the file back in the source location • File System • SharePoint • Office365
Filter Locations by Regular Expression	Regular expressions can be used to filter the input files that are processed by kingfisher.
Exclude Locations	Add paths that you want kingfisher to skip.
Retain PDF Metadata	Switch this on to retain the PDF metadata of the source PDF file.
Retain SharePoint Metadata	Switch this on to retain the SharePoint metadata of the source PDF file, note that the destination library must have the same column types and names as the source library for this to work.
Log File	The path to the log file for the current rule, you can use the %TIMESTAMP% and %DATESTAMP% variables to create log files based on dates or time.

2.2.2.1 Location Types

Kingfisher gives you an option of 3 different location types, this section explains the different ways there are added or edited in a job.

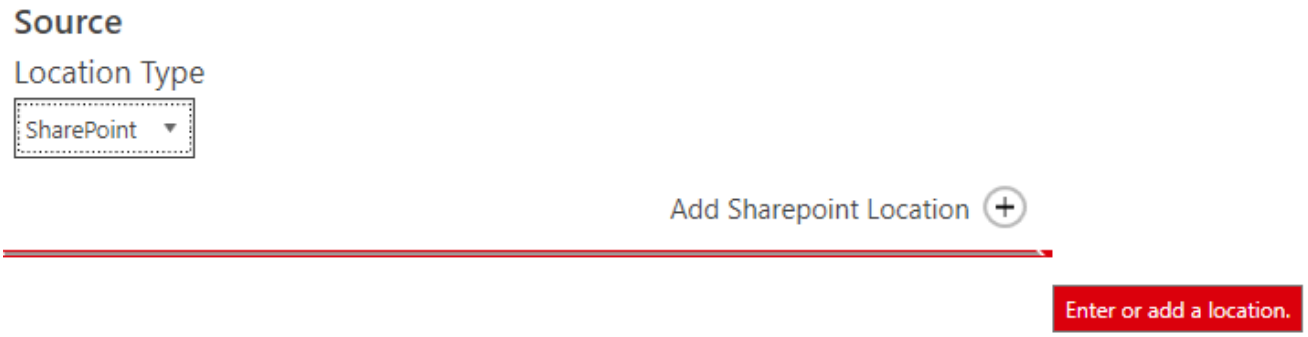
2.2.2.1.1 File System

This is straight forward, choose between processing a “Single File”, “Folder” or “Tree”. After chose the file system type, you can use the browse button to select the location.

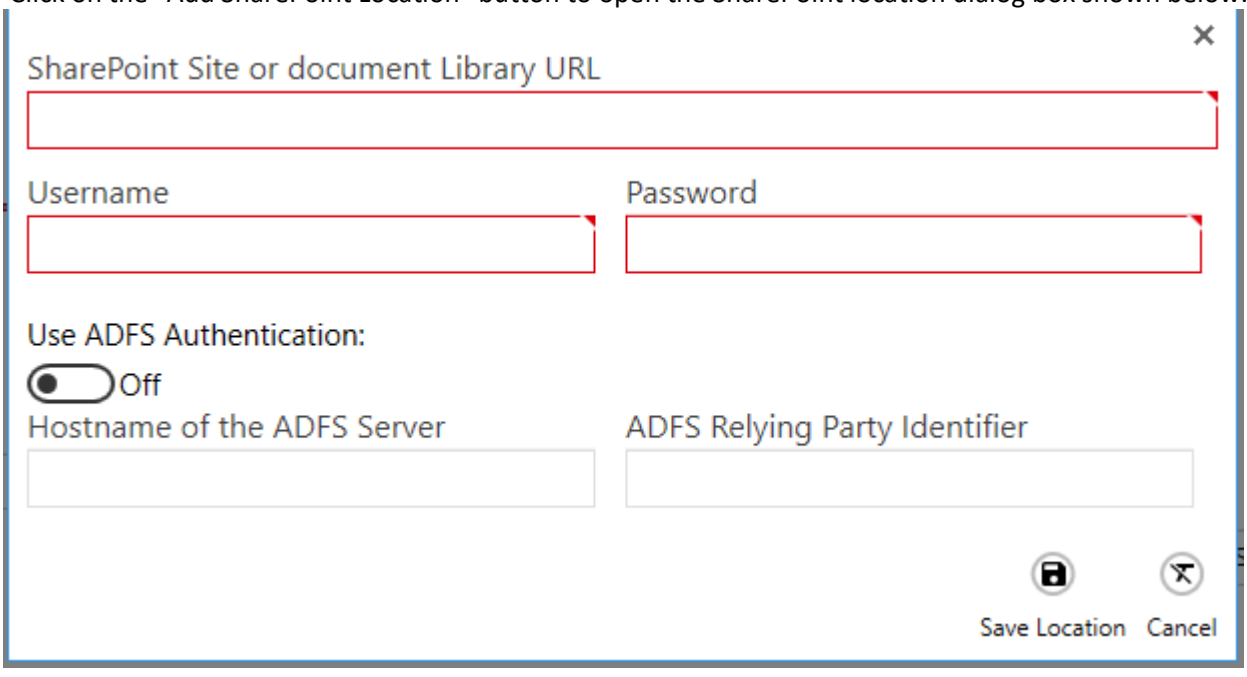


2.2.2.1.2 SharePoint

Adding a SharePoint location is a bit different, after select the “SharePoint” option, you will see the button below.



Click on the “Add SharePoint Location” button to open the SharePoint location dialog box shown below.



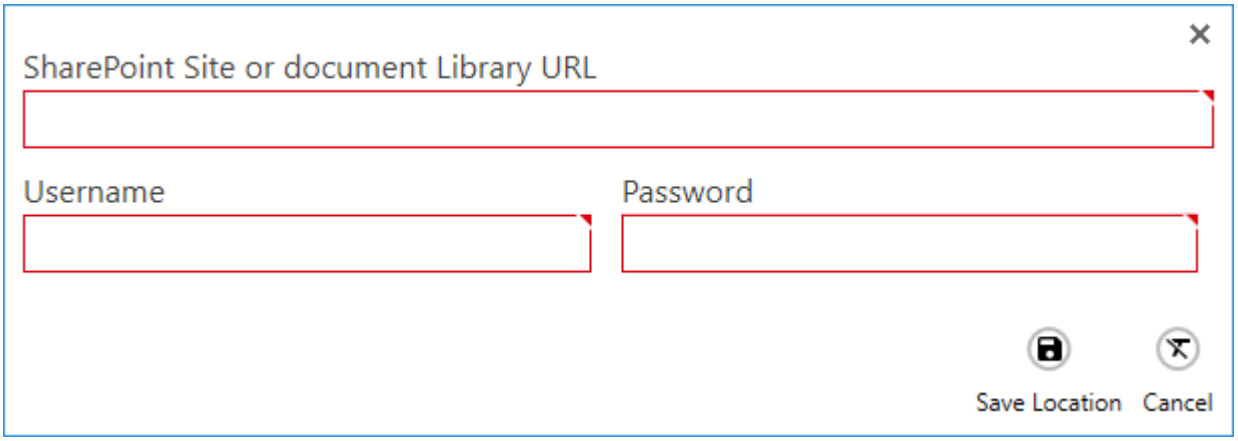
Screen Field / Button	Description
SharePoint URL	Provide the URL of the SharePoint site or document library you want kingfisher to process.
UserName	Provide the username of a user that has access to the URL you provided above.
Password	Provide a password for the User above.
Use ADFS Authentication	Switch this on if you use Active Directory for your SharePoint User Management.
Hostname of the ADFS Server	Provide the name of the Active Directory server.
ADFS Relying Party Identifier	Provide the Relying Party Trust identifier for your SharePoint .

After you enter your SharePoint details, click the “Save Location” button and the locations will be validated as shown below.

After saving the location, you can always edit by clicking the “Edit SharePoint Location” shown below.

2.2.2.1.3 Office365

Adding a new Office 365 location is like adding a new SharePoint location, the only difference is the filed types as shown in the screenshot and table below.



Screen Field / Button	Description
SharePoint URL	Provide the URL of the SharePoint site or document library you want kingfisher to process.
UserName	Provide the username of a user that has access to the URL you provided above.
Password	Provide a password for the User above.

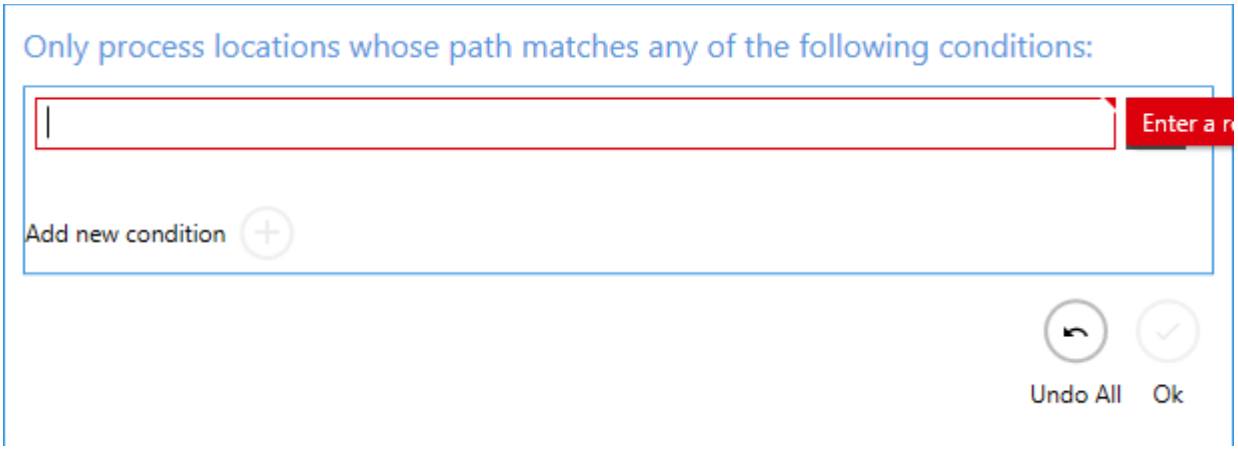
Follow the same process as the previous section if you want to save or edit the location.

2.2.2.2 Apply Input Filters

Kingfisher give you the ability to filter or exclude source files/paths based on difference criteria.

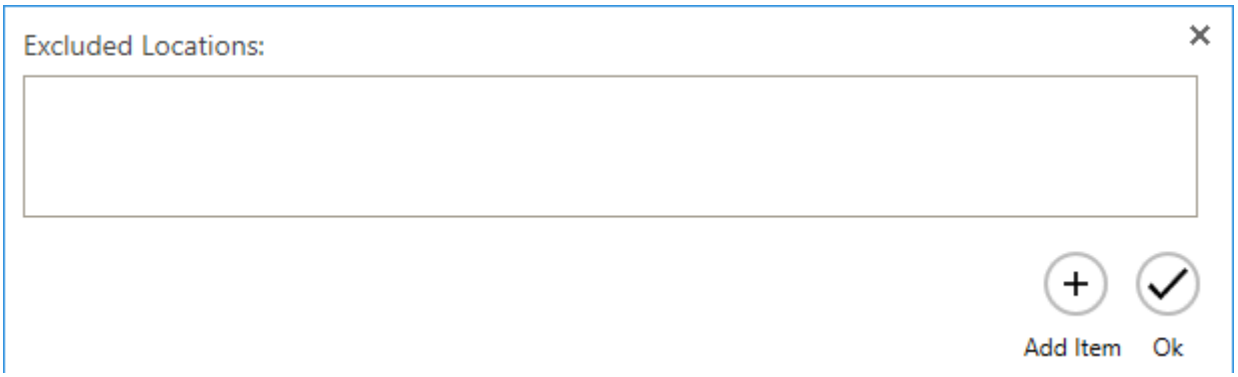
2.2.2.2.1 Filter Locations by Regular Expression

If you provide any values for this, kingfisher will only process files that match the regular expression provided. To add values, click the “Filter Locations by Regular Expression” link, this will show the dialog below. Then click the “Add new Location” button.



2.2.2.2.2 Exclude Locations

Any path you provide here will be skipped by kingfisher, to add values to this property, click the “Exclude Location” link. The Dialog below will be shown.



Click on the “Add Item” button to add a new value, after you are done click the Ok button.

2.2.3 Select Variables

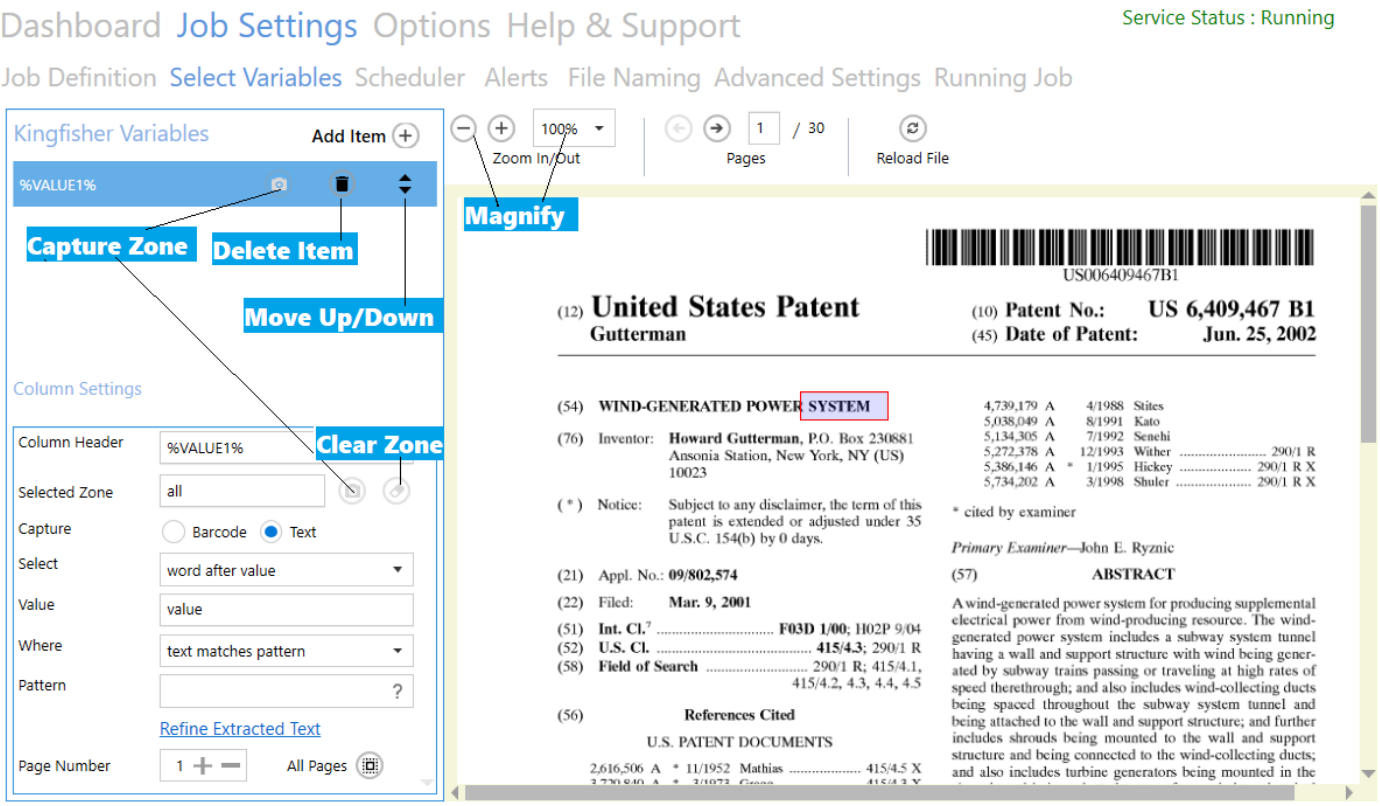
One of the most attractive features of kingfisher is the ability to select text or barcode at a position in a PDF page, the selected text or barcode can be used to:

- Decide whether to split the document at the page where the text of barcode was found.
- Decide whether to extract the page where the text of barcode was found.
- Use these extracted text or barcode values to rename the file

The “Select Variable” tab is used to select four variables named:

- “%VALUE1%”: This is used to decide the split or extract page, it is also used to generate the output file name, note this must be selected for some jobs to run.
- “%VALUE2%”: This is used to generate the output file name.
- “%VALUE3%”: This is used to generate the output file name.
- “%VALUE4%”: This is used to generate the output file name.

To assign a zone to a variable, you should first choose the zone on the PDF page with your mouse, click the camera icon associated to the variable you want to assign the zone to. If you do not assign a zone to the variable, kingfisher will assume you want to use the whole page as a zone. You will notice the “Selected Zone” Text Box change from “all” to the coordinates you selected.

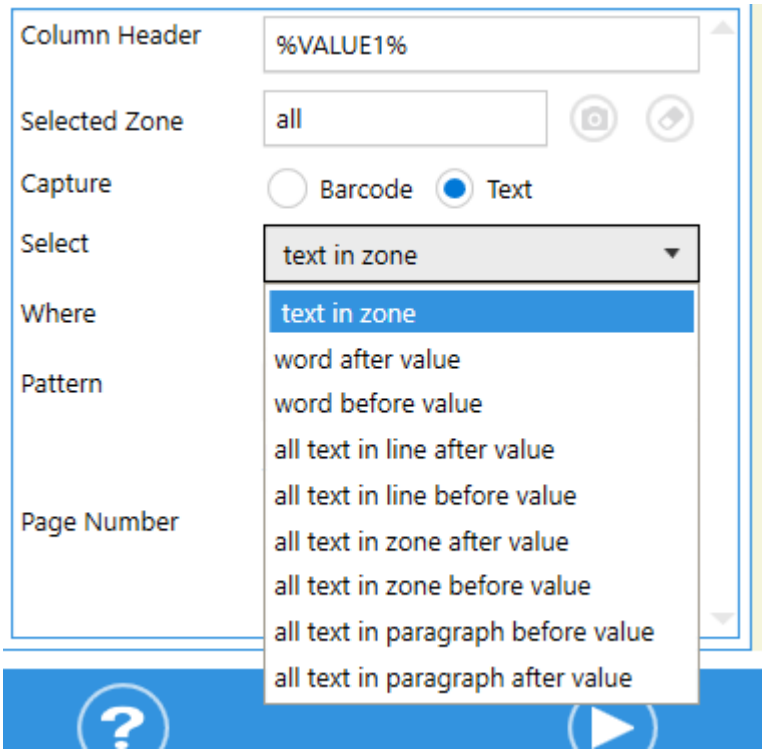


Screen Field / Button	Description
Magnifier	Use the +/- buttons to zoom in or zoom out
Next	Click to go to the next page of the PDF file
Back	Click to go to the previous page of the PDF file
Delete Item	Click this button to delete the variable or item.
Move Up/Down	You can use this to rearrange the order of the variables.
Selected Zone	Click the camera icon to assign a selected zone to a variable
Clear Zone	Click this button set the selected zone to the whole page
Text	Check this radio button to extract text values
Barcode	Check this radio button to extract barcode values

Select	Use this drop-down box to specify the text you want to extract relative to a word/value. See section 2.2.2.1 for more details.
Where	Use this Drop-down box to specify if a regular expression should be used to match the selected text. See section 2.2.2.3 for more details.
Refine Extracted Text	Click this link to refine the text extracted. See section 2.2.2.4 for more details
Page Number	You can specify specific pages to want Kingfisher to check for certain values.
All Pages	Click this button to make Kingfisher search for the values on all pages.

2.2.3.1 Text Match Option

When extracting text, Kingfisher gives you the ability to select text that is relative to a word/pattern. The table below shows the options available and what they mean.



Screen Field / Button	Description
Select	• **text in zone:** This option will select all the text that was extracted. • **word after value**: If this option is selected, kingfisher will return the word that appears immediately after the word supplied in the text box. • **Word Before value:** If this option is selected, kingfisher will return the word that appears immediately before the word supplied in the text box. • **all text In Line after value:** If this option is selected, kingfisher will return all the words that appear on the same line after the word supplied in the text box. • **all text In Line before value:** If this option is selected, kingfisher will return all the words that appear on the same line before the word supplied in the text box.

	• **all text In Zone after value:** If this option is selected, kingfisher will return all the words that appear in the selected zone after the word supplied in the text box. • **all text In Zone before value:** If this option is selected, kingfisher will return all the words that appear in the selected zone before the word supplied in the text box. • **all text in paragraph before value:** If this option is selected, kingfisher will return all the words that appear in the paragraph before the word supplied in the text box. **all text in paragraph after value:** If this option is selected, kingfisher will return all the words that appear in the paragraph after the word supplied in the text box.
Pattern	• **Where text matches any pattern:** This option will not match the text to any pattern • **Where text matches pattern:** Provide a Regular expression here to match the extracted text to, this will be explained more below.

2.2.3.2 Barcode

When extracting Barcode, you can choose the barcode format you are interested in.

Screen Field / Button	Description
Barcode Type	This combo box allows the user to select the barcode the they want to extract, below is the list of barcode types that kingfisher can recognize • Aztec 2D barcode format. • CODABAR 1D format. • Code 39 1D format. • Code 93 1D format. • Code 128 1D format. • Data Matrix 2D barcode format. • EAN-8 1D format. • EAN-13 1D format. • ITF (Interleaved Two of Five) 1D format (Code 25). • MaxiCode 2D barcode format. • PDF417 format. • QR Code 2D barcode format. • RSS 14 • RSS EXPANDED • UPC-A 1D format. • UPC-E 1D format. • MSI • Plessey

2.2.3.3 Text Pattern Option

Kingfisher allows the user to define the text pattern of the text to be extracted from the PDF document. This helps the user make sure only text that match a certain pattern will be extracted. For example, a user might want a phone number, but a postcode gets extracted just because it was present in the same location as the postcode on a different page. Kingfisher provide pattern matching so that a user can make sure the text that is extracted is properly validated before it is accepted.

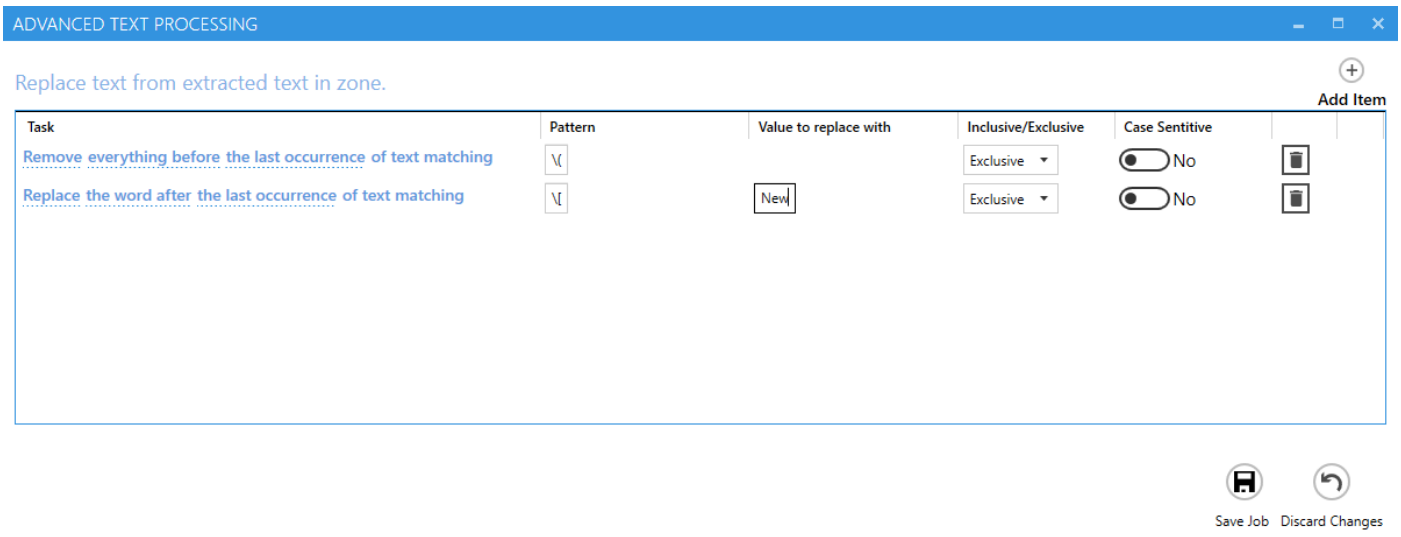
Patterns are used to match parts of the document. Some useful examples are below. For a complete reference on patterns including advanced examples see the link below.

https://docs.microsoft.com/en-us/dotnet/standard/base-types/regular-expression-language-quick-reference

Pattern	Matches	Examples
invoice	The word "invoice"	
invoice no:	The phrase "invoice no:"	
.	Any 1 character	A 9 %
….	Any 4 characters	12ab $5.00
Invoice….	The word invoice followed by any 4 characters. A space is a character.	Invoice 123 Invoice No: Invoice A9-
^invoice	The word invoice at the start of a line	
invoice$	The word invoice at the end of a line	
[0-9] [0-9] [0-9]	Any 3 digits	123 900
[a-z] [a-z] [a-z]	Any 3 alphabetic characters	abc Inv

2.2.3.4 Refine Extracted Text

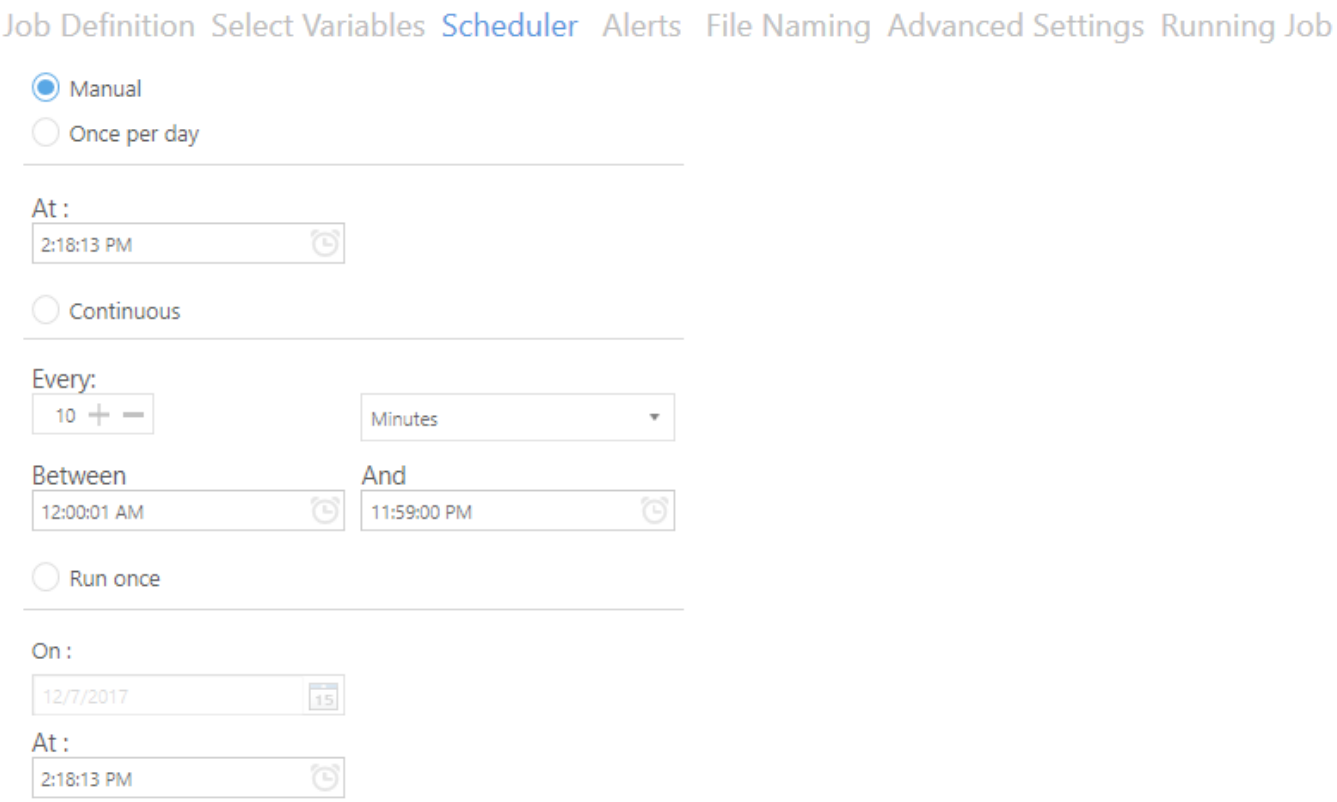
We have added a more sophisticated way for users to refine text in the new version of kingfisher. This simply entails removing/replacing certain text patterns or values.



Option	Explanation
Task	The first part of the task is whether to remove or replace the pattern provided, the next part deals with the position and the frequency of the removal or replacement.
Pattern	The pattern to search for, this can be a regular expression, Use the “\” literal to escape special characters.
Values to replace with	If you choose to replace a pattern, you will have to provide a substitute value in this text box.
Inclusive/Exclusive	When you chose Inclusive, kingfisher will remove all the text specified including the pattern being matched. When you choose Exclusive, kingfisher will remove all the text specified without the pattern being matched.
Case Sensitive	This option tells kingfisher whether to use a case sensitive approach or not.

2.2.4 Scheduler

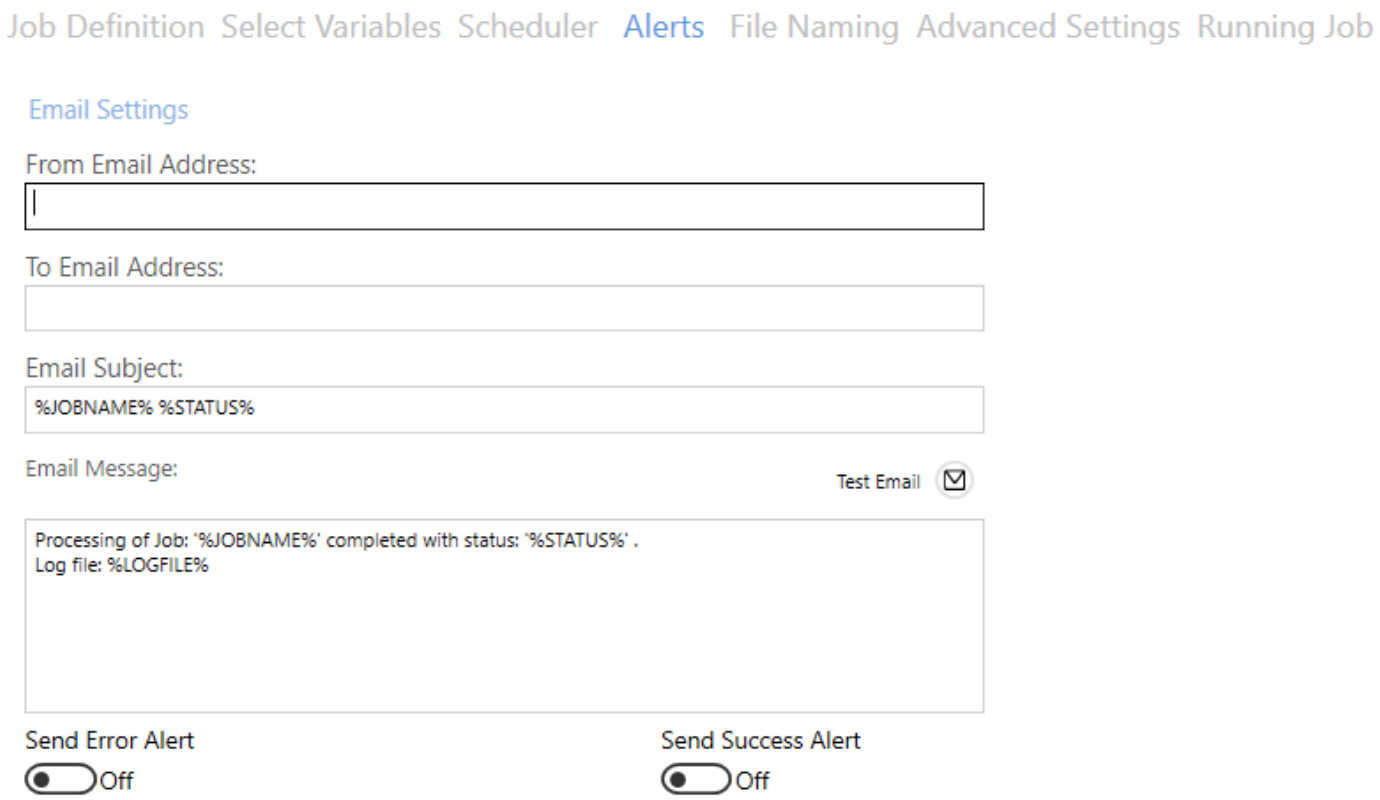
This tab allows the user to schedule when jobs are executed.



Setting	Description
Manual	This means that the document library has to be run manually by clicking on the “Run” button on the dashboard.
Once per day	This allows the document library to be scheduled to run at a specified time each day.
Continuous	This allows the document library to be scheduled to run periodically between a start time and end time each day. The periods may be minutes, hours, days or months. For example, a document library may be specified to run every 1 hour between 9:00 and 17:00.
Run Once	This allows the document library to be scheduled to run only once at a specified time.

2.2.5 Alerts

Use this tab to set email alerts for jobs.



Setting	Description
From Email Address	The email address to send the email from.
To Email Address	The email address to send the email to.
Email Subject	The email subject. You can use the following templates: %JOBNAME% - will be replaced by the name of the library %STATUS% - will be replaced by “success” or “error” depending on whether the job ran successfully or not
Email Message	The email message to send. You can use the following templates: %JOBNAME% - will be replaced by the name of the library %STATUS% - will be replaced by “success” or “error” depending on whether the job ran successfully or not %LOGFILE% - will be replaced by the path of the log file.
Send Error Alert	Switch this on to send an alert when a job ends in an error.
Send Success Alert	Switch this on to send alerts when the job successfully completes.

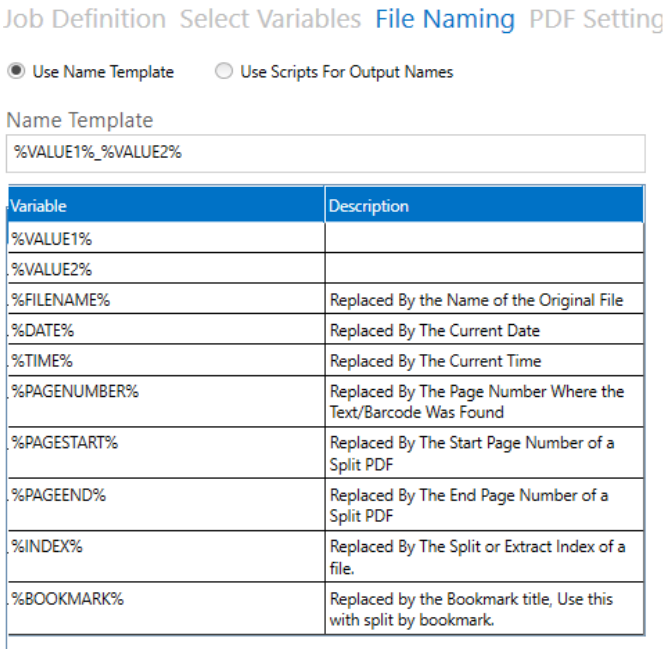
2.2.6 File Naming

Renaming output file is one of the most attractive and robust features of Kingfisher. Kingfisher offers various variables that allows the user to customize the final output name. These variables range from Text/Barcode values extracted from the PDF document to other predefined variables. Kingfisher also gives the user the ability to write scripts to generate the file name.
In kingfisher, output file names can be generated in two ways:

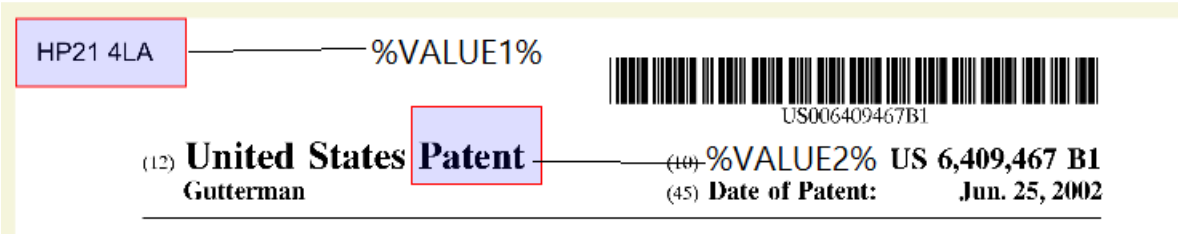
2.2.6.1 Name Template

You can enter a string in the name template text box, any occurrence of variables in the table below will be replaced by the appropriate value at runtime.

%VALUE1%	The primary barcode/text value.
%VALUE2% %VALUE3% %VALUE4%	The barcode/text value(s) found.
%INDEX%	The index of the output file in a split or extract page operation.
%PAGENUMBER%	The page number of the extracted page in the original PDF file.
%FILENAME%	The original file name without the .pdf extension.
%PAGESTART%	Start page number .
%PAGEEND%	End page number.
%BOOKMARK%	This represents the bookmark name and can be used with the “Split by Top Level Bookmarks” split option only.
%TIME%	The current Time.
%DATE%	The current Date.
%RUNDATE%	The Time at which the job was executed, the difference between this and the %TIME% variable is that this value will be same for all files processed in a single job run.
%JOBNAME%	The Name of the Job.

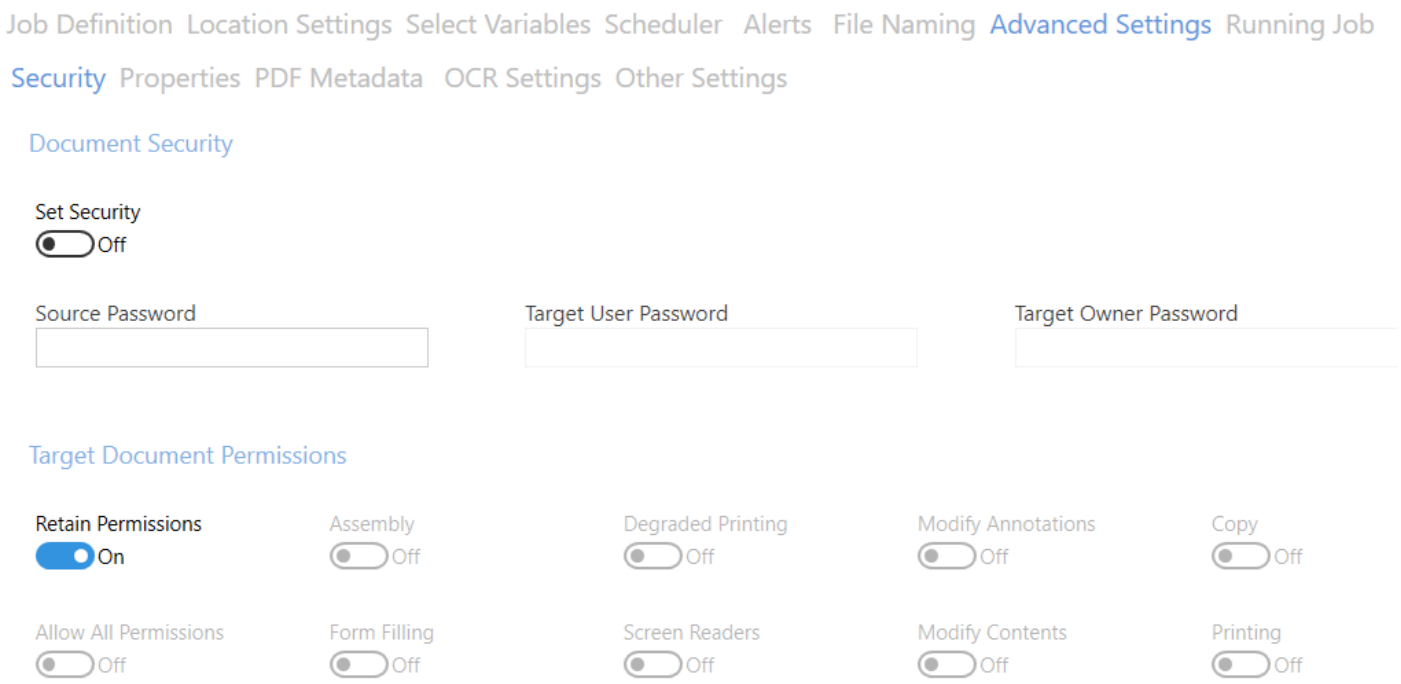


Using the image below as an example, if the document below is to be renamed,
Using “%VALUE1%\\Technical%VALUE2%” as the File Template will create a file named “TechnicalPatent” in a folder named “HP21 4LA”.



2.2.7 Advanced Settings

2.2.7.1 Security



2.2.7.1.1 Source Document Security

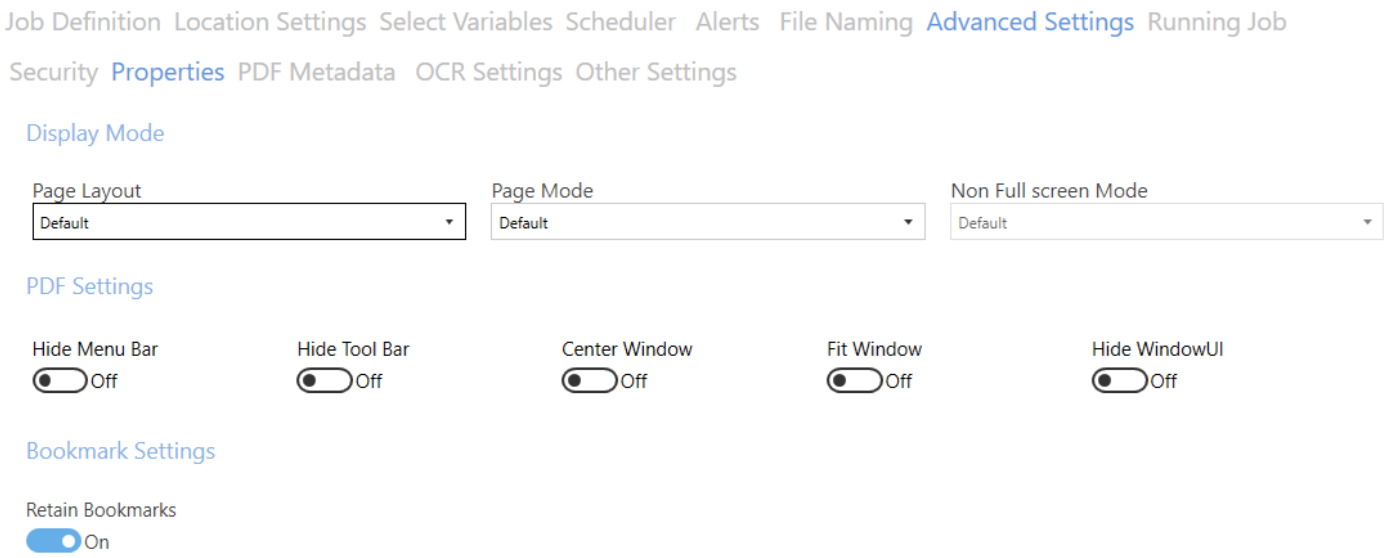
Screen Field / Button	Description
Source Password	If the source document requires a password it should be entered here.

2.2.7.1.2 Target Document Security

Option	Description
User Password	A password that will be required to open the document.
Owner Password	A password that will be required to change the document permissions.
Permissions	

Permission	Description
Allow ALL Permissions	All the permissions below.
Printing	Allow high-quality printing.
Modify Contents	Allow assembly (see below) and other document medications.
Copy	Allow text and graphic copying and extraction.
Modify Annotations	Allow modification of annotations.
Form Filling	Allow filling of form fields.
Screen Readers	Allow extraction of text and graphics in support of accessibility.
Assembly	Allow rotation, insertion or deletion of pages.
Degraded Printing	Allow low-quality printing.

2.2.7.2 Properties



2.2.7.2.1 Display Modes

Option	Description
Page Layout	The setting for initial document page display. One of: • Single Page • Continuous • Continuous Facing (odd pages left) • Continuous Facing (odd pages right)
Page Mode	The setting for initial viewer mode. One of: • Neither Bookmarks nor Thumbnails Open • Bookmarks Open • Thumbnails Open • Bookmarks & Thumbnails Open • Full Screen
Non-Full Screen Mode	Only applicable where Page Mode=Full Screen. The setting for document page display when exiting Full Screen mode. One of: • Neither Bookmarks nor Thumbnails Open • Bookmarks Open • Thumbnails Open

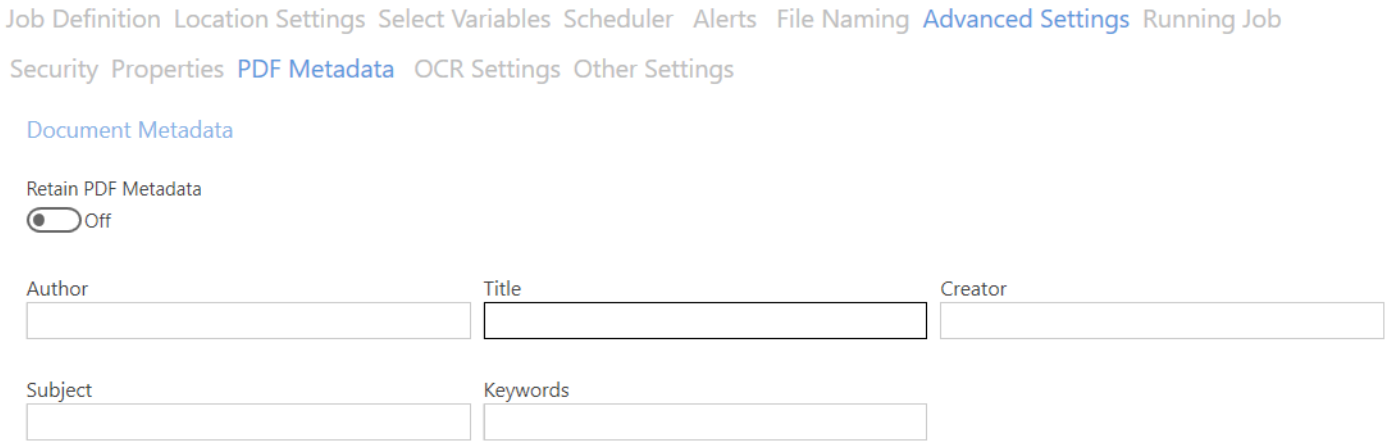
2.2.7.2.2 Settings

Option	Description
Hide Tool Bar	The viewer's tool bar will be hidden.
Hide Menu Bar	The viewer's menu bar will be hidden.
Hide Window UI	The viewer's UI elements (scrollbars etc.) will be hidden.
Fit Window	The viewer will resize the document's window to fit the size of the first displayed page.
Centre Window	The document window will be positioned in the centre of the screen.

2.2.7.2.3 Bookmark Settings

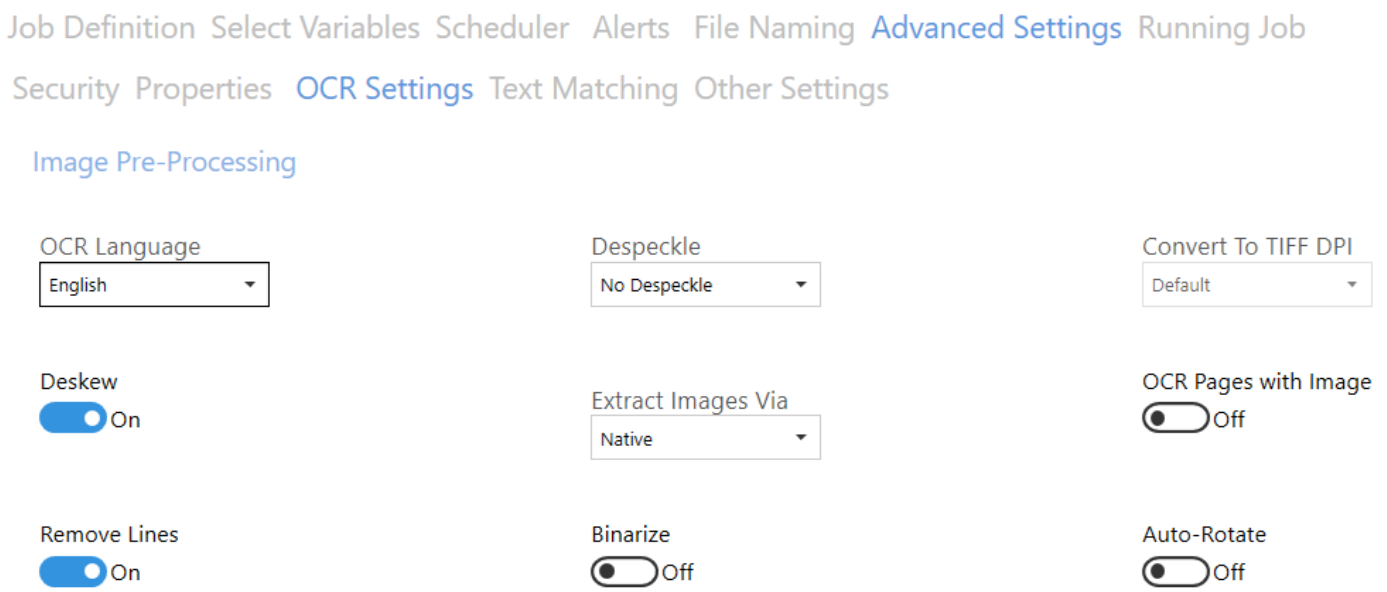
Option	Description
Retain Bookmarks	Select this option to carry over the bookmarks in the original PDF to the split PDF. Default is set to true.

2.2.7.3 PDF Metadata



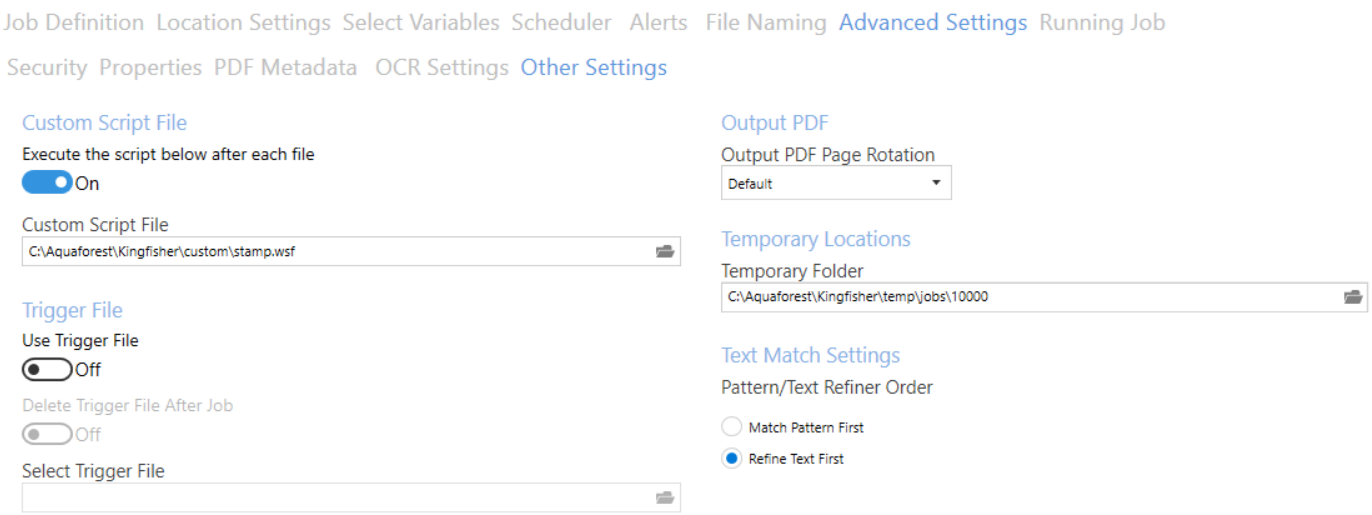
Option	Description
Retain PDF Metadata	Set to yes if you want the output PDF file to retain the metadata of the input PDF, when you set this to true, kingfisher will disable all the metadata options below.
Author	Provide a new value for the **Author**
Title	Provide a new value for the **Title**
Creator	Provide a new value for the **Creator**
Subject	Provide a new value for the **Subject**
Keywords	Provide a new value for the **Keywords**

2.2.7.4 OCR Settings



Option	Description
Convert To TIFF	Set to yes if you us to convert the input PDF to a TIFF file before performing OCR when extracting text from the PDF.
OCR Pages with Image	If this is set to true, kingfisher will OCR all pages with images in them by default, select this if you have partially searchable documents.
Despeckle	Despeckle the image – The method removes all disconnected elements within the image that have height or width in pixels less than the specified figure. The maximum value is 9 and the default value is 0 .
Auto Rotate	Auto rotates images to an orientation where more text can be recognized.
Remove Lines	Remove lines from images.
Binarize	Use This option to process coloured files.
Deskew	Deskew (straighten) the image.

2.2.7.5 Other Settings



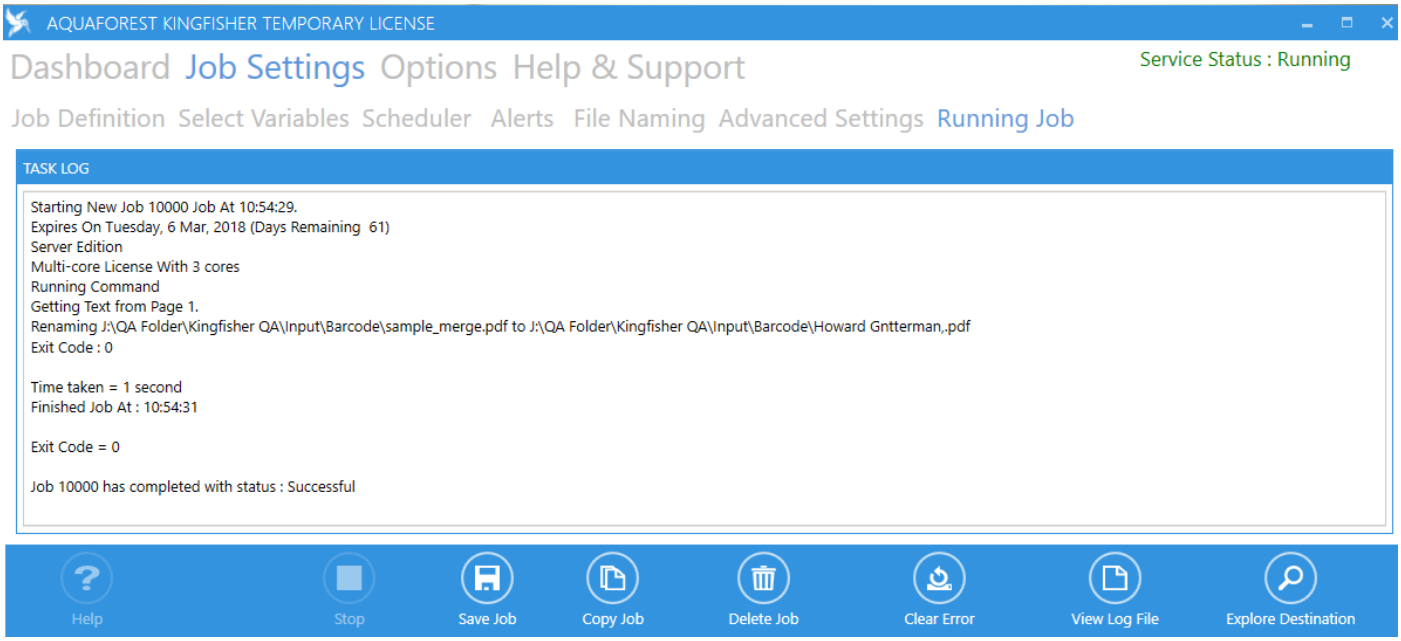
Option	Description
Custom Script File	
Execute the script below after each fie	Switch this on if you want to execute a .wsf file after each output file is produced. We send the following arguments: “newfilename” and “oldfilename”.
Custom Script File	The location of the script file.
Trigger File	
Use Trigger File	When this option is switched on, a kingfisher job will not start if a trigger file specified is not available.
Delete Trigger File After Job	King fisher will delete the trigger file after the job completes if this option is turned on.
Select Trigger File	Use this control to provide the trigger file name and location.
Output PDF	
Output PDF Page Rotation	Choose the rotation of the output PDF pages.

Retain Bookmarks	Set to true to retain the bookmarks in the output PDF file.
Temporary Folder	You can change the temp folder using this directory browser control.
Csv Log File	The csv log files contain a summary of the new file names generated by kingfisher.
Text Match Settings	
Match Pattern First	Kingfisher will match the extracted text with the provided pattern before it applies the text refiner
Refine Text First	Kingfisher will apply the text refiner to the extracted text before matching it to the pattern.

2.2.8 Running Job

This tab has two elements explained below:

- **Task Log:** This text area displays the progress of the current job being run.
- **Explore Destination:** This button opens the destination folder if it exists.

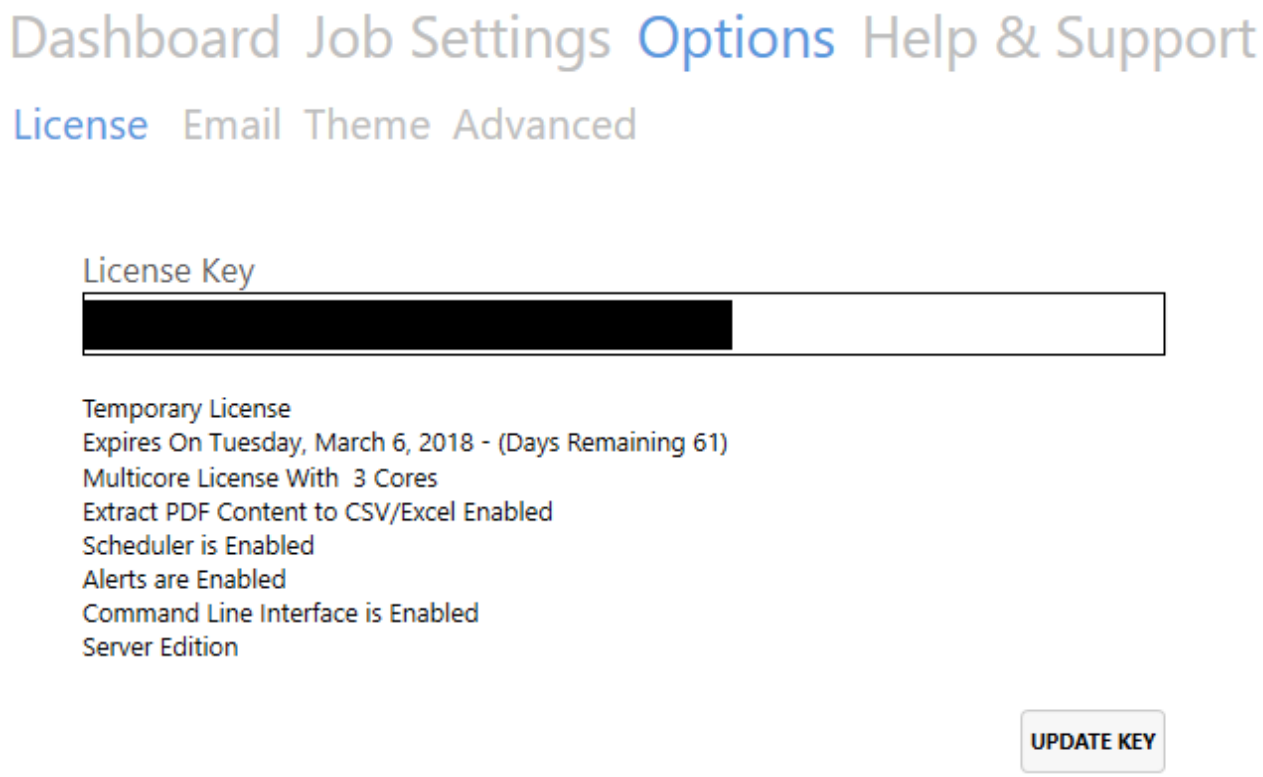


2.3 Options

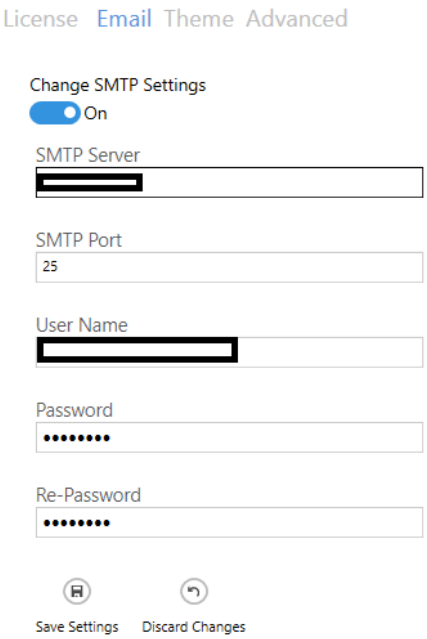
The options tab contains general settings in kingfisher like license keys, Smtp Settings, Themes etc.

2.3.1 License Keys

Enter the license key in the License key text box and click the “UPDATE KEY” button



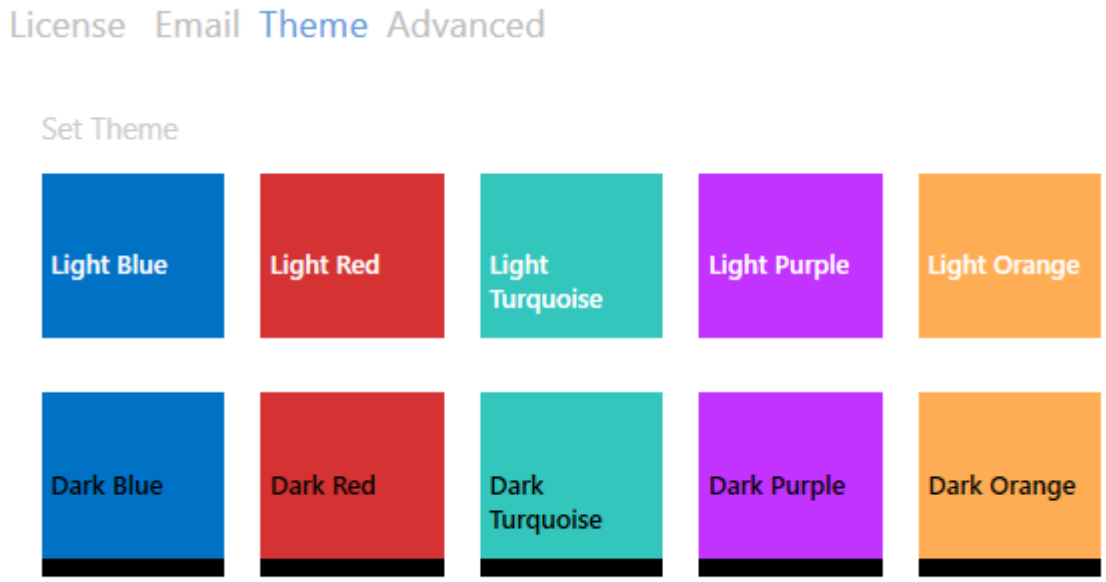
2.3.2 Email



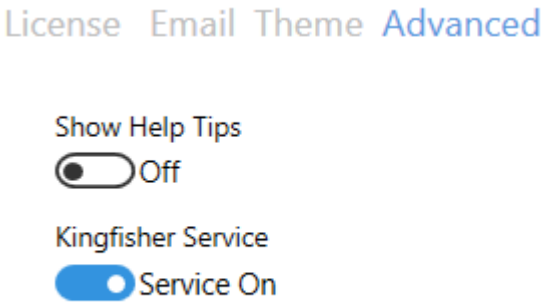
Option	Description
Smtp Server	Enter the address of your SMTP server.
SMTP Port	Enter the SMTP port number .
User Name	Enter your SMTP User Name.
Password	Enter your smtp password.
Save Settings	Click to update the SMTP settings.
Discard Changes	Click this button to revert to the last saved SMTP settings.

2.3.3 Theme

Use this tab to switch between the different themes available in Kingfisher.



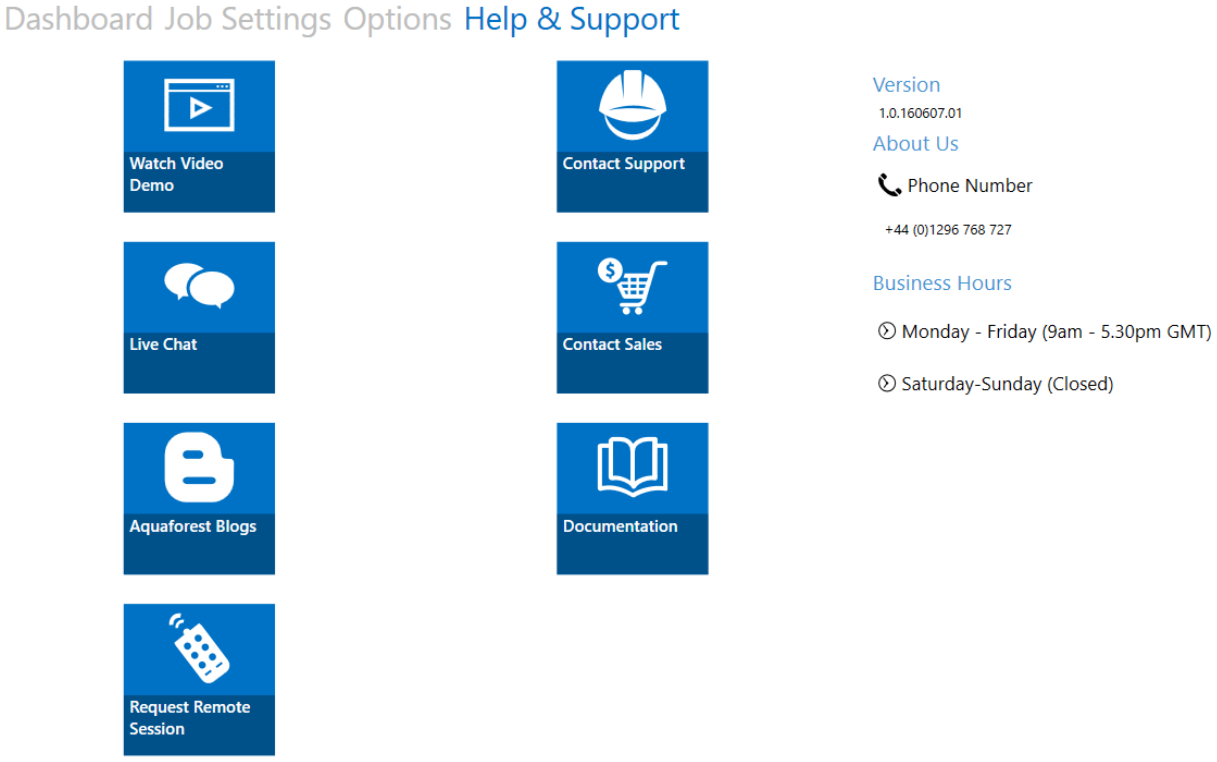
2.3.4 Advanced Settings



Option	Description
Show Help Tips	Switch this on if you want kingfisher to show you tips as you use it.
Kingfisher Service	Use this switch to On or Off the kingfisher service.

2.4 Help

Our help Tab give you access to all the different ways you can get support on how to use Kingfisher.



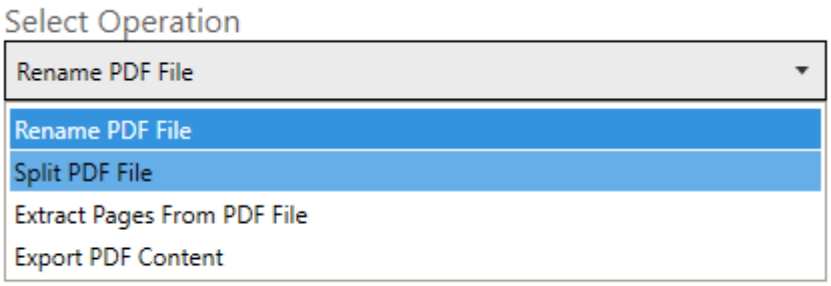
Option	Description
Watch Video Demo	This will launch a video showing the user how to create and run jobs in Kingfisher.
Request Remote Session	Request a remote session and we will contact you to arrange a suitable time.
Contact Support	Launches the default Email Client with the email address of our support team.
Contact Sales	Launches the default Email Client with the email address of our sales team.
Live Chat	Opens our live chat in your default web browser.
Aquaforest Blogs	This takes the user to the Aquaforest Blogs.
Documentation	This Open the directory contain the Documentation of kingfisher.

3 KINGFISHER OPERATIONS

Kingfisher has 4 main operations which are:

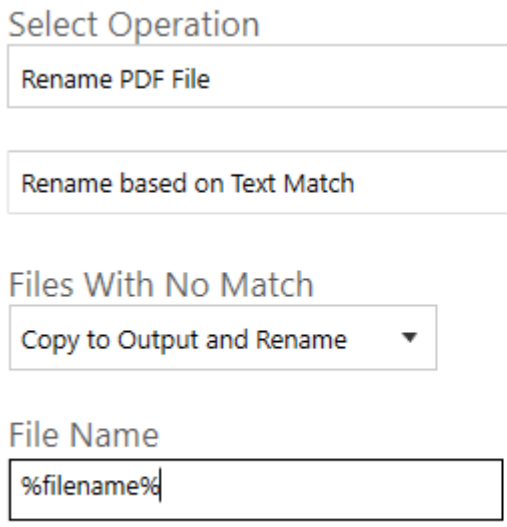
- Rename PDF File
- Split PDF File
- Extract Pages from PDF File
- Extract Content from PDF File

To select an operation, select one of the Main Operations in the “Select Operation” combo-box.



Each of the operation listed above are further divided into sub-categories, the section below explains each of this operations in more details.

3.1 Rename PDF File



Option	Description
Files With No Match	This gives the user three options of what to do with files that do not match the text or barcode criteria specified by the user. • Do Not Copy to Output • Copy to Output • Copy to Output and Rename
File Name	If the “Copy to Output and Rename” option is selected you will need to provide a filename mask like the one in the “File Naming” tab, the table below gives you a breakdown of the variables available. %FILENAME% — The original file name without the .pdf extension.

	%BOOKMARK%	This represents the bookmark name and can be used with the “Split by Top Level Bookmarks” split option only.
	%TIME%	The current Time.
	%DATE%	The current Date.
	%RUNDATE%	The Time at which the job was executed, the difference between this and the %TIME% variable is that this value will be same for all files processed in a single job run.
	%JOBNAME%	The Name of the Job.

3.1.1 Rename Based on Barcode

This Operation will take a PDF file that has barcodes in its pages as an input and renames the PDF file based on the value of the barcode. You can also specify Areas on the PDF page; kingfisher will check these Areas and use the barcode present there. If no Areas are selected, kingfisher will use the first barcode it encounters in the PDF document.

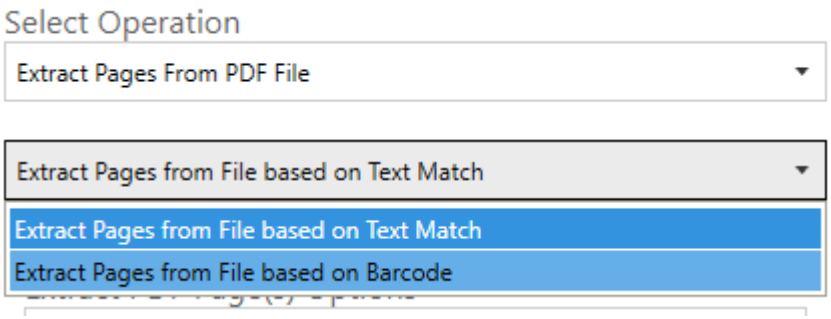
3.1.2 Rename Based on Text Match

This step is like the step above except that it uses text in the PDF document to rename the file.
Unlike the barcode step, an Area must be selected for this step to work. You can also specify Regular Expressions and Words that appear before/after the word you want the file to be named to. These extra features are to help you validate the text you select.

3.1.3 Rename Based on Barcode and Text Match

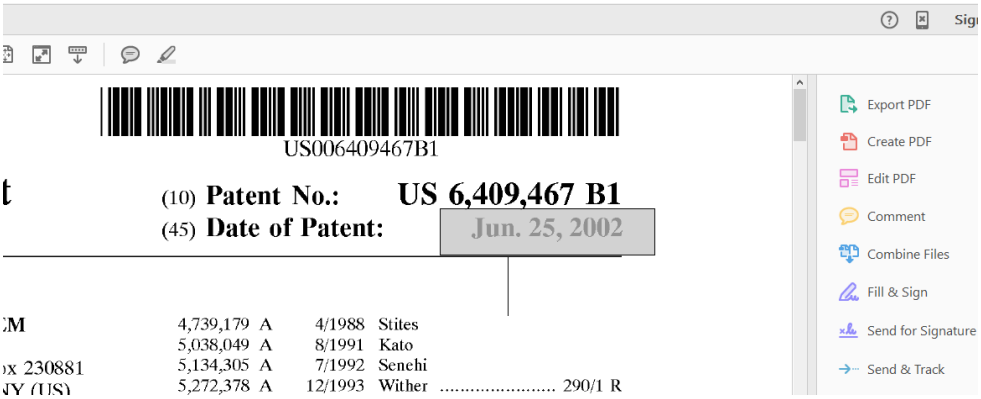
This step is a combination of the two steps above, it allows a user to specify barcodes and text on a PDF page that the file can be renamed to.

3.2 Extract Pages from PDF File

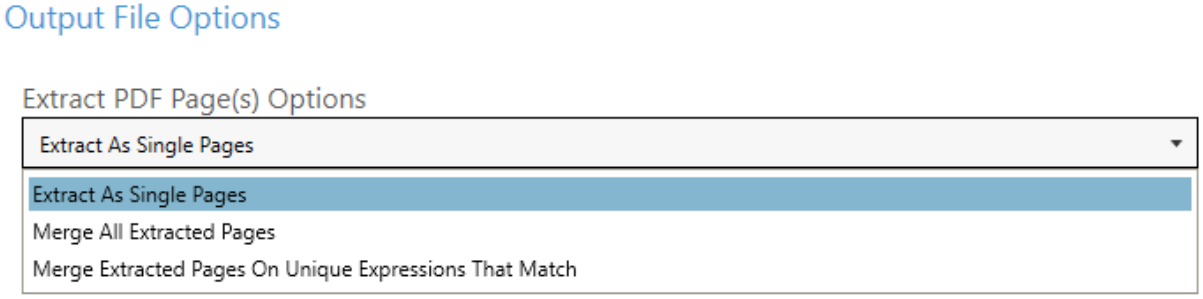


3.2.1 Extract Pages from File based on Text Match

This step allows you to extract single pages form a PDF file based on the presence of a text value in a zone. For example, in the image below you might want to extract any page that has a date value (Jun. 25, 2002) at that location, all you need to do is to is to set up an “Extract Pages from File based on Text Match” job, select the zone and provide some pattern matching to make sure a date is extracted.
Kingfisher also give you the ability to rename the output file base on the text extracted from the PDF page.



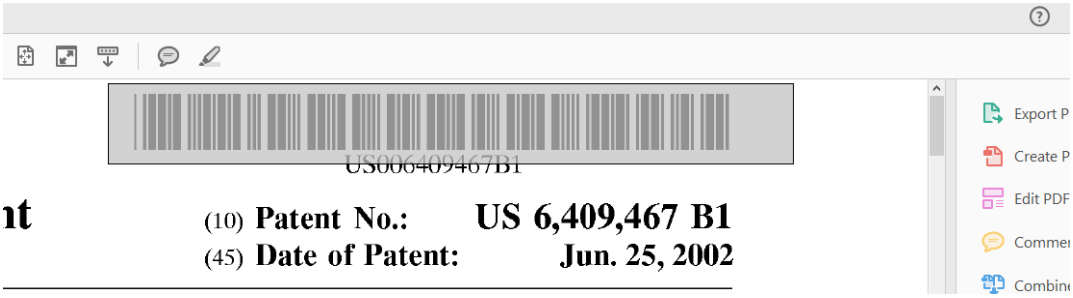
The extract step has the following settings



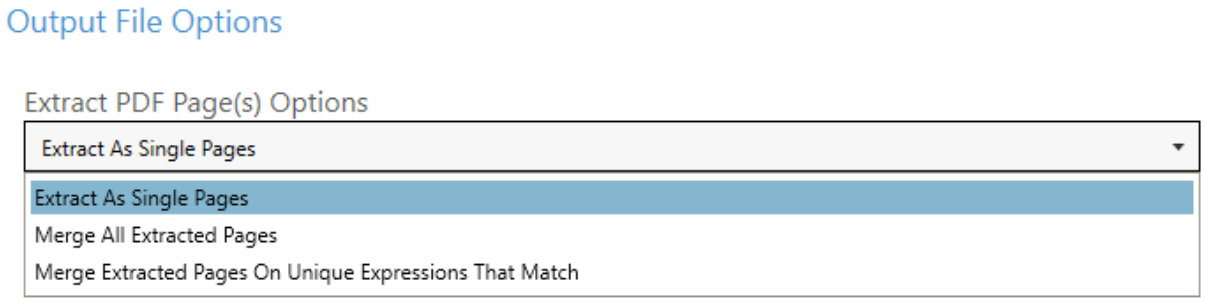
Screen Field / Button	Description
Extract as Single Pages	This option copies all the extracted pages as single page PDF files to the output folder.
Merge All Extracted Pages	Merges all the pages that were extracted into one document.
Merge Extracted Pages on Unique Expressions That Merge	Merges All the Extracted pages that have equal “%VALUE1%” values

3.2.2 Extract Pages from File Name based on Barcode

This step allows you to extract single pages form a PDF file based on the presence of a barcode value in a zone. For example, in the image below you might want to extract any page that has a barcode at that location, all you need to do is to set up an “Extract Pages from File based on Barcode” job and select the zone you are interested in. If no zone is selected the first barcode found in a page will be used.
Kingfisher also give you the ability to rename the output file base on the value extracted from the PDF page.

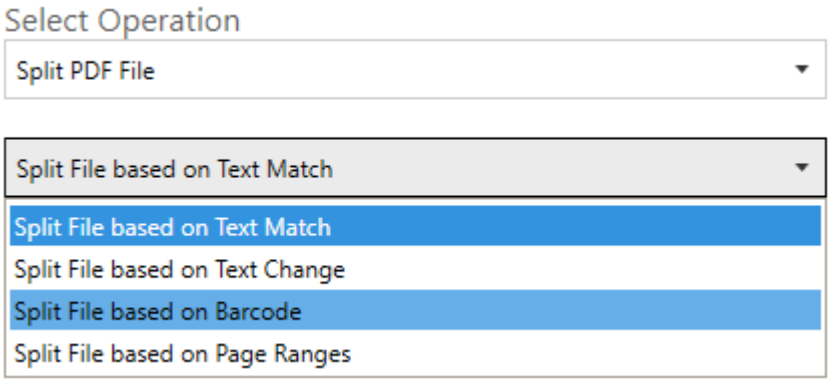


The extract step has the following settings



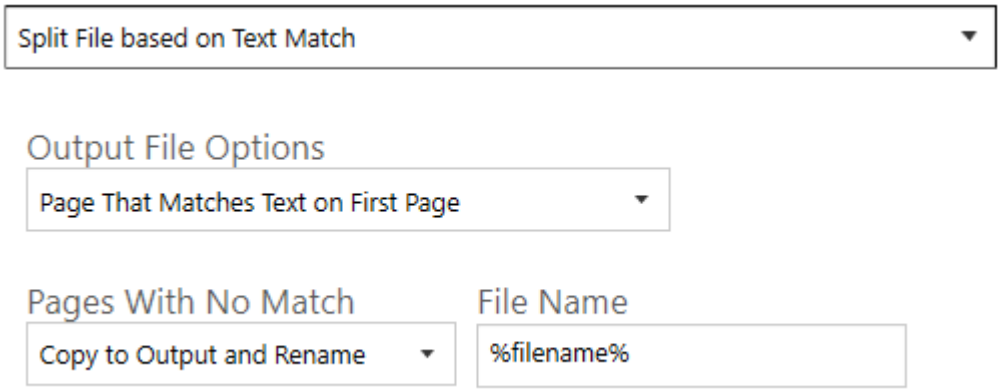
Screen Field / Button	Description
Extract as Single Pages	This option copies all the extracted pages as single page PDF files to the output folder.
Merge All Extracted Pages	Merges all the pages that were extracted into one document.
Merge Extracted Pages on Unique Expressions That Merge	Merges All the Extracted pages that have equal “%VALUE1%” values.

3.3 Split PDF File



3.3.1 Split File based on Text Match

The split option is like the extract option except that instead of extracting a single page, it will split the file into pages using the pages that matches the text value as a divider.

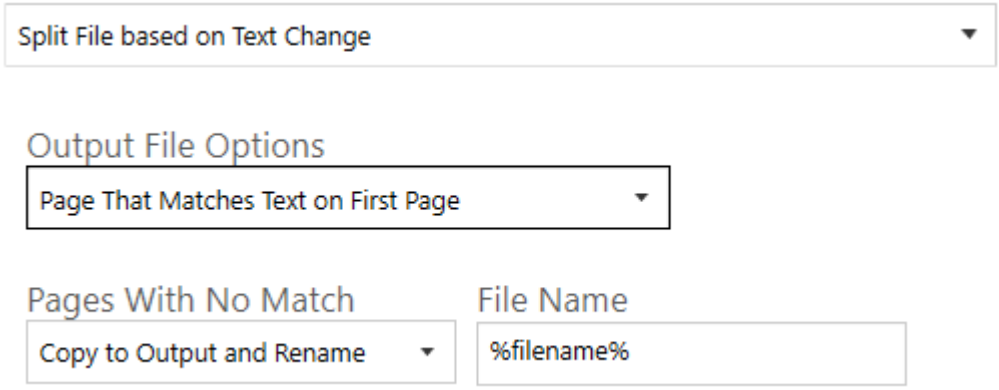


Screen Field / Button	Description
Pages That Matches on First Page	This option puts the page that matches the text value as the first page in the output file.
Pages That Matches on Last Page	This option puts the page that matches the text value as the last page in the output file.
Remove Page That Matches Text	This option removes the pages that match the text expressions in the output files.
Pages with No Match	This gives the user three options of what to do with pages that do not match the text or barcode criteria specified by the user. • Do Not Copy to Output • Copy to Output • Copy to Output and Rename
File Name	If the “Copy to Output and Rename” option is selected you will need to provide a filename mask like the one in the “File Naming” tab, the table below gives you a breakdown of the variables available.

	%FILENAME%	The original file name without the .pdf extension.
	%BOOKMARK%	This represents the bookmark name and can be used with the “Split by Top Level Bookmarks” split option only.
	%TIME%	The current Time.
	%DATE%	The current Date.
	%RUNDATE%	The Time at which the job was executed, the difference between this and the %TIME% variable is that this value will be same for all files processed in a single job run.
	%JOBNAME%	The Name of the Job.

3.3.2 Split File based on Text Change

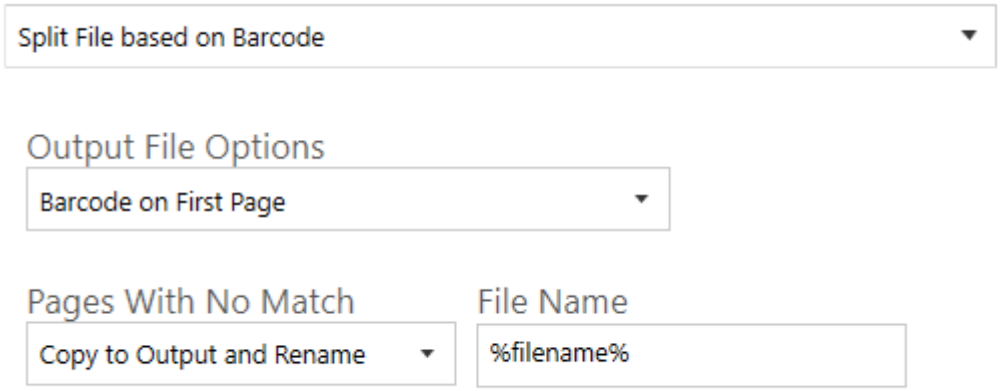
The split option is like the “Split File based on Text Match” option except that instead splitting every where there is a text match, it waits until the text that was matched has changed from the one on the previous page.



Screen Field / Button	Description
Pages That Matches on First Page	This option puts the page that matches the text value as the first page in the output file.
Pages That Matches on Last Page	This option puts the page that matches the text value as the last page in the output file.
Remove Page That Matches Text	This option removes the pages that match the text expressions in the output files.

3.3.3 Split File based on Barcode

The split option is like the extract option except that instead of extracting a single page, it will split the file into pages using the pages that matches the barcode value as a divider.



Screen Field / Button	Description
Barcode on First Page	This option puts the page that matches the barcode value as the first page in the output file.
Barcode on Last Page	This option puts the page that matches the barcode value as the last page in the output file.
Remove Barcode Page	This option removes the pages that match the barcode value in the output files.

3.3.4 Split File based on Page Ranges

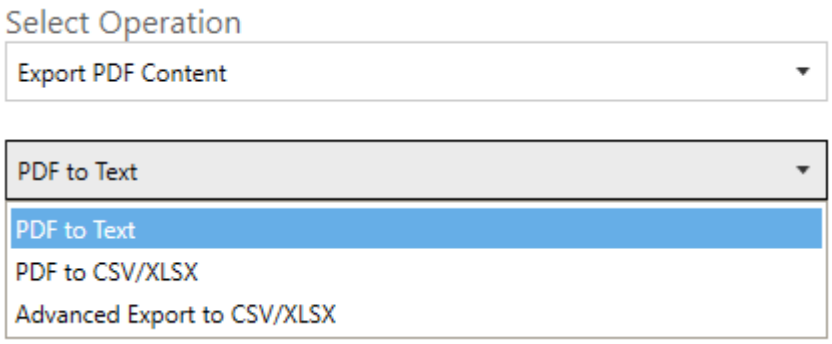
This split operation uses characteristics like page numbers and bookmarks are to split PDFs documents, Has four split options.

- Split into Single Pages: Split the PDFs into single pages
- Split by Top Level Bookmarks
- Split by Page Ranges
- Split by Repeating Page Ranges

<table>
<tr><th>Screen Field / Button</th><th>Description</th></tr>
<tr><td>Split Type</td><td>One of:
Split into single pages
Split by top-level bookmarks
Split by page ranges (See below)
Split by repeating page ranges (See below)</td></tr>
<tr><td>Page Ranges</td><td>Set of page ranges separated by commas that defines which pages from the original should be extracted. The following types of page ranges are allowed:

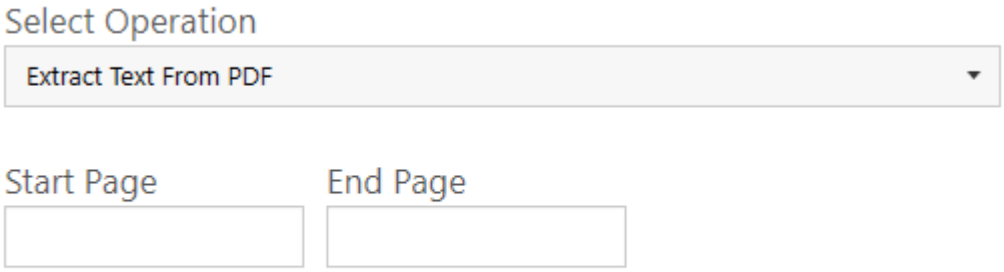
<table><tr><td>1</td><td>Specifies a single page.</td></tr><tr><td>1-3</td><td>Specifies a range of pages.</td></tr><tr><td>10-20even</td><td>Specifies even pages within the range.</td></tr><tr><td>10-20odd</td><td>Specifies odd pages within the range.</td></tr><tr><td>20-lastpage</td><td>Note that lastpage can be used to represent the last page in the document.</td></tr></table></td></tr>
<tr><td>Repeating Range</td><td>Apply the page range at an interval of the repeating number. For example, if 2-4 is specified for page ranges, and 4 is specified as the repeating range, then the range is re-applied every 4 pages.
Hence the file is split such that the first output file contains pages 2-4 from the original document, the second contains pages 6-8 and so on.</td></tr>
</table>

3.4 Extract Content from PDF File



3.4.1 Extract Text from PDF File

This step simply extracts all the text in a PDF file, Kingfisher is intelligent enough to detect image PDF pages and OCR before extracting any text from it. The only type of files we can't extract meaningful text from by default are the ones with font encoding, we advise users to switch OCR for these file types.



Screen Field / Button	Description
Start Page	Page number of the page you want Kingfisher to start extracting text from.
End Page	Page number of the page you want Kingfisher to stop extracting text from.

3.4.2 PDF to CSV/XLSX

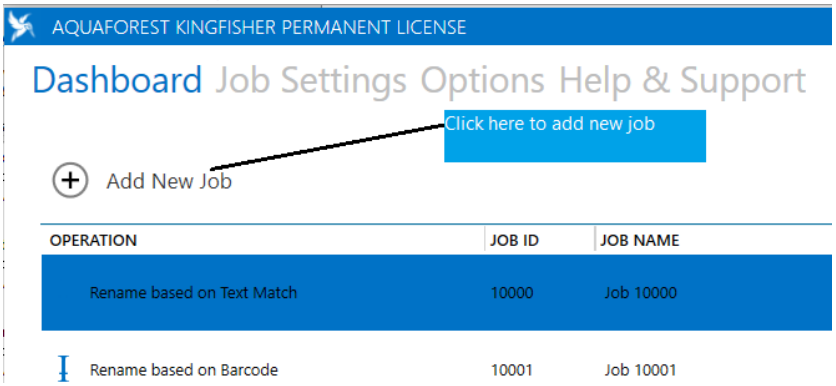
This step is used to extract tabular data from PDF files, see section 5 for more details.

3.4.3 Advanced Export to CSV/XLSX

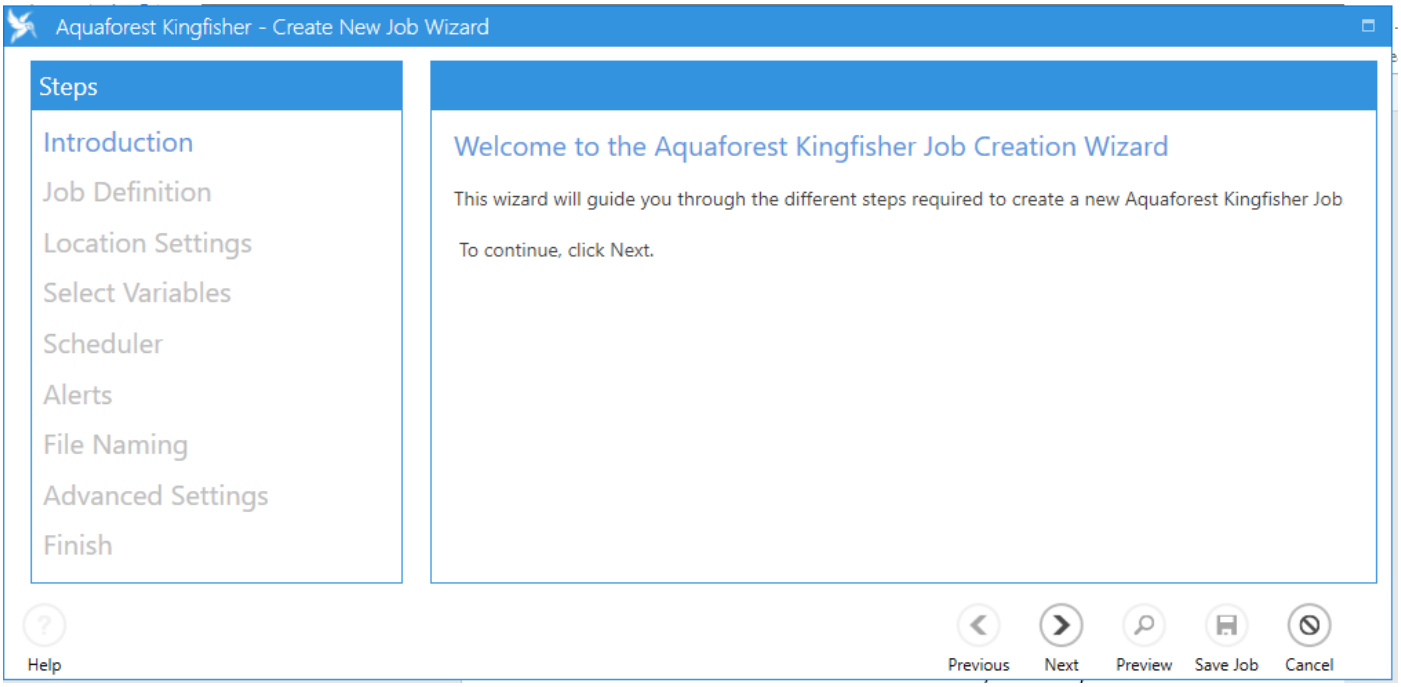
This step extracts text that appears before/after certain expressions, see section 6 for more details.

4 CREATING AND RUNNING JOBS WITH THE GUI

4.1 Creating a new Job

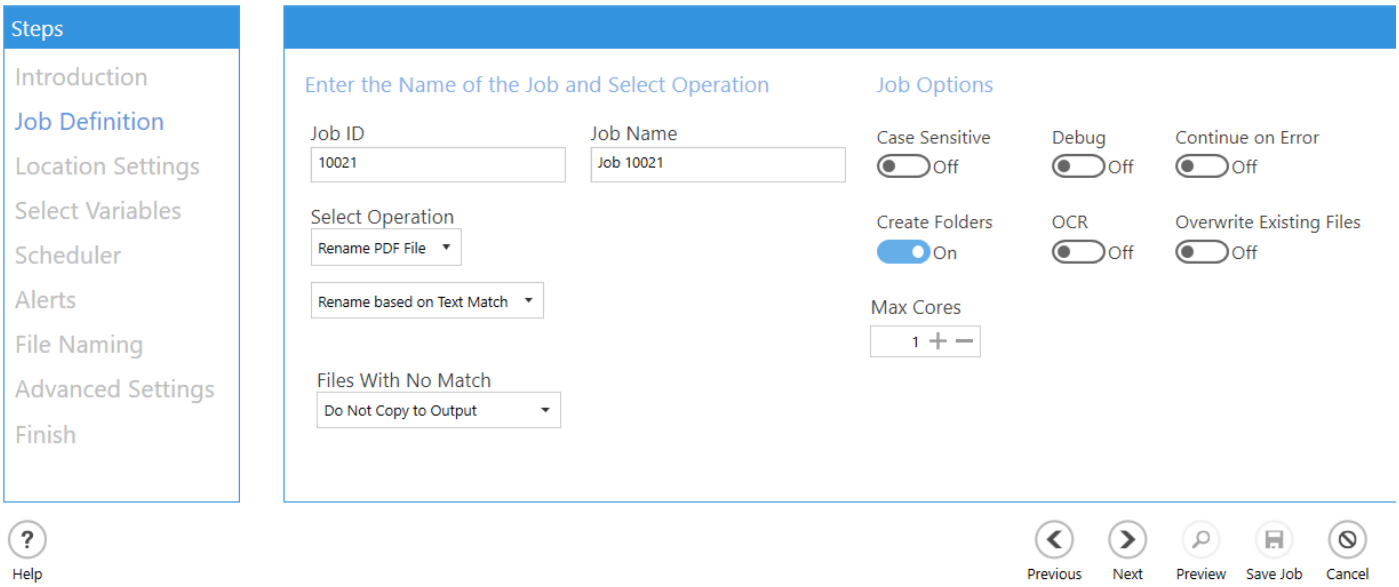


To create a new job in Kingfisher, click on the "Add New Job" button, this will launch the new job wizard shown below.

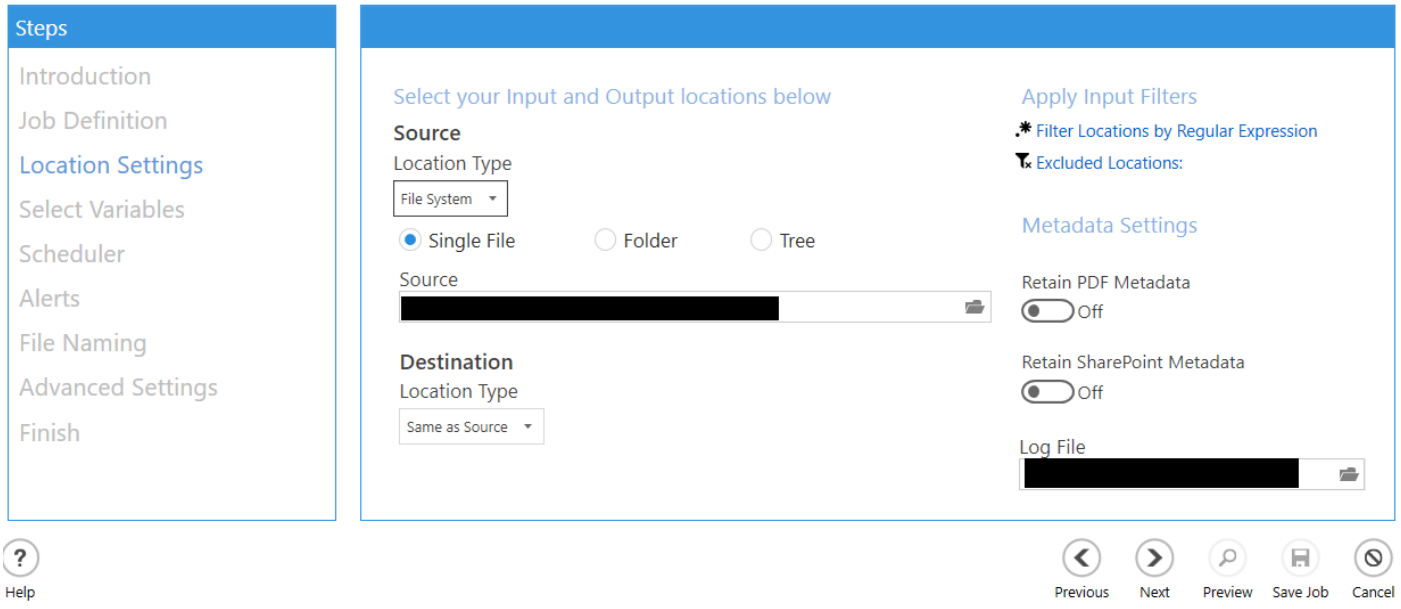


Follow the steps below to create a new job

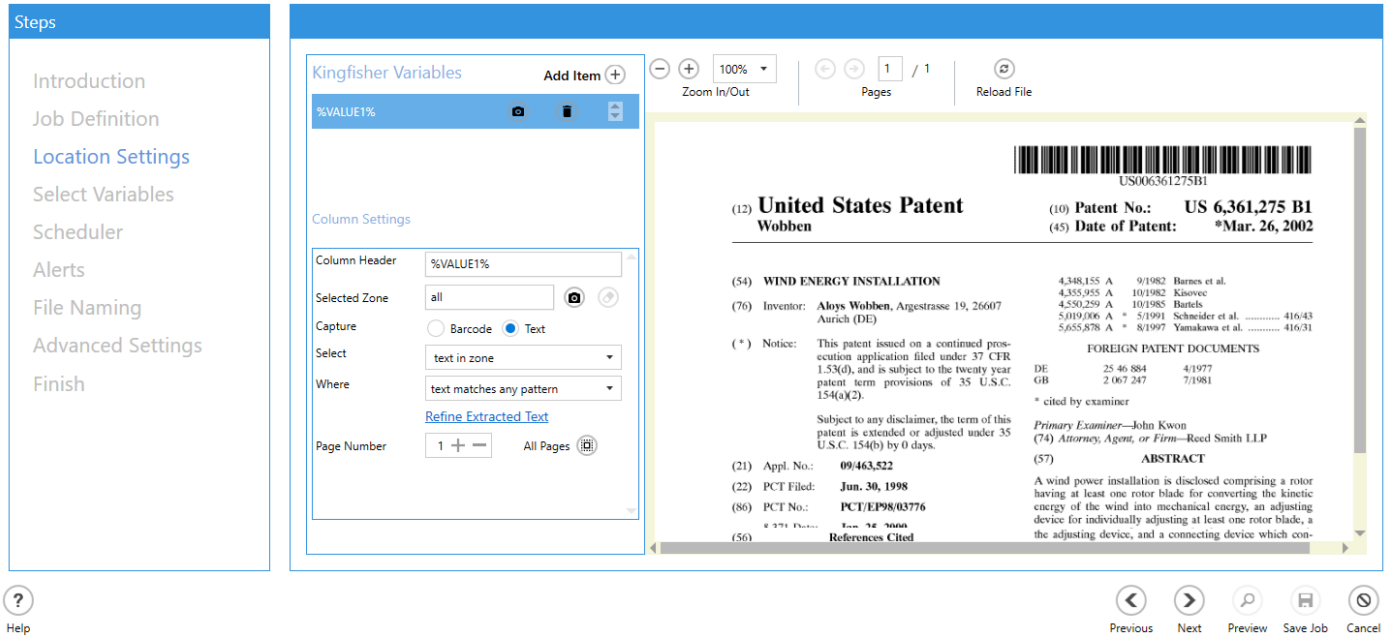
1. Click on the next Button at the bottom of the wizard, this takes you to the “Job Definition Tab”. Check section 2.1.3 for explanation of the fields in this window, choose the suitable options for the job and press next.



2. Click on the next Button at the bottom of the wizard, this takes you to the “Location Settings Tab”. Check section 2.2.2 for explanation of the fields in this window, choose the suitable options for the job and press next.

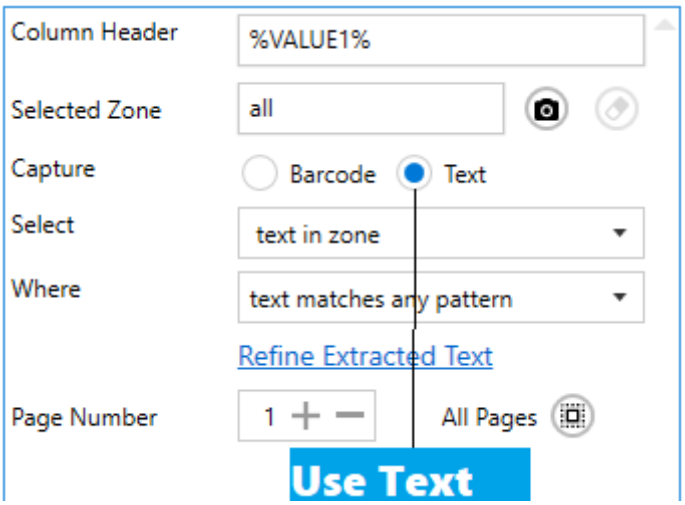


3. One of the main advantages of Kingfisher is the ability to process pdf files based on the file content. To do this, we have created a zone definer to allow users select areas on the pdf page they want to extract text or barcode values from, the “Select Variable” contains the zone definer as shown below.
4.

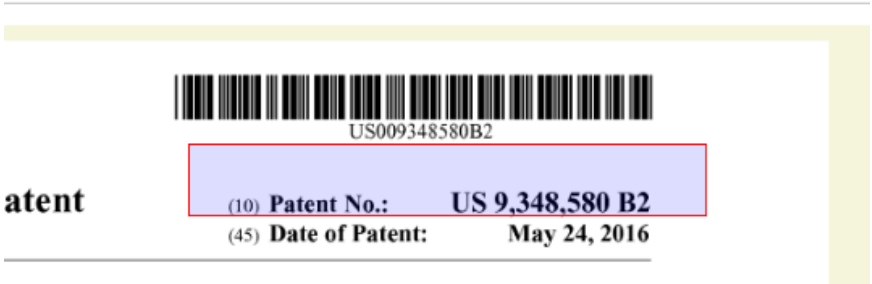


To assign the patent number at the top right corner of the page to the variable named “%VALUE1%”, follow the steps below

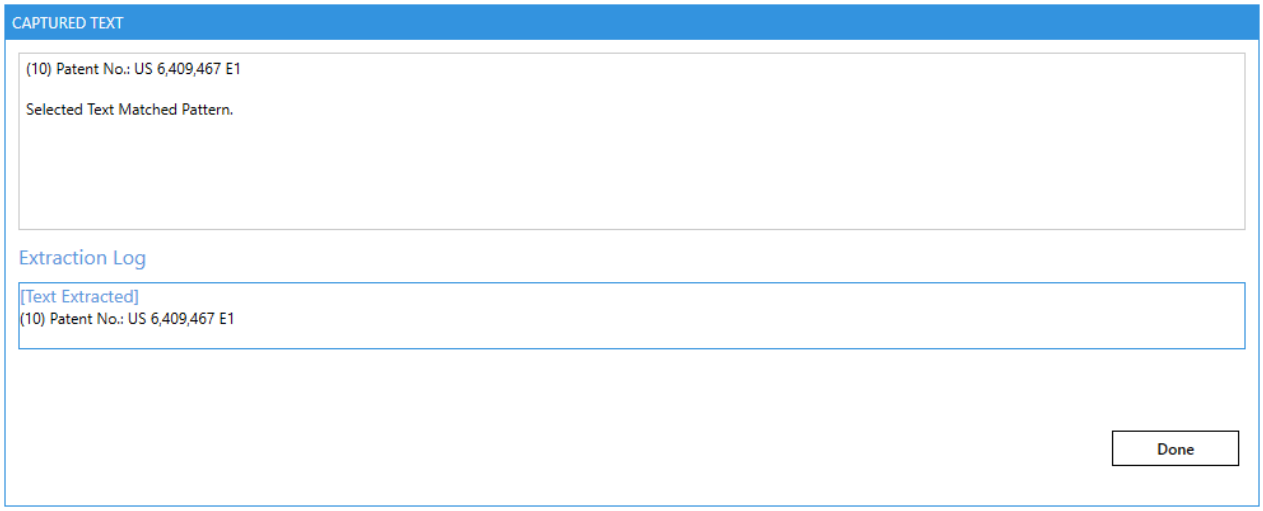
- Select the “Use Text” option associated with “%VALUE1%”.



- Now select the zone at the top right corner

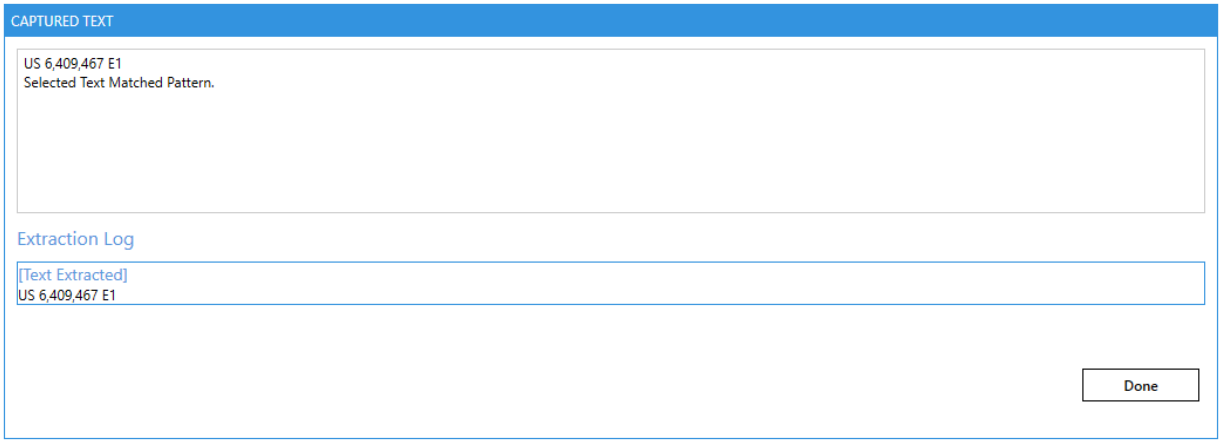


- Click on the “Camera Icon” in the line on the “Selected Zone” Text Box, you will see the text below, if you want to extract the number alone you can refine your option further by clicking the “settings” button.



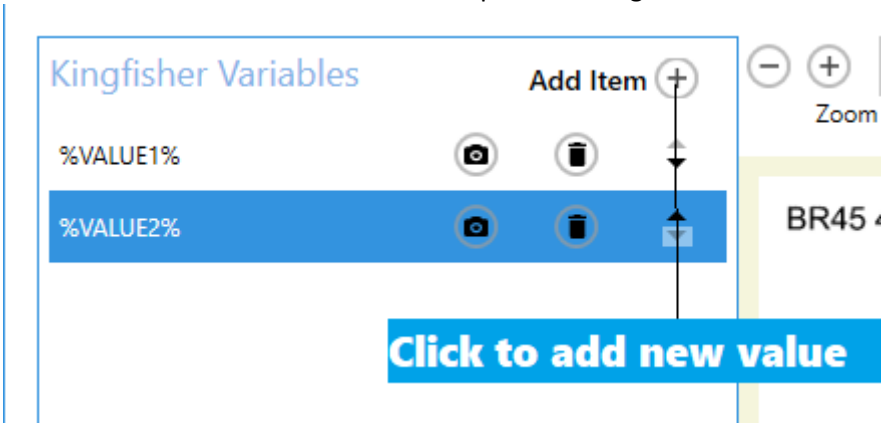
- Click on the Select text in zone combo box, choose the “Select all text in zone after value” option and provide “Patent No.:” as the value.

- Click on capture again to make sure the changes you made worked.

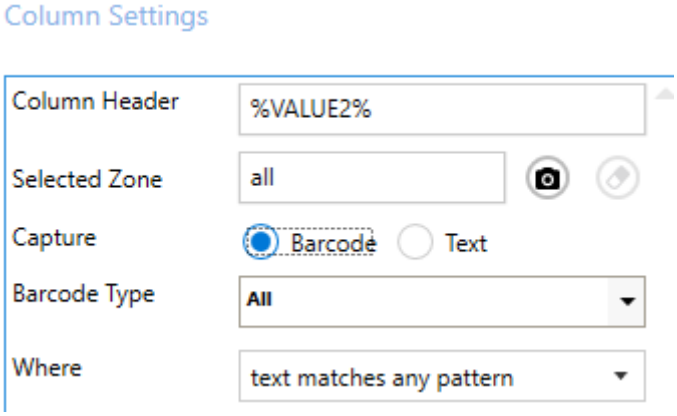


To assign the barcode at the top right corner of the page to the variable named “%VALUE2%”, follow the steps below

- Click on the “Add Item” Button at the top of the “Kingfisher Variables” Group Box.

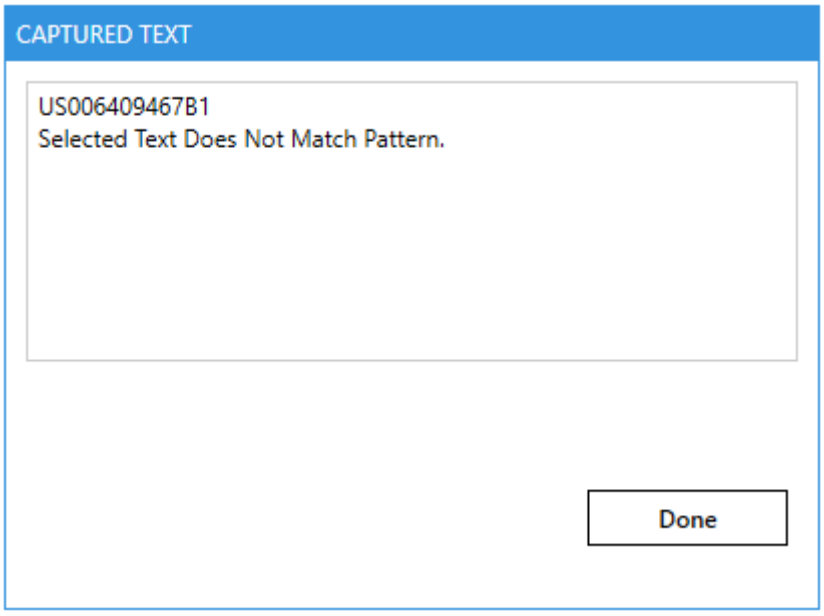


- Select the “Barcode” option associated with “%VALUE2%”.

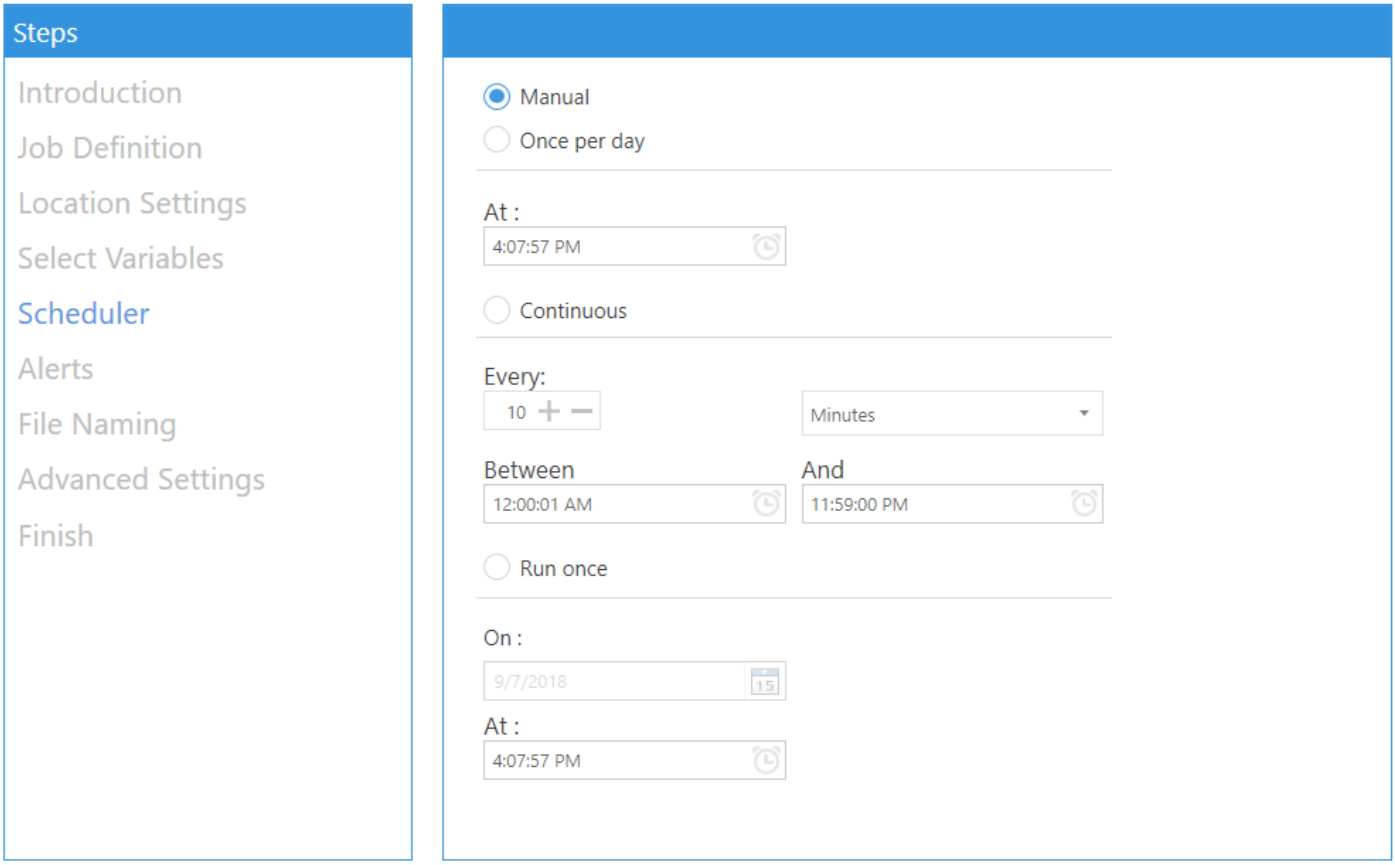


- Now select the zone at the top right corner

- Click on the “Camera Icon” button, you will see the text below

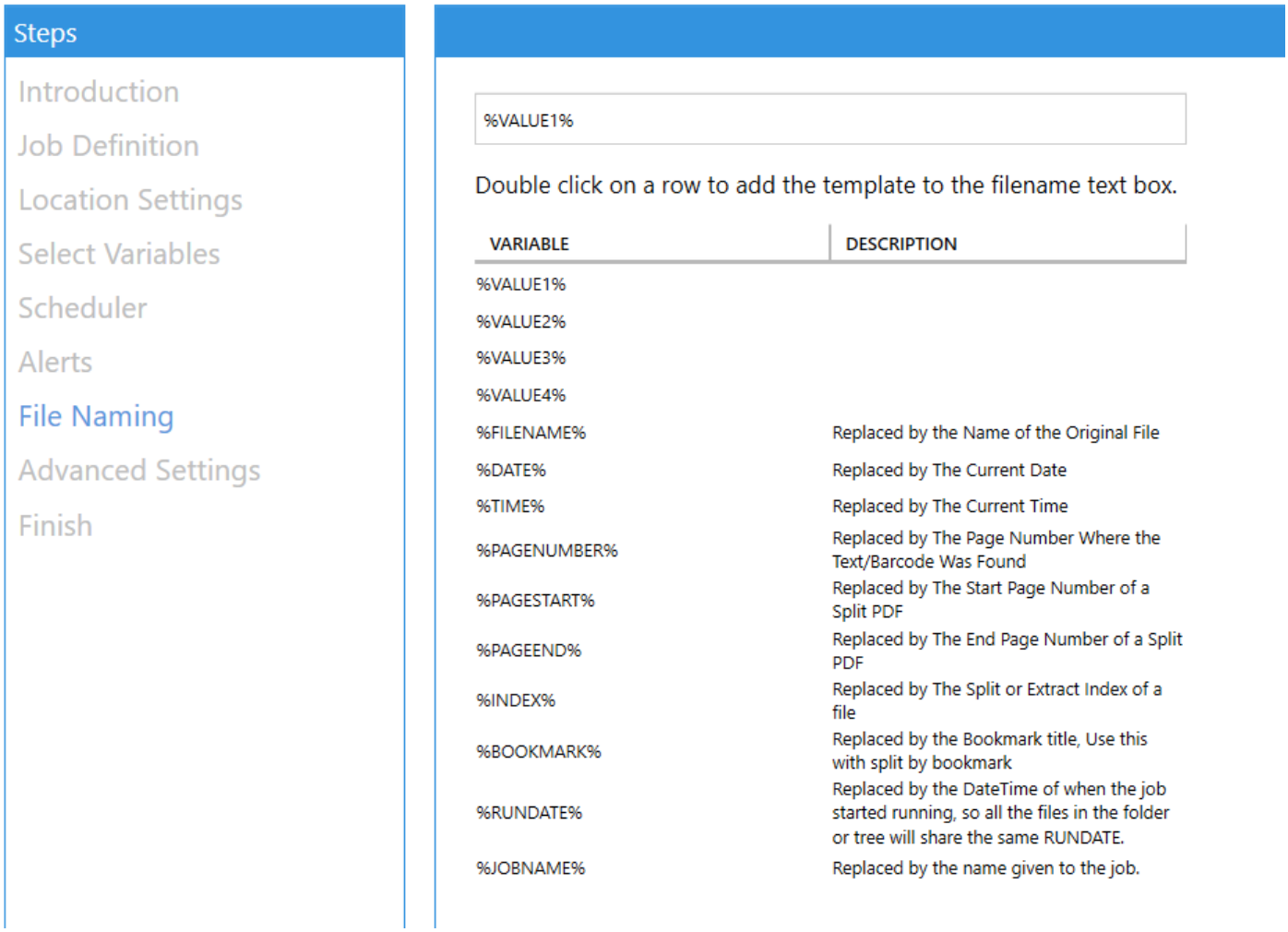


5. Set up the scheduler, we will use manual for this test.

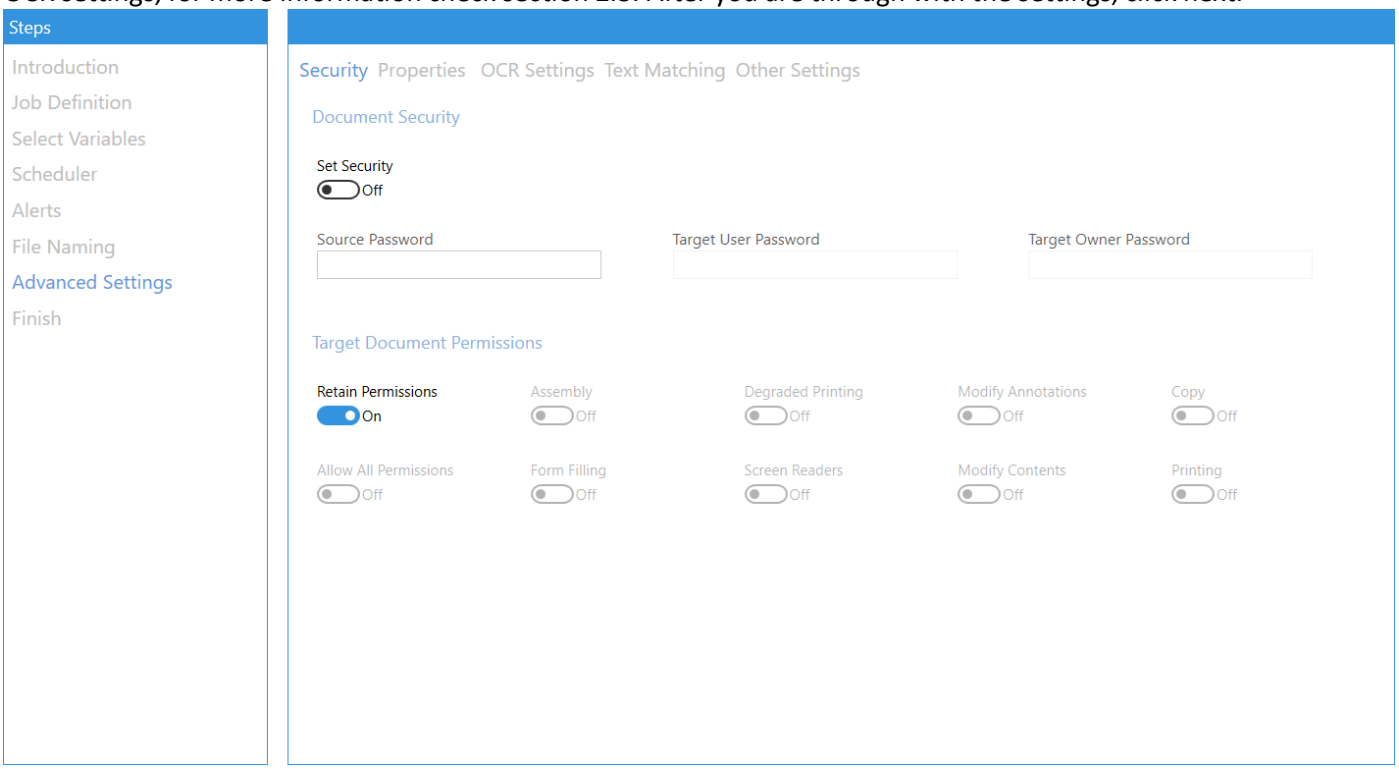


6. Set the alerts, we will not use alerts for this example.

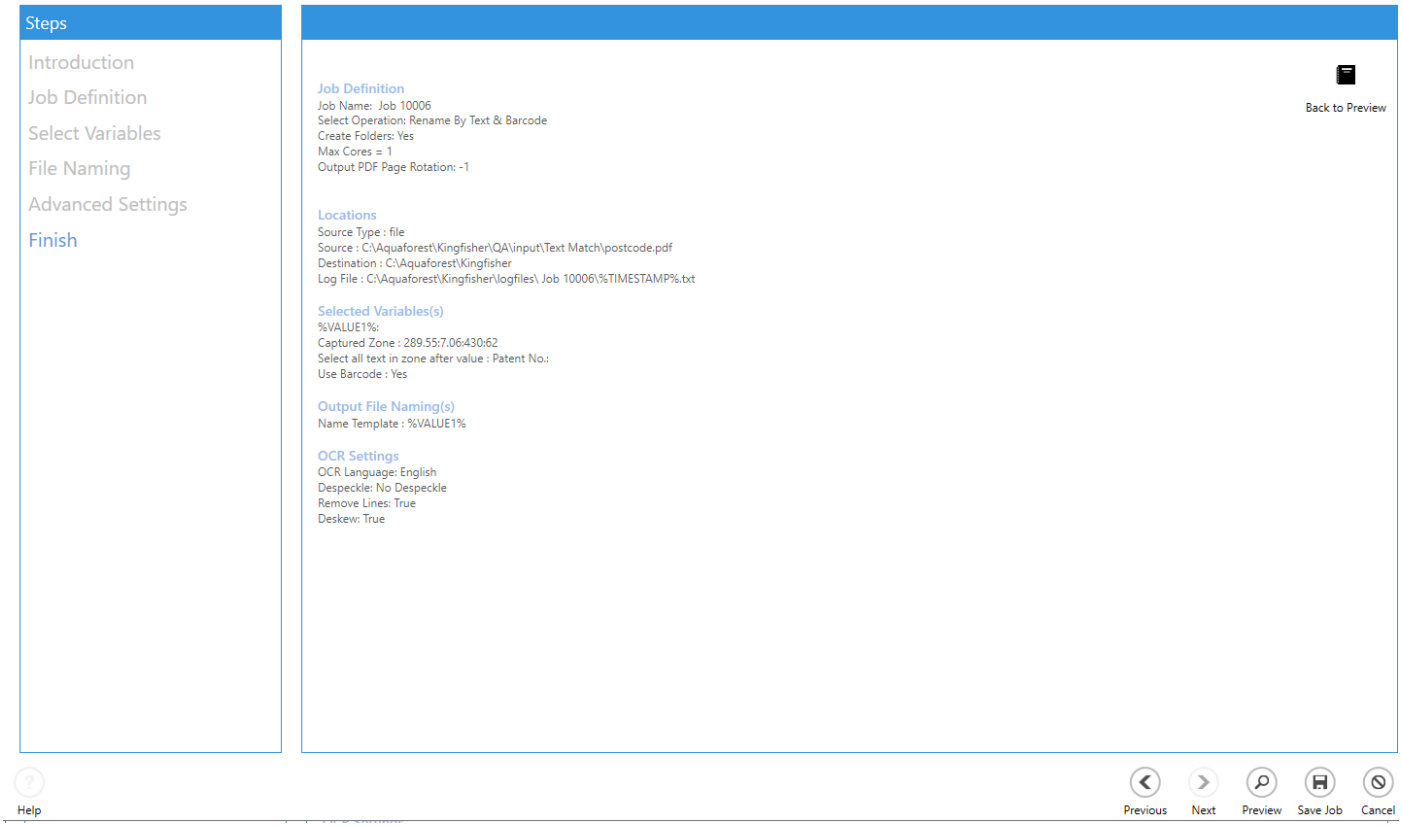
7. After setting the “Scheduler” and “Alerts”, click next to go to the “File Naming” tab, this allows you to specifiy a template for the output file, for this example, we are goin to use the two variables we selected in the previous tab, “%VALUE1%_%VALUE2%”, after you choose a suitable template, click the “next” button.



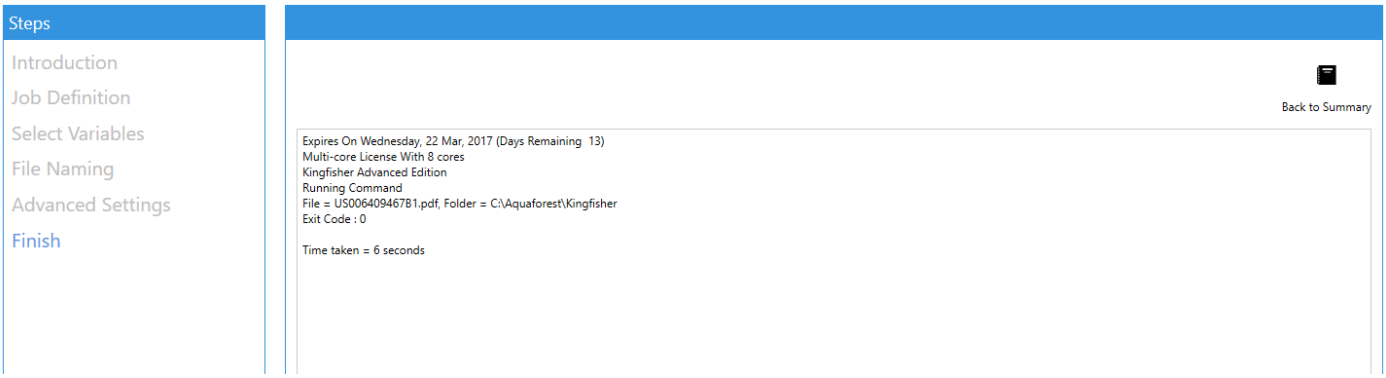
8. The “Advanced Settings” tab allows the user to provide properties of the output PDF file, it also provides a few OCR settings, for more information check section 2.3. After you are through with the settings, click next.



9. The “Finish” tab will show you a summary of the job you set up



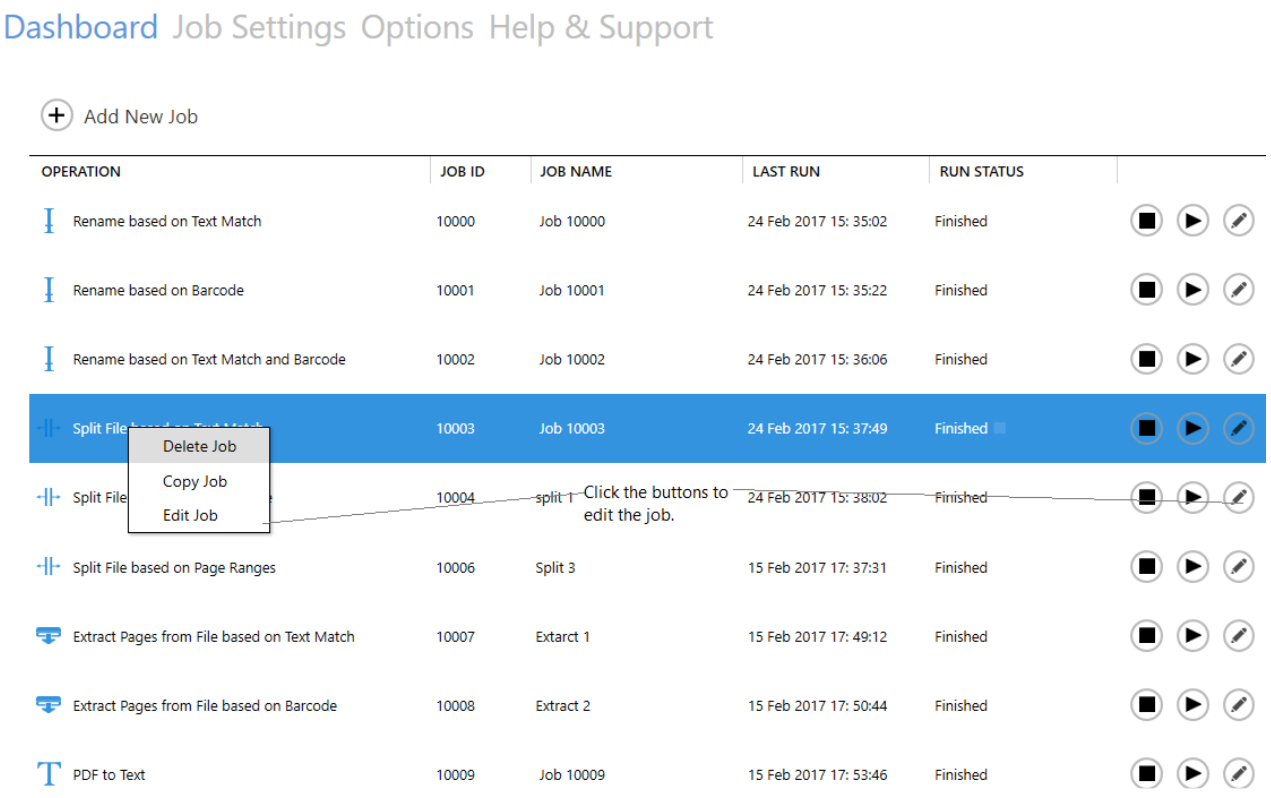
10. To confirm that you are happy with the job, click the preview button to see what the outputs generated will look like, note this does not with the “Extract PDF Contents” step.



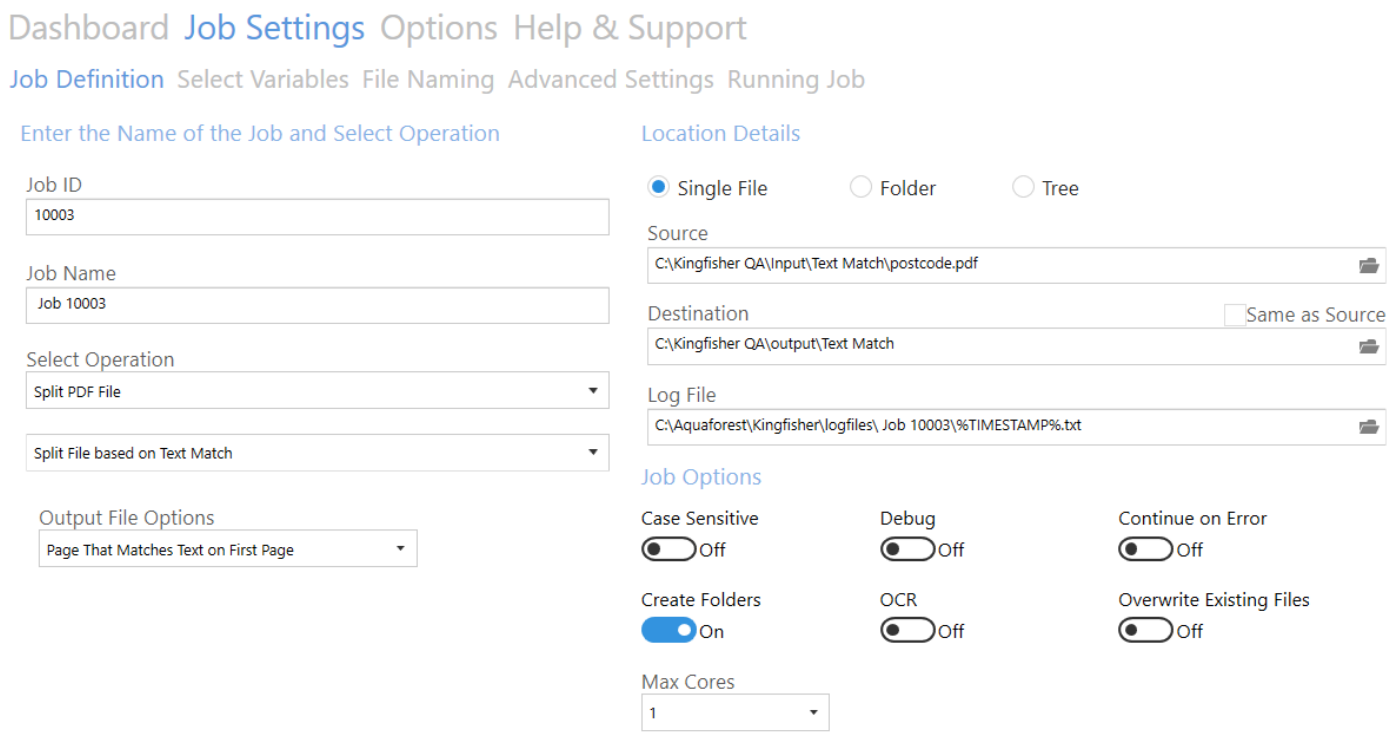
11. Click on the “create” button to create the new job.

4.2 Editing a job

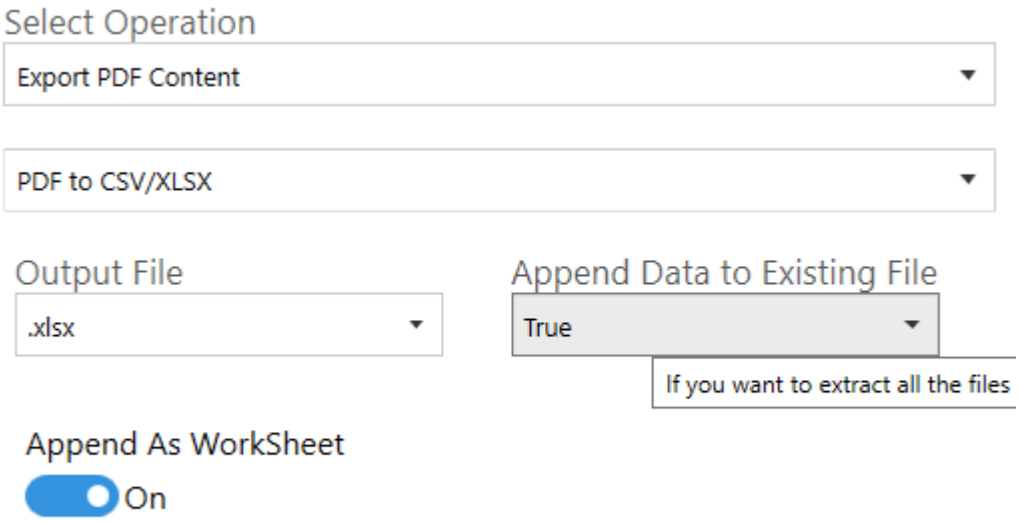
To edit a job in Kingfisher, you should go to the “Dashboard”, select the job you want to edit and go to the “Job Settings” tab, you can also achieve this by double clicking on the job you want to edit, you can also edit a job by clicking the edit button beside the Job name or by clicking the edit option in the context menu as shown below.



The “Job Settings” tab is explained in section 2.2 it contains all you need to know about editing jobs in Kingfisher, after making changes to the job, click on the save button to update the Job definition file.



5 EXTRACT TABULAR DATA FROM PDF



Screen Field / Button	Description
Output file	• csv: Produces a simple csv file. • xlsx: Produces and Excel file.
Append Page Data to Existing File	If set to true, instead of overwriting the output file, kingfisher will append the contents to the end of the current file.
Append as WorkSheet	When on, this will make kingfisher add an extra worksheet for each PDF file, instead of appending the data at the end on the first Worksheet.

Aquaforest Kingfisher has the capability of recognizing tables in PDF files with very minimal user intervention. The tabular data it extracts is usually written to a csv or xlsx file.

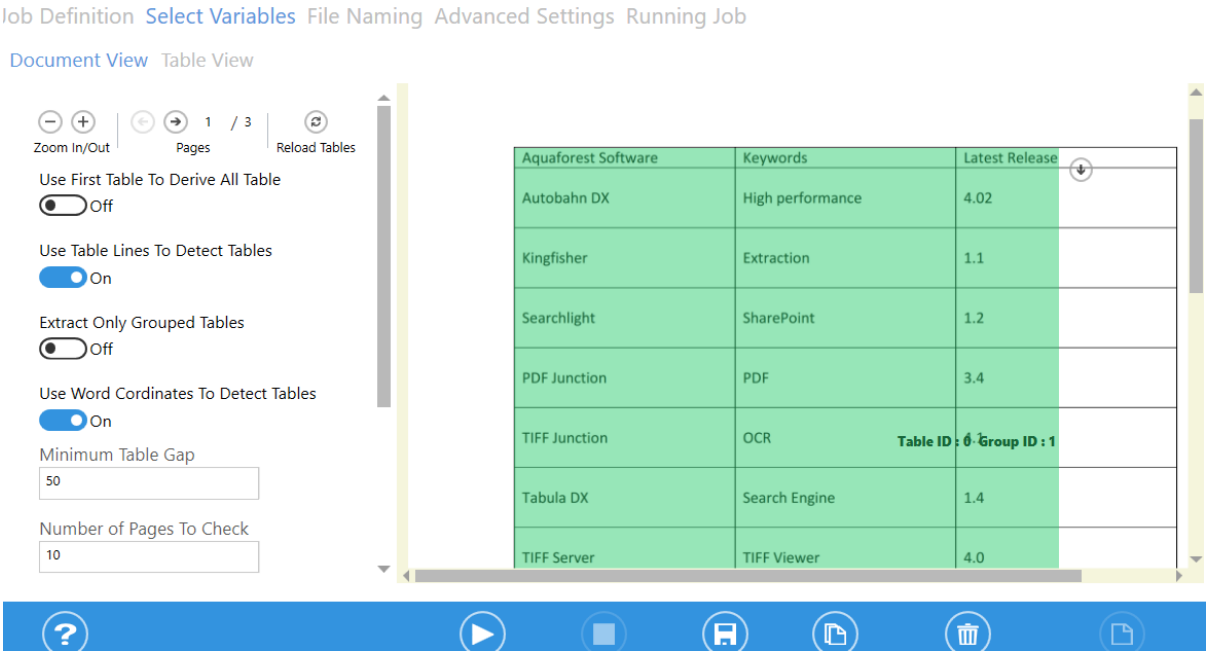
To extract all the tables recognized in the PDF file to the csv file you can just load the input file and run the job. If you want to tailor the extraction further and group the tables to your own liking, you can read the section below.

The Kingfisher Table control is divided into 2 tabs called the “Document View” and the “Table View”

5.1 The Document View

This tab shows a visual representation of the PDF file and the table recognized on the file, the table are usually highlighted. Below is a screen shot showing the Document View tab.

The document view usually shows the table recognized in the first 10 pages, this value can be changed as shown in the table below. It also provides some properties that could be changed to alter the way the tables are recognized, read the table below to see what each property represents.



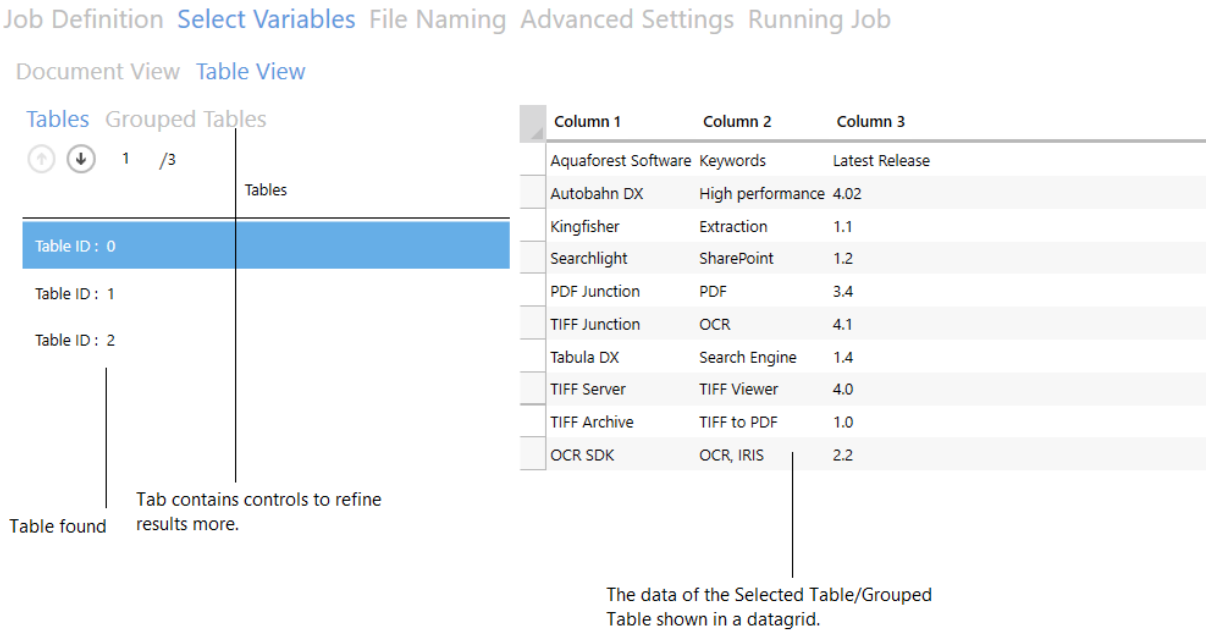
Value	Details
Zoom in/Out	You can use this to zoom in or zoom out of the file loaded in the table viewer control.
Pages	Use this control to move through the pages of the loaded file.
Reload Tables	Clicking this button will make Kingfisher reload the tables in the PDF file, this allows you to change some settings and reload the tables in the file after the file has been processed.
Crop Table	Switching this on will allow the user to specify an area where kingfisher should check for tables, this method produces more accurate result because it eliminants any possible noise.
Do Not Write Header to Output	Switching this on will not write the Group Header to the output file, this is important if you want to write similar table data from multiple files to one output file, it will preserver continuity.
Use First Table to Derive All Tables	This setting instructs Kingfisher to use the first table in the file as a template to recognise all the other tables in the file.
Use Table Lines to Detect Tables	If set to true, Kingfisher will use the PDF graphic line to find the tables in the file.

Use Word Coordinates to Detect Tables	This setting instructs Kingfisher to use the coordinates of words on a page to recognise tables.
Extract Only Grouped Tables	Kingfisher gives the user the ability to group tables depending on different rules, turning this setting on will write only the tables grouped to the csv file.
Number of Pages to Check	The process of table recognition is CPU intensive; thus, kingfisher limits the number of tables displayed in the GUI to the first 10 pages. You can increase or reduce it using the textbox.
Minimum Table Gap (pt)	If the space between two preceding lines is greater than the value given here, that space will be interpreted as a table break.
Table Space (pt)	If the space between two words is greater than this value, kingfisher will recognise that space as a column. Note if the value is 0 or less, kingfisher will ignore this and work out the space automatically.
Minimum Number of Rows	Kingfisher will only recognize tables with a higher number of rows than this value.
Minimum Number of Columns	Kingfisher will only recognize tables with a higher number of columns than this value.

5.2 Table View

The Table View shows the data that was successfully extracted from the PDF file, it is divided into two parts explained below:

- The controls on the left are divided into two tabs named “Tables” and “Grouped Tables”, the tabs. They contain a list of tables found and some other controls for grouping tables.
- The right part contains a Data grid which displays the data in the Table/Grouped Table Selected.



The next section will dive deeper into the “Table View”.

5.2.1 Tables

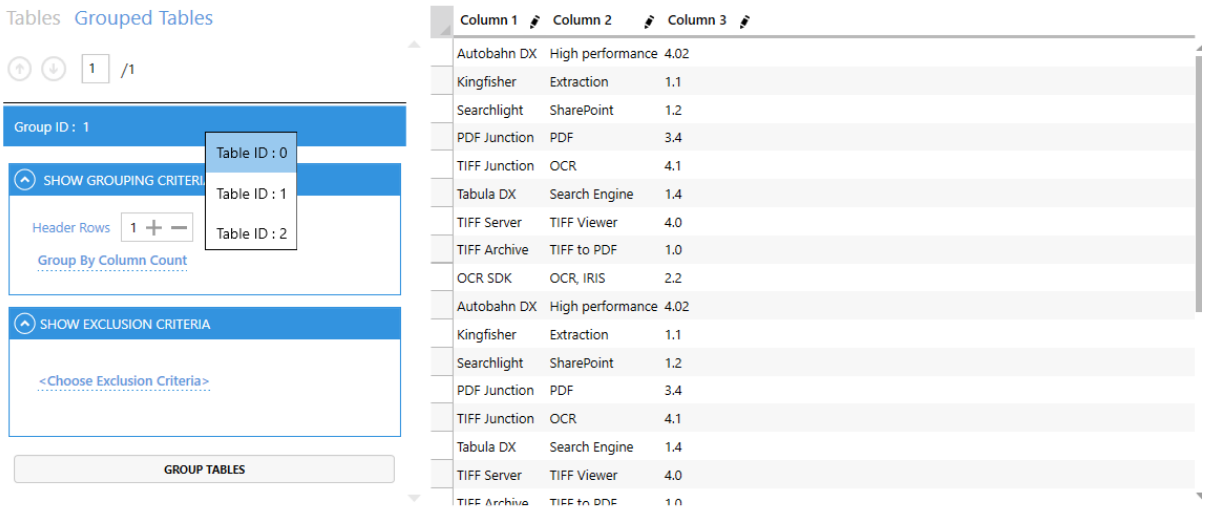
This is a very straight forward tab; it contains only two controls.

Value	Details
UP/Down Control	Use these buttons to select the next table in the list. Note the data grid on the right shows the contents of the selected table.
The table list	This is the second control, it contains a list of all the tables recognized in the document. Clicking on any one of them will load the selected table in the data grid on the right.

5.2.2 Grouped Tables

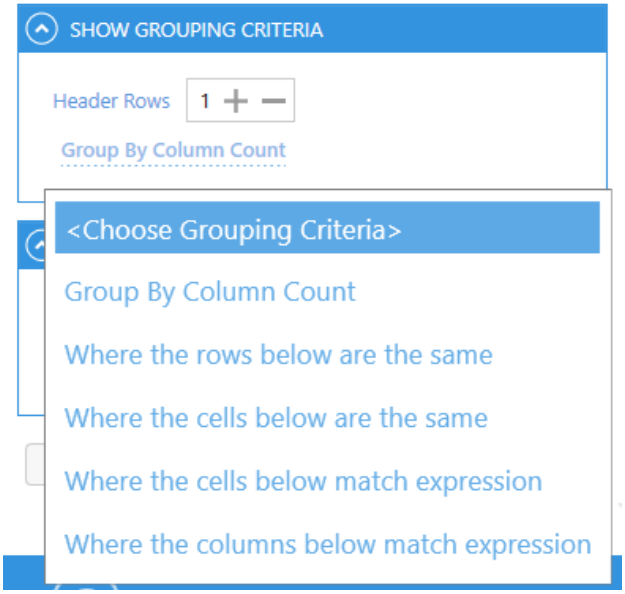
By default, Kingfisher can extract all structured data in a PDF file and write it to a ".csv" or ".xlsx" file. We also understand that users might not want to extract all the tables found in a document or the user might want to group some similar tables into one group with a common header.

Because of the issues raised in the paragraph above, we have decided to give the user the ability to group/exclude tables and edit the headers. The following sections will explain this in more depth.



5.2.2.1 Grouping Criteria

This section explains how users can group various tables based on the grouping criteria offered by Kingfisher.



- Header Rows: If you are writing a group of say 5 tables to a ".csv" file, the table header will be repeated for each table thus the final CSV file won't be very clean. As result, Kingfisher will skip the first "n" rows provided in the header rows control.

5.2.2.1.1 Group by Column Count

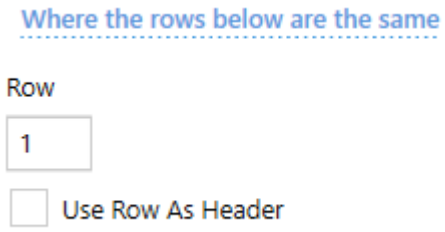
This is a straight forward option; it groups all the tables with the same number of columns. This option will also allow to change the provide the header rows (See 1.2.3).

5.2.2.1.2 Where the rows below are the same

This criterion will check all the tables for the row specified and group the tables based on identical rows. The example below will group all the tables with the same first row as one group.

Unfortunately, you can't save headers for this type of criteria, this is because the table types recognized here are not very predictable and will vary across different files. We decided to disable that feature to avoid putting wrong headers on tables.

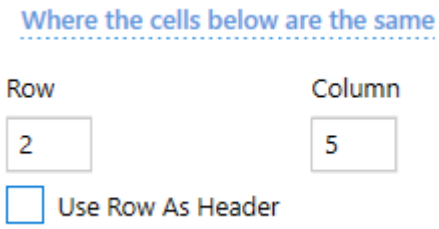
Checking the Use Row as Header Check Box will use the Row Matched as the header row.



5.2.2.1.3 Where the cells below are the same

This is the same as the previous option, the only difference is that it compares a single cell instead of the whole row. Unfortunately, you can't save headers for this type of criteria, this is because the table types recognized here are not very predictable and will vary across different files. We decided to disable that feature to avoid putting wrong headers on tables.

Checking the Use Row as Header Check Box will use the Row Matched as the header row.

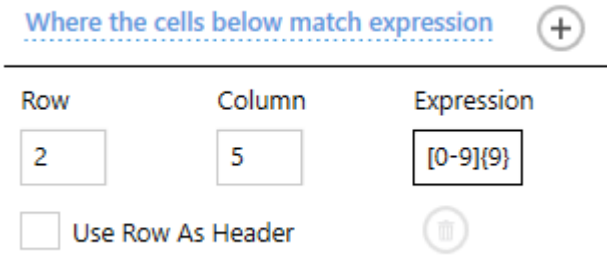


5.2.2.1.4 Where cells below match expression

This option groups all the tables where the cells provided below matches the Regular Expression given. Note, you can use the '+' button to add more items

This option will also allow to change the provide the header rows (See 1.2.3).

Checking the Use Row as Header Check Box will use the Row Matched as the header row.

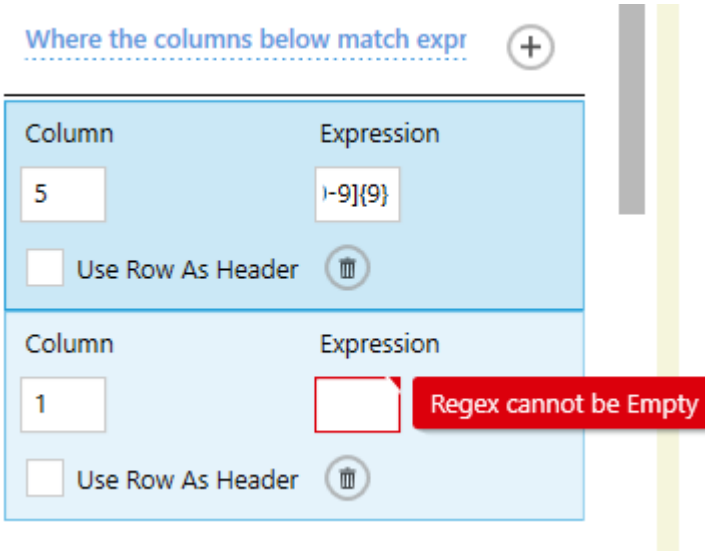


5.2.2.1.5 Where Columns below match expression

This option groups all the tables where the columns provided below matches the Regular Expression given. Note, you can use the '+' button to add more items

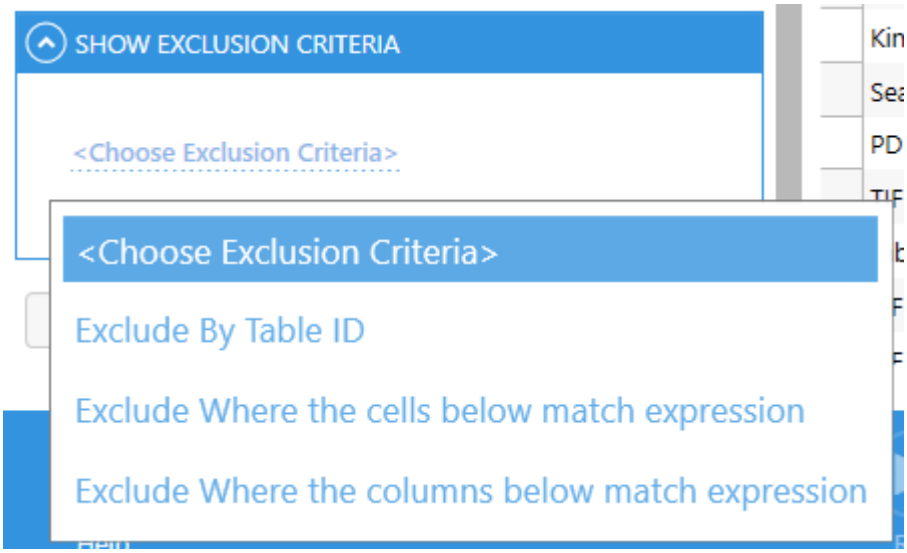
This option will also allow to change the provide the header rows (See 1.2.3).

Checking the Use Row as Header Check Box will use the Row Matched as the header row.



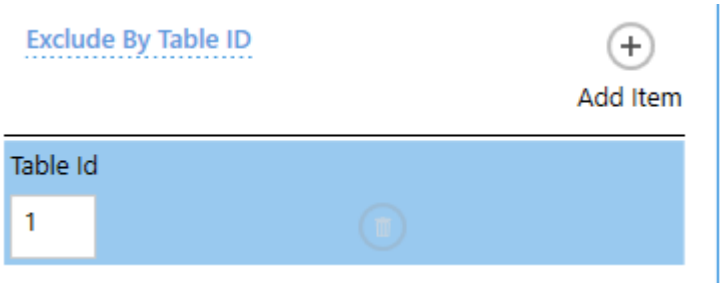
5.2.2.2 Exclusion criteria

Kingfisher also allows the Users to exclude table with characteristics they are not interested in, to use this feature you will have click on the "Show Exclusion Criteria" Expander to view the various option



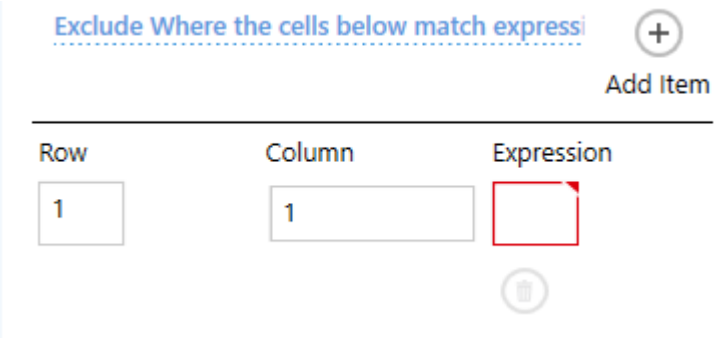
5.2.2.2.1 Exclude by Table ID

This option will exclude all the tables that their ID's are provided below.



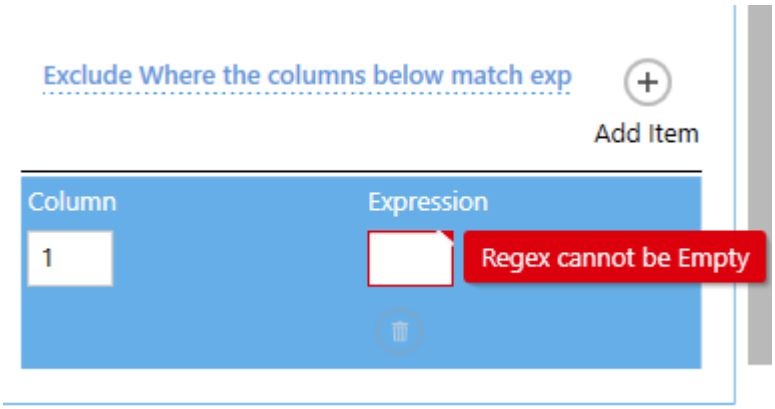
5.2.2.2.2 Exclude cells below match expression

This option excludes all the tables where the cell provided below matches the Regular Expression given. Note, you can use the '+' button to add more items



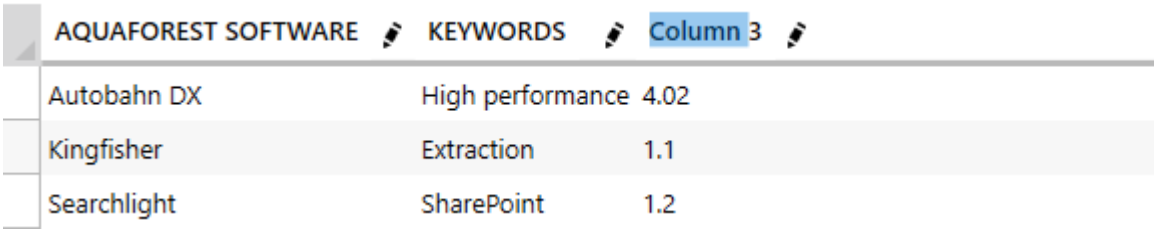
5.2.2.2.3 Exclude Where Columns below match expression

This option excludes all the tables where the columns provided below matches the Regular Expression given. Note, you can use the '+' button to add more items.



5.2.3 Data grid

The main use of the data grid is to show the users the data that has been extracted from the PDF file, you can also use it to Edit headers (If the pen symbol appears next to the header) as shown Below. After editing, just click the save job button.

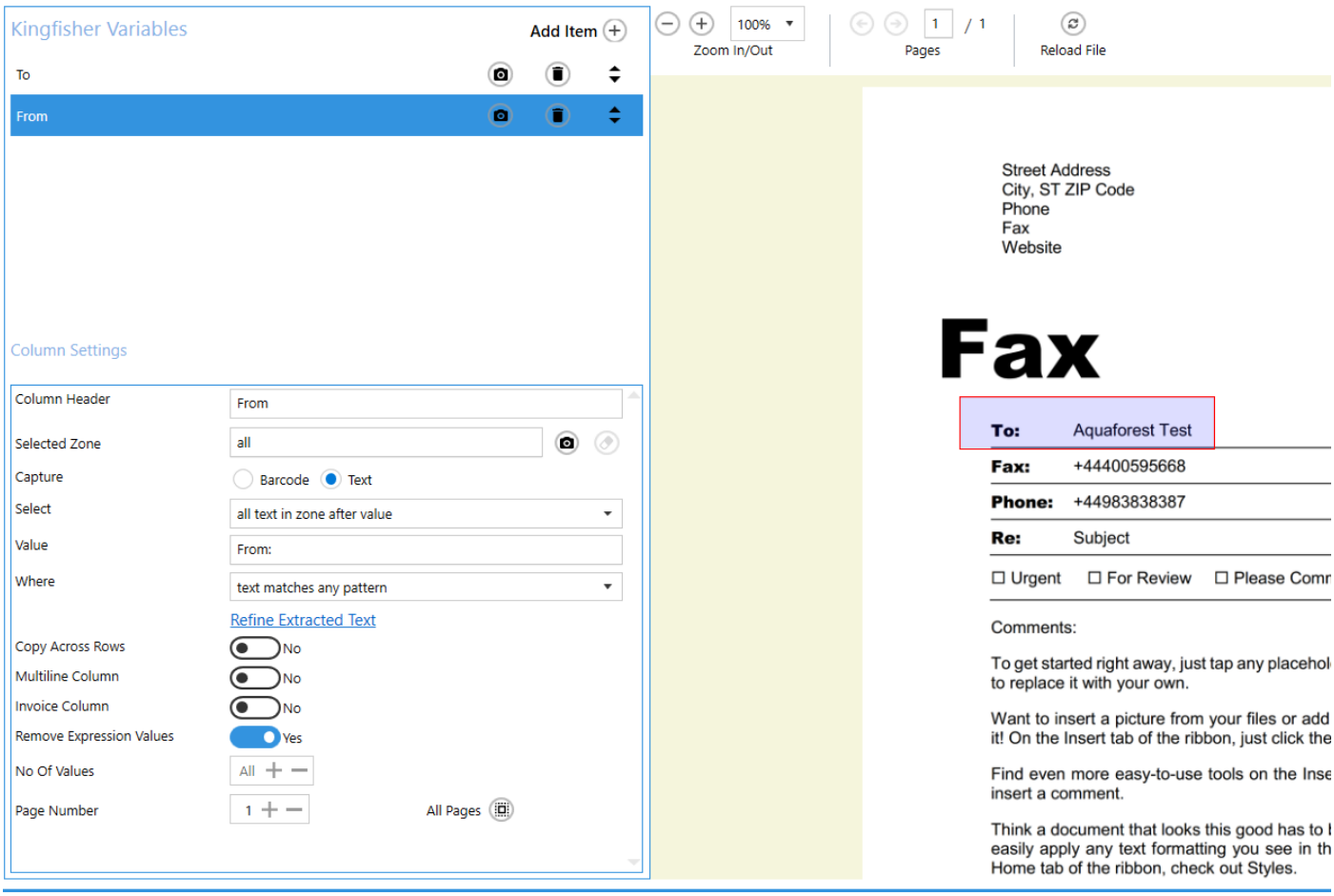


6 ADVANCED EXPORT TO CSV/XLSX

This is step extracts text based on certain keywords in a PDF file. It is used to extract text that appears after or before a certain expression. This can be used to process a collection of invoices, receipts and forms. The impressive part of it is that it uses OCR if need be so users with image PDF files need not worry. The values extracted will be written to a csv file where the expression will be the column header.

Below is a step by step guide to using the Smart Text Extraction

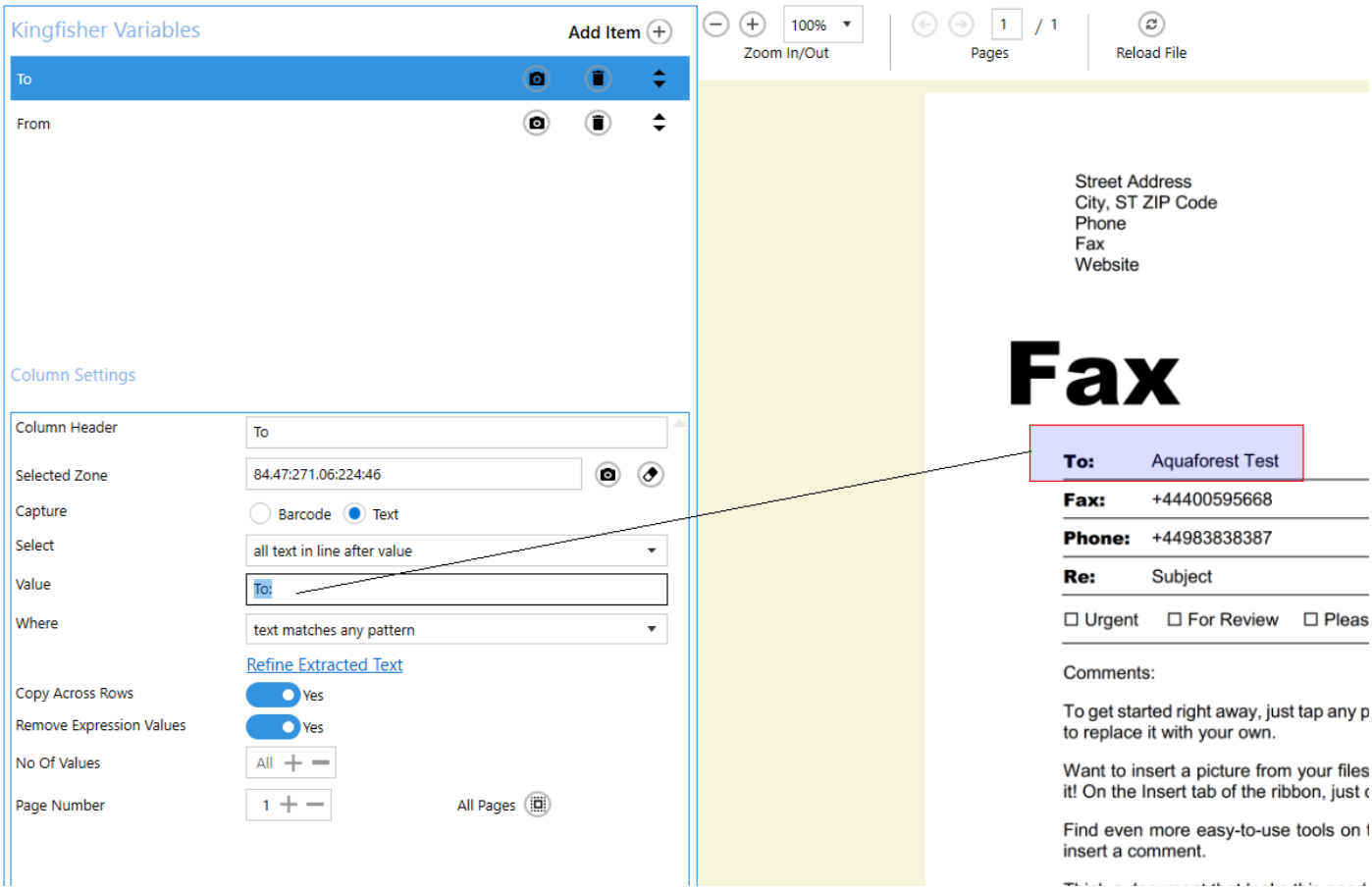
1. Choose the "Export PDF Content" → "Advanced Export to CSV/XLSX" step then go to the "Select Variable" tab. As you can see it is very similar to the select variables tab in section 2.2.2 but for a few extra options explained in the table below...



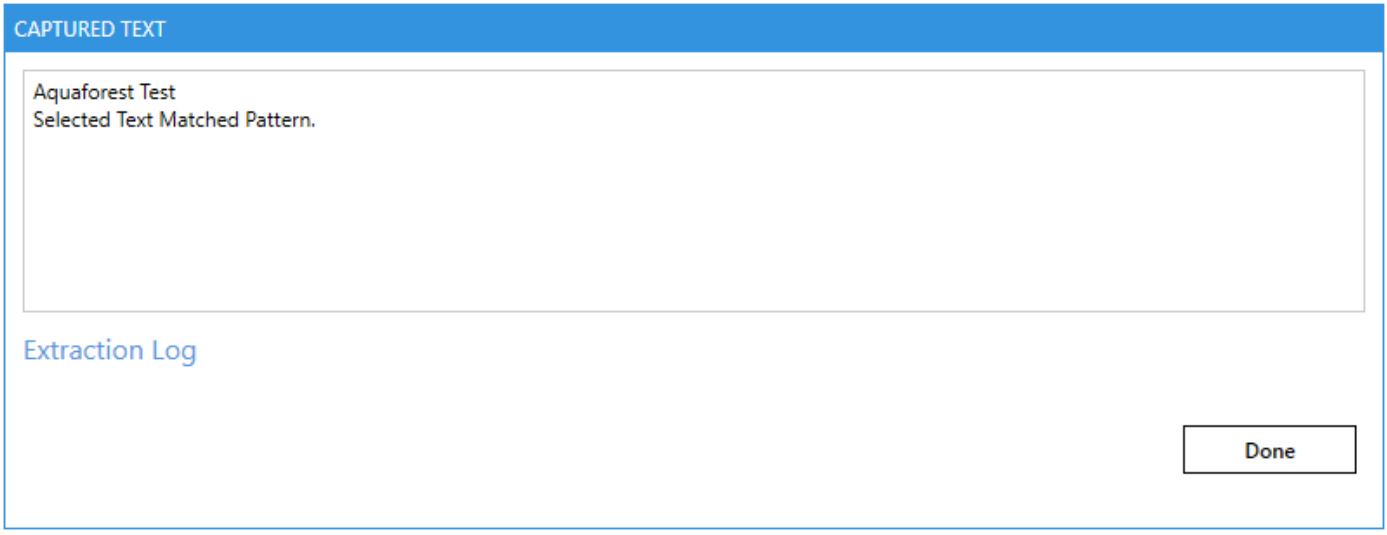
Value	Description
Copy Across Rows	Check this box if you want a value like "Invoice Number" that appears only once in a page to be copied across all rows in that page.
Multiline Column	Check this box if you want the different lines in a zone to be extracted as different values.
Match Multiple Values	Turn this on if you want the pattern matching to return more than one match in a zone.
Invoice Column	Select this option if the current column is part of an invoice, this will make sure kingfisher gets the multiline values. Note Kingfisher will use the first column you add as a marker for the rest in the invoice.

Remove Expressions Values	This option will trim the text extracted to the first occurrence of any "Column Expression Value" in the job
No of values	Say there is more than one occurrence of an expression in the selected zone, this number control will you to set the maximum amount of values that should be extracted per zone.
Pages	This allows the user browse through the loaded pages.

2. Select a zone you want to extract the text from, enter the Expression which you want to extract the text next to it, to extract the name “Aquaforest Test”, enter “To:”.

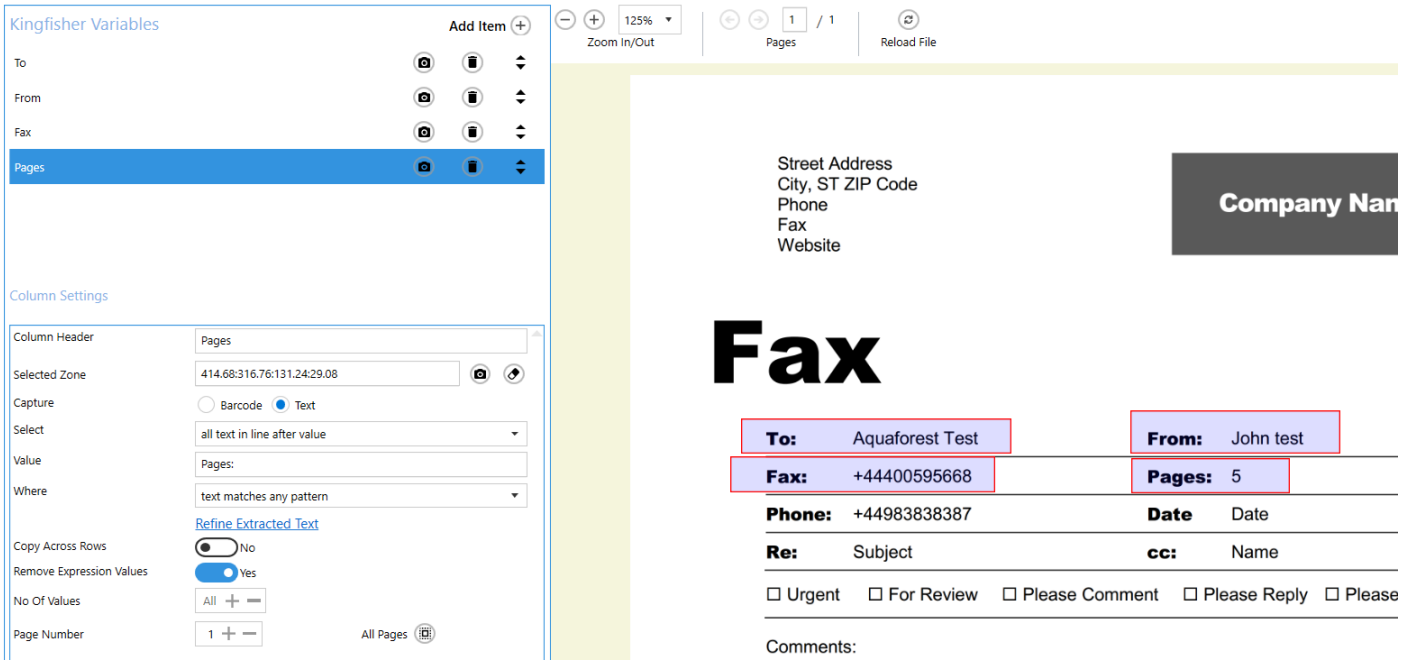


3. After you enter the Expression, click on the camera Icon button next to the item or in the second line of the “Column Settings” Group Box.

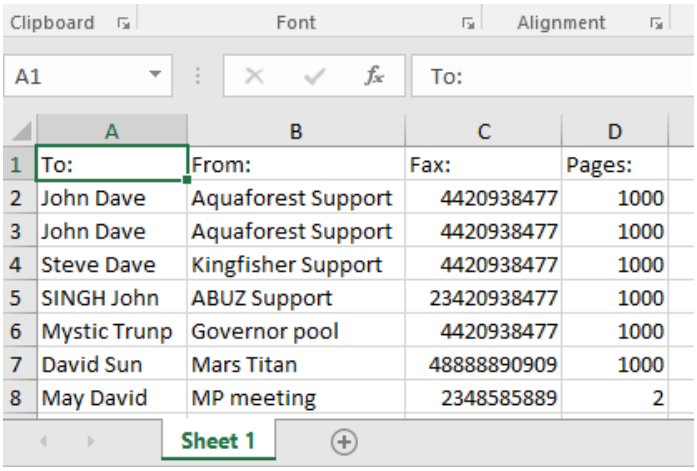


4. As you can see by default Kingfisher will extract the word that appears after the Expression provided. The rest of the Interface works the same as the “Select Variables” tab explained in section 2.2.2

5. You can add several Expressions before you run the job.



6. Below is the expected output



	A	B	C	D
1	To:	From:	Fax:	Pages:
2	John Dave	Aquaforest Support	4420938477	1000
3	John Dave	Aquaforest Support	4420938477	1000
4	Steve Dave	Kingfisher Support	4420938477	1000
5	SINGH John	ABUZ Support	23420938477	1000
6	Mystic Trunp	Governor pool	4420938477	1000
7	David Sun	Mars Titan	48888890909	1000
8	May David	MP meeting	2348585889	2

7 CUSTOM SCRIPTS

In kingfisher, the user can execute custom scripts after each file is processed, this gives the user a chance to do some post processing on the file input or output files. This gives the user the power to perform any type of operations that are not available in kingfisher.

We use **Windows Script Host** to execute these scripts, below is example the command line generated to execute a custom script.

```
cscript "C:\Aquaforest\Kingfisher\custom\stamp.wsf" "/Newfilename:XYZ Company.pdf" "/Oldfilepath:J:\QA Folder\Kingfisher QA\Input\Text Match\479.pdf" "/TempTarget:C:\Aquaforest\Kingfisher\temp\jobs\10000\a244c740-f0e3-4ff1-9de9-9fd5c86097a9\target" "/FinalTarget:J:\QA Folder\Kingfisher QA\Input\Text Match" "/JobSource:J:\QA Folder\Kingfisher QA\Input\Text Match\479.pdf" "/value1:XYZ Company"
```

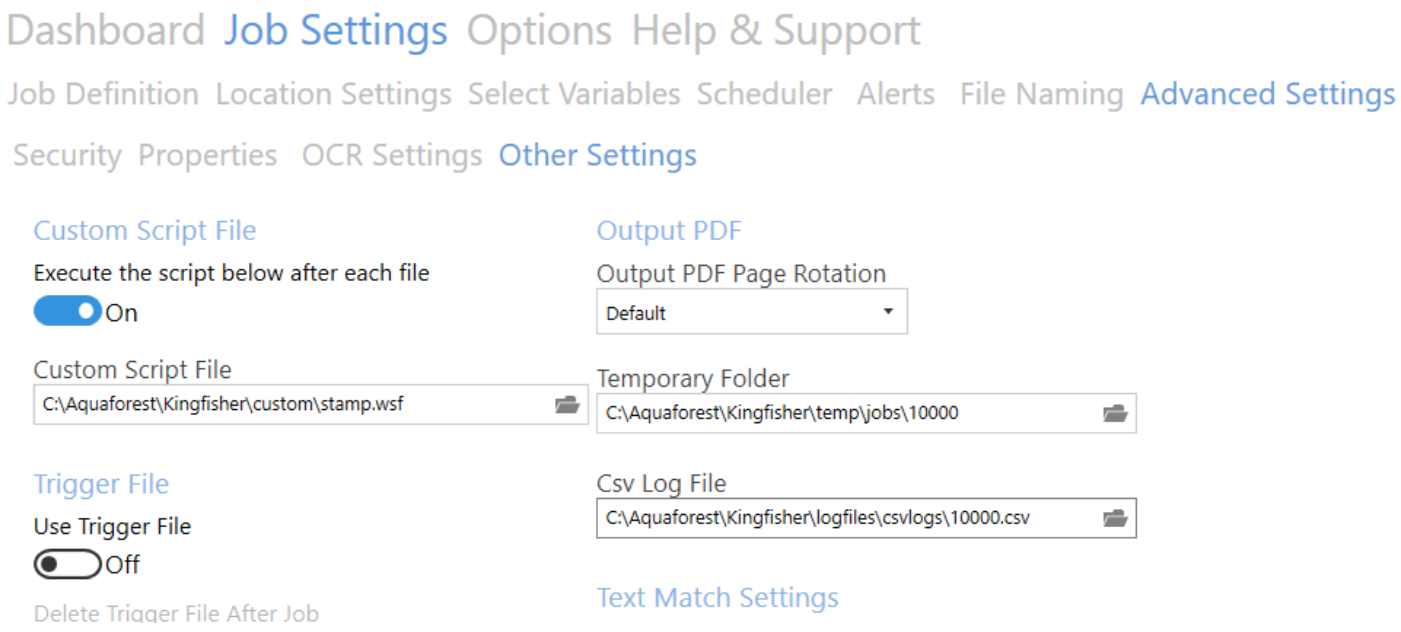
7.1 Custom Script Arguments

Value	Description
cscript	Path to a valid WSF file.
Newfilename	The file generated by kingfisher, this does not contain the whole path.
Oldffilepath	The input file path, in a SharePoint step this will the temporary download location.
TempTarget	This is the temporary folder where kingfisher stores all the target files until the job is finished, this is where you will get the output file. **i.e. outputfilepath= TempTarget+ Newfilename**
`TempOutputFile`	This is the temporary output file, it is the same as concatenating the **TempTarget** and **Newfilename** variables above.
FinalTarget	This is the destination path.
JobSource	This is the source path that the user provided when setting up the job, it can be a file name, a folder name or a tree name.
value1	This represents the “%VALUE1%” selected variable
value2	This represents the “%VALUE2%” selected variable
value3	This represents the “%VALUE3%” selected variable
value4	This represents the “%VALUE4%” selected variable

7.2 Enabling Custom Scripts

To enable custom scripts in a job you should

- Go to **Job Settings → Advance Settings → Other settings.**
- Switch the “Execute the script below after each file” switch on.
- Enter a valid “.wsf” file in the “Custom Script File” file browser.



The settings above will execute a script named “C:\Aquaforest\Kingfisher\custom\stamp.wsf” after each file.

7.3 Custom Scripts Samples

We have shipped a few sample scripts to get you started, to access these scripts, extract the “custom.zip” to a suitable location. You will notice three files that have the “.wsf” extension.

7.3.1 archiveinputs.wsf

This script copies the input file to an archive location. This has been tested with the file system job only. Open the file with a text editor of your choice and apply the following changes to use the script.

- Line 14: var ArchiveTarget = "%ARCHIVELOACTION%".
- Line 15: var move = "false". Set to true to move the file instead of copying it.

7.3.2 stamp.wsf

This script applies a stamp to the first page of the output PDF file, this stamp can contain the values extracted from the file. Open the file with a text editor of your choice and apply the following changes to use the script.

- Line 18: var stampExe = "C:\\Aquaforest\\Kingfisher\\custom\\stamp\\stamp.exe". Provide the path to the stamp.exe file, this can be found in the “stamp” subfolder of the custom.zip file.

- Line 19: var stampText = "Processed by %value1%". This is the text that will be stamped on the PDF file, the script will replace the “%valuex%” in the string with the values extracted from the document by kingfisher. E.g. if the text extract from the file is “Admin Department”, the stamp will be “Processed by Admin Department”

7.3.3 converttoxml.wsf

This script converts a CSV output file from kingfisher to an XML and copies the file to the target directory. This has been tested with the file system job only. Open the file with a text editor of your choice and apply the following changes to use the script.

- Line 14: var XmlRoot = "books". This represents the root xml node.

- Line 15: var XmlElement = "book". This is the name of the node that represents each record in the CSV.

- Line 16: var HasColumnNames="true". This will use the first row of the CSV file as the names of the xml nodes that represent the CSV Records. If you use false, the xml nodes will be named column1,column2 etc.

- Line 17 var csv2XmlExe = "C:\\Aquaforest\\Kingfisher\\custom\\csvtoxml\\CSV2XML.exe". Provide the path to the CSV2XML.exe file, this can be found in the “csvtoxml” subfolder of the custom.zip file.

8 XML JOB DEFINITION FILES

Any job created in kingfisher is saved in a job definition file, these are xml files usually created in the conditions folder. These files can be used to execute jobs from the Command line using the –JobFile parameter.

Developers may wish to create or generate XML job files using the information below.

A sample job file is shown here:

```
<pdfsplitterJob>
 <operation>renametext</operation>
 <jobid>10000</jobid>
 <conditionname> Job 10000</conditionname>
 <jobspecialoptions>
 </jobspecialoptions>
 <joboptions> -language 0</joboptions>
 <pagerangeoption>0</pagerangeoption>
 <splitrange>
 </splitrange>
 <frompage>1</frompage>
 <topage>9999</topage>
 <splitmode>
 </splitmode>
 <splitrepeatingrange>
 </splitrepeatingrange>
 <maxcores>1</maxcores>
 <condition>rename</condition>
 <usenametemplate>False</usenametemplate>
 <outputnamemask>%VALUE1%</outputnamemask>
 <sourcefiles>C:\Aquaforest\Kingfisher\Input\Text Match\postcode.pdf</sourcefiles>
 <password>
 </password>
 <target>C:\Aquaforest\Kingfisher\output\Text Match</target>
 <sourcetype>file</sourcetype>
 <casesensitive>False</casesensitive>
 <advancedflags>
 </advancedflags>
 <metadata>
 </metadata>
 <security>
 </security>
 <docoptions>
 </docoptions>
 <value1location>1.57:8.73:120.4:43.87</value1location>
 <nomatchfilename>
 </nomatchfilename>
 <nomatchoption>1</nomatchoption>
 <uselines>False</uselines>
 <extractgroup>False</extractgroup>
 <writehead>False</writehead>
 <appendcsv>false</appendcsv>
 <usefirsttableforall>False</usefirsttableforall>
 <usewords>False</usewords>
 <useworksheets>False</useworksheets>
 <mincolumns>1</mincolumns>
 <removelines>False</removelines>
```

```
 <deskew>False</deskew>
 <converttotiff>0</converttotiff>
 <converttotiffdpi>0</converttotiffdpi>
 <retainbookmarks>False</retainbookmarks>
 <retainmetadata>False</retainmetadata>
 <binarize>False</binarize>
 <language>0</language>
 <despeckle>0</despeckle>
 <tablespace>1</tablespace>
 <pagestocheck>10</pagestocheck>
 <minrows>1</minrows>
 <outputscript>
 </outputscript>
 <jobtemp>J:\dev\Source Control\Kingfisher
2.0\PDFjunction.Deployment\temp\aquaforest\kingfisher\jobs\10000</jobtemp>
 <onerrorcontinue>false</onerrorcontinue>
 <errors>
 </errors>
 <mergeextarctedpagesmatch>False</mergeextarctedpagesmatch>
 <pagerotation>-1</pagerotation>
 <csvoutput>xlsx</csvoutput>
 <tablebreak>50</tablebreak>
 <mergeextarctedpages>
 </mergeextarctedpages>
 <logfile>C:\Aquaforest\Kingfisher\logfiles\ Job 10000\%TIMESTAMP%.txt</logfile>
 <lastlogfile>C:\Aquaforest\Kingfisher\logfiles\ Job 10000\16-13-33.txt</lastlogfile>
 <kingfisherlivelog>J:\dev\Source Control\Kingfisher
2.0\PDFjunction.Deployment\logfiles\livelogs\10000.log</kingfisherlivelog>
 <runstatus>C:\Aquaforest\Kingfisher\status\10000.txt</runstatus>
 <lastrundate>09 Jun 2017 16: 13:32</lastrundate>
 <overwriteexisting>False</overwriteexisting>
 <overwriteexistingcurrent>False</overwriteexistingcurrent>
 <debug>False</debug>
 <doocr>False</doocr>
 <createfolders>True</createfolders>
 <schedulecontinuousand>23:59:00</schedulecontinuousand>
 <schedulecontinuousbetween>00:00:01</schedulecontinuousbetween>
 <schedulecontinuousevery>10</schedulecontinuousevery>
 <schedulecontinuouseveryunit>minutes</schedulecontinuouseveryunit>
 <scheduleonceperdayat>15:31:51</scheduleonceperdayat>
 <schedulerunonceat>15:31:51</schedulerunonceat>
 <schedulerunonceon>1/4/2018 3:31:51 PM</schedulerunonceon>
 <scheduletype>Manual</scheduletype>
 <emailbody>Processing of Job: '%JOBNAME%' completed with status: '%STATUS%' .
Log file: %LOGFILE%
</emailbody>
 <emailsubject>%JOBNAME% %STATUS%</emailsubject>
 <toemail>
 </toemail>
 <fromemail>
 </fromemail>
 <sendemailonerror>False</sendemailonerror>
 <sendemailonsuccces>False</sendemailonsuccces>
 <continueonerror>False</continueonerror>
 <removewhitespaces>False</removewhitespaces>
 <usetriggerfile>False</usetriggerfile>
```

```
 <deletetriggerfile>False</deletetriggerfile>
 <triggerfile>
 </triggerfile>
 <csvlogfile>J:\dev\Source Control\Kingfisher
2.0\PDFjunction.Deployment\logfiles\csvlogs\10000.csv</csvlogfile>
 <kingfisherversion>2.0.220817</kingfisherversion>
 <TextExtraction>
  <ExtractionItem Barcode="false" Regex="0" CopyValuesToAllRows="false" SelectMultipleLines="false"
RemoveKeywords="true" BarcodeFormat="" PageNumber="" ValueCount="" Expression="" Format=""
HeaderName="%VALUE1%" Location="1.57:8.73:120.4:43.87" Position="6" />
 </TextExtraction>
</pdfsplitterJob>
```

Attribute	Description
operation	Must be split, extract or rename.
sourcetype	Source type (file, folder, tree). The default value is file.
condition	
target	Target directory for result files. If not specified, the source directory will be used as the target.
Joboptions	Set of options for the specified operation. Described in detail in section 6.2
jobspecialoptions	Set of options for the specified operation. Described in detail in section 6.2.1 and 6.2.2
outputnamemask	The Name Template explained in section 4.1.5
usenametemplate	If the value is true, the Name Template will be used instead of the script.
Metadata	Metadata settings for result PDF files. By default, metadata fields are left blank. Described in detail in section 6.2.3
Security	Security settings for result PDF files. By default, files are not secured. Described in detail in section 6.2.4
docoptions	Document option settings for result PDF files. Described in detail in section 6.2.5
Sourcefiles	Source PDF file, folder or tree. May be multiple files for merge operations.
outputscript	The file naming script.
advancedflags	Used to pass extra parameters, do not use unless directed by Aquaforest Technical team.
logfile	If specified, output will be logged to a file with this name in .
TextExtraction	This contains the definitions of text to be extracted .

9 USING KINGFISHER FROM THE COMMAND LINE

Kingfisher can be accessed via the command line to run rules that have already been created in the GUI. You can also run kingfisher directly from the command line.
The executable used for running Kingfisher from the command line is “Kingfisher.Commandline.exe”

9.1 Using Job Definition File

This is the preferred and safest method to run kingfisher from the command line, this method ensures you set up the job correctly and you are sure of what it does.
If you create a Job in Kingfisher, an xml Job Definition file is automatically created in the “conditions” folder in the installation Directory.
Note: When using the command line option, the xml file can actually be stored in a different location if required. e.g. c:\mydocuments\1001.xml. Thus, a working copy of definitions can be stored in a different (possibly less secure) location to the original.

Usage:

Kingfisher.Commandline.exe -jobfile=“..\\conditions\\1001.xml”

10 AUTOBAHN DX INTEGRATION

Autobahn DX is another software provided by Aquaforest, it is a Document Processing product designed to fit into an organization's document workflow. It can process and convert a variety of different types of document including TIFF images, Microsoft Office documents and HTML pages, convert those documents to searchable PDFs and perform a variety of other customizable activities. For more information on Autobahn DX, check the product page by clicking the link below.

https://www.aquaforest.com/en/autobahn.asp

A Kingfisher Job can now be set up in Autobahn DX, for more details, see the blog post below.

11 KINGFISHER FILES AND DIRECTORIES

After installation, the following subdirectories are created in the Kingfisher install directory:

Directory	Contents
bin	Executables and Scripts.
docs	Reference Guide, Sample videos
conditions	Standard directory for XML Job definition files.
logfiles	Log files.
regex	This folder contains files that can be used to store predefined regular expressions.
license	License File.
samples	This folder contains sample files used in the sample jobs.
theme	This folder contains the file that keeps track of the current theme.
Status	Contains status file, for progress reports.

12 SUPPORT

Any problems running this application should give evident error messages, but if you are unable to resolve a problem please contact support@aquaforest.com

13 ACKNOWLEDGEMENTS

This product makes use of many Open Source components which are included in binary form. The appropriate acknowledgements and copyright notices are given below.

Name	Homepage
Cuneiform	n/a (Copyright (c) 1993-2008, Cognitive Technologies)
FreeImage.NET	Homepage
IKVM.NET	Homepage | Sourceforge
Leptonica	Homepage
Libjpeg	Homepage
Libpng	Homepage
Libtiff	Homepage
PDFBox	Homepage
Zlib	Homepage
ZXing.NET	Homepage