# Kingfisher Reference Guide



# Version 1.1 May 2017

© Aquaforest Limited 2001-2017 Web: <u>www.aquaforest.com</u> E-mail: <u>info@aquaforest.com</u> Δquoforest

### CONTENTS

1	INTRO	DUCTION	.3
	1.1 IN	STALLATION	.3
	1.2 Sy	STEM REQUIREMENTS	.3
	1.3 Li	cencing and Purchasing	.3
	1.4 Vi	ersion Breakdown	.3
2	KINGF	ISHER WINDOWS INTERFACE	.4
	2.1 D	ASHBOARD	.5
	2.2 Jo	B SETTINGS	.6
	2.2.1	Job Definition	
	2.2.2	Select Variables	
	2.2.3	File Naming	
	2.2.4	PDF Settings	
	2.2.5	Running Job	
		PTIONS	
	2.4 H	ELP	20
3	KINGF	ISHER OPERATIONS	21
	3.1 Ri	ENAME PDF FILE	21
	3.1.1	Rename Based on Barcode	
	3.1.2	Rename Based on Text Match	
	3.1.3	Rename Based on Barcode and Text Match	
		RTRACT PAGES FROM PDF FILE	
	3.2.1	Extract Pages from File based on Text Match	
	3.2.2	Extract Pages from File Name based on Barcode	
	3.3.1	Split File based on Text Match	
	3.3.2	Split File based on Text Change	
	3.3.3	Split File based on Barcode	
	<i>3.3.4</i> 3.4 E>	Split File based on Page Ranges	
	3.4 EX 3.4.1	Extract Text from PDF File	
	3.4.1 3.4.2	PDF to CSV/XLSX	
	3.4.2	Advanced Export to CSV/XLSX	
	00		
4		ING AND RUNNING JOBS WITH THE GUI	
		REATING A NEW JOB	
	4.2 E	DITING A JOB	36
5	EXTRA	CT TABULAR DATA FROM PDF	37
	5.1 TH	ie Document View	38
	5.2 TA	ABLE VIEW	40
	5.2.1	Tables	40
	5.2.2	Grouped Tables	
	5.2.3	Data grid	45
6	ADVA	NCED EXPORT TO CSV/XLSX	46
7	XML J	OB DEFINITION FILES	50
8	USING	KINGFISHER FROM THE COMMAND LINE	53
	8.1 U	sing Job Definition File	53

8	.2 (	Command Line Syntax	53
	8.2.1	Split Operation Parameters	56
	8.2.2	Extract Job Parameters	57
	8.2.3	Extract Text Job Parameters	57
	8.2.4	Metadata Specification	57
	8.2.5	Security Specification	58
	8.2.6	Document Options Specification	58
8	.3 (	Command Line Examples	60
	8.3.1	Command line to split all PDFs in a folder tree by Barcode	60
	8.3.2	Command line to extract text from all the PDF files in a folder	60
	8.3.3	······································	
	8.3.4	Command line to rename a PDF file Based on text match	60
	8.3.5	Command line to rename all the PDF files in a folder tree to the value of the first barcode encountered	1
	in the	e file	60
	8.3.6		
	8.3.7	Command line to split all the PDF files in a folder by barcode	61
	8.3.8	Command line to rename a PDF file based on Text and barcode	61
	8.3.9	Command line to split a PDF file by top level bookmarks and rename the PDF files to the bookmark tit	les
		61	
	8.3.1	0 Command line to split a PDF file by a change in the text in a particular on the page	61
9	KING	FISHER FILES AND DIRECTORIES	62
10	SU	PPORT	62
11	AC	KNOWLEDGEMENTS	62

# **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. Our OCR engine ensures that this text regions are extracted as strings even though the PDF is an image only file, this also makes kingfisher have an edge over similar tools.

# **1.1 Installation**

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

# **1.2 System Requirements**

- Server 2008, Server 2012, Server 2016,
- Windows 7 (64 bit), Windows 8 (64 bit), Windows 8.1 (64 bit), Windows 10 (64 bit).
- Microsoft .net framework 4.6.1
- 250Mb of disk space is required.

# **1.3 Licencing and Purchasing**

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

When a license is purchased (see <u>http://www.aquaforest.com/en/tif\_pdf\_buy.asp#collapseEleven</u>) 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

Feature	Basic	Standard	Advanced
Split by PDF Content, Barcode or Page	$\checkmark$	$\checkmark$	$\checkmark$
Ranges			
Extract Pages by PDF Content or Barcode	$\checkmark$	$\checkmark$	$\checkmark$
Extract Text (including OCRed text)	$\checkmark$	$\checkmark$	$\checkmark$
Rename by PDF Content or Barcode	$\checkmark$	$\checkmark$	$\checkmark$
Graphical User Interface	$\checkmark$	$\checkmark$	$\checkmark$
Command Line Interface		$\checkmark$	$\checkmark$
Extract PDF Content to CSV		$\checkmark$	$\checkmark$
Number of Cores Supported	1	2	8

# **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.

# Dashboard Job Settings Options Help & Support

+ Add New Job					
OPERATION	JOB ID	JOB NAME	LAST RUN	RUN STATUS	
E Rename based on Text Match	10000	Job 10000	24 Feb 2017 15: 35:02	Finished	
E Rename based on Barcode	10001	Job 10001	24 Feb 2017 15: 35:22	Finished	
E Rename based on Text Match and Barcode	10002	Job 10002	24 Feb 2017 15: 36:06	Finished	
-  - Split File Delete Job	10003	Job 10003	24 Feb 2017 15: 37:49	Finished	
	10004	split 1	24 Feb 2017 15: 38:02	Finished	
- I+ Split File based on Page Ranges	10006	Split 3	15 Feb 2017 17: 37:31	Finished	
👽 Extract Pages from File based on Text Match	10007	Extarct 1	15 Feb 2017 17: 49:12	Finished	
😴 Extract Pages from File based on Barcode	10008	Extract 2	15 Feb 2017 17: 50:44	Finished	
T PDF to Text	10009	Job 10009	15 Feb 2017 17: 53:46	Finished	

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.

# 2.2 Job Settings

Job Definition Select Variables File Naming Advanced Settings Running Job

Enter the Name of the Job and Select Operation	Location Details			
Job ID	<ul> <li>Single File</li> </ul>	O Folder	◯ Tree	
10010	Source			
Job Name	C:\Kingfisher QA\Input	\Tables\PDF Extractor Tes	t Files\01240682090.pdf	<b>_</b>
Job 10010	Destination		Same	As Source
Select Operation	C:\Kingfisher QA\Input	\Tables\PDF Extractor Tes	t Files	<b>_</b>
Rename PDF File	Log File			
Rename based on Text Match	C:\Aquaforest\Kingfishe	er\logfiles\ Job 10010\%T	TIMESTAMP%.txt Select a logfile	
	Job Options			
	Case Sensitive	Debug Off	Continue on Erro	r
	Create Folders	OCR Off	Overwrite Existing	g Files
			Oli	
	Max Cores			
	1	•		
(?)			(m) (1	<u>۹</u>
Help Run	Stop Save Job	Сору Јов	$\sim$	Log File

Screen Field / Button	Description
View Log File	Clicking this button will display the last generated log file of the selected rule.
Save	Clicking this button will save all the changes in the rule.
Run	Clicking this button will execute the rule on the selected files.
Stop	Clicking this button will halt the processing of the current batch of files at the earliest convenience.
Delete Job	Clicking this button will delete the currently selected job.
Сору Јођ	Clicking this button will to make a duplicate of the currently selected job.
Explore Destination	Clicking this button will 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
Source Type	File Folder Tree
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.
Source	The source PDF file or directory.
Destination	The location where the generated PDF file(s) will be placed. This can be set to be the same as the source location. Note that Kingfisher cannot overwrite source files, a new file is always made.
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 Select Variables

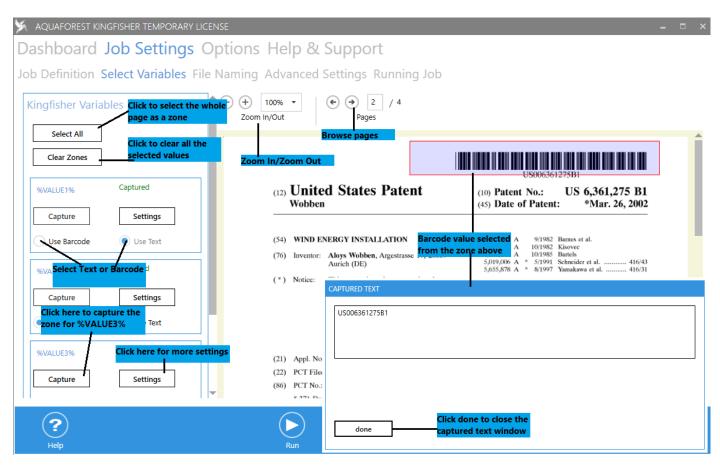
One of the most attractive features of kingfisher is the ability to select text or barcode at a position in a PDF page, the selected text or barcode can be used to:

- 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 capture button under the variable name you want to assign the zone to, you will notice the "Not Captured" red text change to a "Captured" green text. Check the picture below for more details.



Screen Field / Button	Description
Clear Zones	Click this button to clear all the zones selected
Select All	Click this button to select the whole page.
Magnifier	Slide this to the left and right 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
Use Text	Check this radio button to extract text values
Use barcode	Check this radio button to extract barcode values
Capture	Click the capture button to assign a selected zone to the variable
	name above the button.
Settings	Click this button to give more criteria for selecting text.

### 2.2.2.1 Capture Settings

The Settings control can be launched by clicking the "Settings" button shown in the window above, it provides a platform for the user to further describe the type of text of barcode they want to extract. The diagrams below show the text and barcode capture settings.

### 2.2.2.1.1 Text Match Option

SETTINGS
<choose match="" options="" text=""></choose>
Where text matches any pattern
Show Help Tips
apply changes discard changes

Screen Field / Button	Description
<choose match="" options="" text=""></choose>	<ul> <li>Select text in zone: This option will select all the text that was extracted.</li> <li>Select word after value: If this option is selected, kingfisher will return the word that appears immediately after the word supplied in the text box.</li> <li>Select Word Before value: If this option is selected, kingfisher will return the word that appears immediately before the word supplied in the text box.</li> <li>Select 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.</li> <li>Select all text In Line before 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.</li> <li>Select 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.</li> <li>Select 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.</li> <li>Select all text In Zone before 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.</li> </ul>
<choose match<br="" pattern="">Options&gt;</choose>	<ul> <li>Where text matches any pattern: This option will not match the text to any pattern</li> <li>Where text matches pattern: Provide a Regular expression here to match the extracted text to, this will be explained more below.</li> </ul>

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.
	<ul> <li>ITF (Interleaved Two of Five) 1D format (Code 25).</li> </ul>
	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

SETTINGS				
Barcode Type				
Where text matches any pattern				
Show Help Tips				
apply changes discard changes				

### 2.2.2.1.3 Text Pattern Option

Kingfisher allows the user to define the type of text that is extracted from the PDF document. This is to prevent unwanted text to be extracted just because it happens to be in the same location as the text the user is interested in. 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 section \*17 of the reference guide.

Pattern	Matches	Examples
invoice	The word "invoice"	
invoice no:	The phrase "invoice no:"	
	Any 1 character	А
		9
		%
	Any 4 characters	12ab
		\$5.00
Invoice	The word invoice followed by any 4	Invoice 123
	characters. A space is a character.	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 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:

### 2.2.3.1 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%	The barcode/text value(s) found.
%VALUE3%	
%VALUE4%	
%INDEX%	The current index of the
%PAGENUMBER%	the current split index for the current value
	(used with extract)
%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

Job Definition Select Variables File Naming PDF Setting

Use Name Template

Name Template

Variable	Description
%VALUE1%	
%VALUE2%	
%FILENAME%	Replaced By the Name of the Original File
%DATE%	Replaced By The Current Date
%TIME%	Replaced By The Current Time
%PAGENUMBER%	Replaced By The Page Number Where the Text/Barcode Was Found
%PAGESTART%	Replaced By The Start Page Number of a Split PDF
%PAGEEND%	Replaced By The End Page Number of a Split PDF
%INDEX%	Replaced By The Split or Extract Index of a file.
%BOOKMARK%	Replaced by the Bookmark title, Use this with split by bookmark.

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".

HP21 4LA	%VALUE1%		
	United States Gutterman	Patent -	US006409467B1 

#### 2.2.3.2 Script

The user can also write a script to generate file names, to use this feature, select the "Use Scripts for Output Names" radio button

Use Name Template
Use Scripts For Output Names

Select from the scripts below	Tips
	1 You can use JavaScript to specify the output filename.
append constant 💌	2 Any string assigned to the variable "newName" will be used as the new filename
	3 You can also use the variables in the name template section and they will be replaced accordingly
Script	4 The parameters of the method below match the 4 variables ("%VALUE1% - %VALUE4%")
function NewFilename(value1,value2,value3,value4)	extracted from the PDF file
{	5 You should only use single quotes to enclose strings
var newName=' '	6 You can load and edit the scripts or write yours from scratch.
newName="dsAyIsebury\\"+value1; return newName }	*
Update Script         Add New Script         Delete Script	

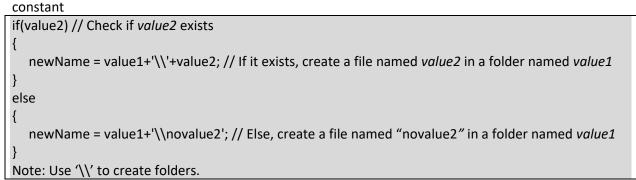
Any string assigned to the variable "*newName*" will be used as the new filename, you can also use the variables discussed in the preceding section and they will be replaced accordingly.

The parameters are the variables that have been extracted from the PDF document.

```
Script
function NewFilename(value1,value2,value3)
{
var newName=' '
return newName
```

}

Below are some examples of scripts that can be used in kingfisher to Check if a value is extracted, if not use a



### **2.2.3.2.1** Use substring to group strings with a common prefix

ſ	if(value2) // Check if value2 exists
l	{
l	// If it exists, create a file named value2 in a folder named as the first three letters value1
l	<pre>newName = value1.substring(0, 3)+'\\'+value2;</pre>
l	}
l	else
l	{
l	// Else, create a file named "novalue2" in a folder named as the first three letters of value1
l	newName = value1.substring(0, 3)+'\\novalue2';
l	1

ermissions		
	Permission	Description
	Allow ALL Permissions	All the permissions below.
	Printing	Allow high-quality printing
	Modify Contents	Allow assembly (see below) and other
		document medications
	Сору	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.

### 2.2.4 PDF Settings

This tab contains some advanced settings that would allow define the output PDF and provide some OCR settings.

### 2.2.4.1 Security Settings

Job Definition Select Variables File Naming Advanced Settings Running Job

Security Properties OCR Settings Other Settings

#### **Document Security**

Set Security Off Source Password	Targ	jet User Password	Target Owner P	Target Owner Password	
Target Document Permissi	Assembly	Degraded Printing	Modify Annotations	Copy	
Allow All Permissions	Form Filling	Screen Readers	Modify Contents	Off Printing Off	

### 2.2.4.1.1 Source Document Security

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

### 2.2.4.1.2 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	
	Сору	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.	
	Modify Annotations Form Filling Screen Readers	Allow modification of annotations Allow filling of form fields Allow extraction of text and graphics in support of accessibility.	

Degraded Printing Allow low-quality printing	Degradeo	d Printing Allow lo	w-quality printing
--	----------	---------------------	--------------------

#### 2.2.4.2 Properties

Job Definition Select Variables File Naming Advanced Settings Running Job

Security Properties OCR Settings Other Settings

#### Document Metadata

Author		Title		Creator	
Subject		Keywords		Retain Metadata	
Display Mode					
Page Layout		Page Mode		Non Full screen Mode	
Default	•	Default	•	Default 👻	
PDF Settings					
Hide Menu Bar	Hide Tool Bar	Center Window Off	Fit Window	Hide WindowUI	

This page allows setting of the generated PDF Document Properties. Custom and XMP metadata may also be set – see section 7 for further details.

The **Document Metadata** Section allows the user to change the metadata to the target PDF file.

#### 2.2.4.2.1 Display Modes

Option	Description		
Page Layout	The setting for initial document page display. One of:		
	Single Page		
	Continuous		
	<ul> <li>Continuous Facing (odd pages left)</li> </ul>		
	<ul> <li>Continuous Facing (odd pages right)</li> </ul>		
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	Only applicable where Page Mode=Full Screen. The setting for document		
Mode	page display when exiting Full Screen mode. One of:		
	<ul> <li>Neither Bookmarks nor Thumbnails Open</li> </ul>		
	Bookmarks Open		
	Thumbnails Open		

### 2.2.4.2.2 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.

### 2.2.4.3 OCR

Job Definition Select Variables File Naming Advanced Settings Running Job

Security Properties OCR Settings Other Settings

#### Image Pre-Processing

OCR Language	Despeckle	Convert To TIFF DPI
English 👻	No Despeckle 🔻	Default 👻
Deskew On	Convert To TIFF	OCR Pages with Image Off
Remove Lines	Binarize	

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	If this is set to true, kingfisher will OCR all pages with images in them by
Image	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
Remove Lines	Remove lines from images.
Binarize	Use This option to process coloured files.
Deskew	Deskew (straighten) the image.

### 2.2.4.4 Other Settings

Security Properties OCR Settings Other Settings

•

Output PDF Page Rotation

Default

Retain Bookmarks

Off

Option	Description
Output PDF Page	Choose the rotation of the output PDF pages.

# 2.2.5 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.

🕺 AQUAFOREST KINGFISHER PERMANEN	IT LICENSE					_ = ×
Dashboard Job Settings C	ptions H	lelp				
Job Definition Select Varia	bles File	Naming	J PDF Set	ttings <mark>Run</mark>	ning Job	
Task Log						
Page 14 Processed.						•
Page 15 Processed.						
Page 16 Processed.						
Analysing OCR results to get split ranges. The split ranges have been computed, starting split proces Genarating 1st file 279330V5149516200410.pdf	iS.					
Genarating 2nd file 279330V5149516200410 -1.pdf						
Genarating 3rd file 279330V5149516200410 -2.pdf						
Genarating 4th file 279330V5149516200410 -3.pdf						
Genarating 5th file 279330V5149516200410 -4.pdf						
Genarating 6th file 279330V5149516200410 -5.pdf						
Genarating 7th file 279330V5149516200410 -6.pdf						
Genarating 8th file 279330V5149516200410 -7.pdf						
Exit Code : 0						_
Time taken = 00+00+02 /782/152						
	$\triangleright$			Ē		P
	Run		Save	Delete Job	View Log File	Explore Destination

# 2.3 Options

# Dashboard Job Settings Options Help & Support

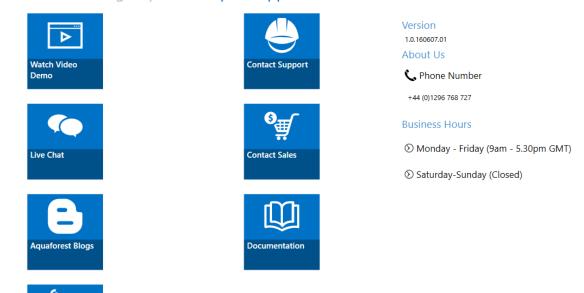
License Key				
Temporary License Expires On Wednesda Multicore License Wi Kingfisher Advanced	th 8 Cores	(Days Remaining 13)	UPDATE	KEY
Show Help Tips	5			
Set Theme				
Light Blue	Light Red	Light Turquoise	Light Purple	Light Orang
		_		_
Dark Blue	Dark Red	Dark Turquoise	Dark Purple	Dark Orang

Option	Description
License Key	Enter the license key in the License key text box and click the "UPDATE
	KEY" button
Set Theme	To change the theme, click on the Tile that corresponds to the colour
	you wish to switch to.
Set Language	To change the language used in the GUI, switch between the options in the "Set
	Language" combo box.
Show Help Tips	Switch this on to view help tips when creating and editing job in Kingfisher

# 2.4 Help

Request Rem Session

Our help Tab give you access to all the different ways you can get support on how to use Kingfisher. Dashboard Job Settings Options Help & Support



Option	Description
Watch Video	This will launch a video showing the user how to create and run jobs in
Demo	Kingfisher
Request Remote	Request a remote session and we will contact you to arrange a suitable time.
Session	
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.

Select Operation	
Rename PDF File	•
Rename PDF File	
Split PDF File	
Extract Pages From PDF File	
Export PDF Content	

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

### Select Operation

Rename PDF File Rename based on Text Match

### Files With No Match

Copy to Output and Rename

File Name

%filename%

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.
	<ul> <li>Do Not Copy to Output</li> <li>Copy to Output</li> <li>Copy to Output and Rename</li> </ul>
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 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

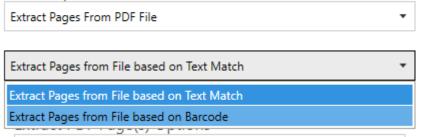
This step is like the step above except that it uses text in the PDF document to rename the file. Unlike the barcode step, an Area must be selected for this step to work. You can also specify Regular Expressions and Words that appear before/after the word you want the file to be named to. These extra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

Select Operation



### 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.

		(?) 🗶 Sigi
	$\heartsuit$ $\checkmark$	
	US006409467B1	<ul> <li>Export PDF</li> <li>Create PDF</li> </ul>
t	(10) Patent No.:       US 6,409,467 B1         (45) Date of Patent:       Jun. 25, 2002	Edit PDF
CM 0x 230881 JY (US)	4,739,179 A 4/1988 Stites 5,038,049 A 8/1991 Kato 5,134,305 A 7/1992 Senehi 5,272,378 A 12/1993 Wither	<ul> <li>Container mes</li> <li>∠ Fill &amp; Sign</li> <li>∠ Send for Signature</li> <li>→- Send &amp; Track</li> </ul>

### The extract step has the following settings Output File Options

#### Extract PDF Page(s) Options

Extract As Single Pages	•
Extract As Single Pages	
Merge All Extracted Pages	
Merge Extracted Pages On Unique Expressions That Match	

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	Merges All the Extracted pages that have
That Merge	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 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.

		?	
¢			
nt	(10) Patent No.: US 6,409,467 B1	Edit PDF	
	(45) Date of Patent: Jun. 25, 2002	🥟 Commer	
	( ),	😲 Combine	

### The extract step has the following settings

### **Output File Options**

Extract PDF Page(s) Options

Extract As Single Pages	•
Extract As Single Pages	
Merge All Extracted Pages	
Merge Extracted Pages On Unique Expressions That Match	

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	Merges All the Extracted pages that have
That Merge	equal "%VALUE1%" values

# 3.3 Split PDF File

Select Operation

Split PDF File	•
Split File based on Text Match	•
Split File based on Text Match	
Split File based on Text Change	
Split File based on Barcode	
Split File based on Page Ranges	

# 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.

Split File based on Text Match

**Output File Options** 

Page That Matches Text on First Page

Pages With No Match		File Name	
Copy to Output and Rename	•	%filename%	

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	<ul> <li>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.</li> <li>Do Not Copy to Output</li> <li>Copy to Output</li> <li>Copy to Output and Rename</li> </ul>
File Name	If the "Copy to Output and Rename" option is selected

the "File Naming" ta breakdown of the v	ab, the table below gives you a ariables 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 very where there is a text match, it waits until the text that was matched has changed from the one on the previous page.

•

Split File based on Text Change



Page That Matches Text on First Page

Pages With No Match

Copy to Output and Rename

File Name

•

%filename%

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.

Split File based on Barcode			•
Output File Options			
Barcode on First Page		•	
Pages With No Match		File Name	
Copy to Output and Rename	•	%filename%	

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

Screen Field / Button	Description			
Split Type	Split by top-leve Split by page rar	One of: Split into single pages Split by top-level bookmarks Split by page ranges (See below) Split by repeating page ranges (See below)		
Page Ranges	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:			
	1	Specifies a single page		
	1-3	Specifies a range of pages		
	10-20even	Specifies even pages within the range		
	10-20odd	Specifies odd pages within the range		
	20-lastpage	Note that <i>lastpage</i> can be used to represent the last page in the document.		
Repeating Range	<ul> <li>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.</li> <li>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.</li> </ul>			

# 3.4 Extract Content from PDF File

Select Operation	
Export PDF Content	•
PDF to Text	•
PDF to Text	
PDF to CSV/XLSX	
Advanced Export to CSV/XLSX	

# 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.

Select Operation		
Extract Text From PDF		•
Start Page	End Page	

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

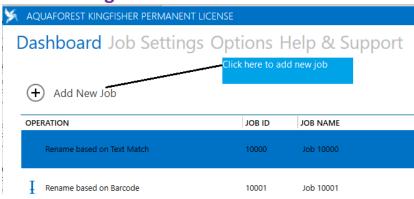
This step is used to extract tabular data from PDF files, see section 5 for more details.

# 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.

🕺 Aquaforest Kingfisher - Create Nev	v Job Wizard
Steps	
Introduction	Welcome to the Aquaforest Kingfisher Job Creation Wizard
Job Definition	This wizard will guide you through the different steps required to create a new Aquaforest Kingfisher Job.
Select Variables	To continue, click Next.
File Naming	
Advanced Settings	
Finish	

#### Follow the steps below to create a new job

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.

ps			
roduction	Enter the Name of the Job and Select Operation	Location Details	
b Definition		Single File	der 🔿 Tree
lect Variables	Job ID 10006	Source	
e Naming			1
vanced Settings	Job Name Job 10006	Destination	Same as Sou
sh		C:\Aquaforest\Kingfisher	Same as Sou
511	Select Operation		
	Rename PDF File	Log File     C:\Aquaforest\Kingfisher\logfiles\ Jo	
	Rename based on Text Match	C:\Aquatorest\kinglisner\logfiles\Jo	b 10006\%TIMESTAMP%.txt
		Job Options	
		Continue on Error	Debug
		Off	Off
		Create Folders	Overwrite Existing Files
		On	Off
		Case Sensitive	OCR
		Off	Off
		—	—
		Max Cores	
		1 -	
			Previous Next Preview Save Joh

2. 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.

Steps	
Introduction	Kingfisher Variables $\bigcirc \bigcirc \bigcirc \bigcirc \bigcirc \bigcirc \bigcirc 1$ / 30
Job Definition	Zoom In/Out Pages
Select Variables	Lear zones HP21 4LA
File Naming	clear zones US006409467B1
Advanced Settings	Not Captured         (12) United States Patent         (10) Patent No.:         US 6,409,467 B1           Gutterman         (45) Date of Patent:         Jun. 25, 2002
Finish	capture settings
	Use Barcode         Use Text         (54)         WIND-GENERATED POWER SYSTEM         4,739,179         A         4/1988         Stites           %VALUE2%         Not Captured         (76)         Inventor:         Howard Gutterman, P.O. Box 230881         5,133,049         A 1/1992         Stotel           %VALUE2%         Not Captured         (76)         Inventor:         Howard Gutterman, P.O. Box 230881         5,134,957         A 1/1992         Stotel           %VALUE2%         Not Captured         (76)         Inventor:         Howard Gutterman, P.O. Box 230881         5,134,957         A 1/1992         Stotel           %VALUE2%         Not Captured         (76)         Inventor:         Howard Gutterman, P.O. Box 230881         5,134,957         A 1/1992         Stotel           %VALUE2%         Not Captured         (76)         Inventor:         Howard Gutterman, P.O. Box 23088         100/18         201 R X
	capture     settings       (*) Notice:     Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.     * cited by examiner <i>Primary Examiner</i> —John E. Ryznic
	Use Barcode  Use Text (21) Appl. No.: 09/802,574 (57) ABSTRACT
	\$\times VALUE3%     Not Captured     (22) Filed:     Mar. 9, 2001     A wind-generated power system for producing supplemental decimation of pr
	%VALUE4%         Not Captured         3350,925 A * 4/1975 Stocket         44/34 X         annote an information what is excited at 3388/376 A * 5/1975 Uzell, L
	Use Barcode Use Text

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%".

%VALUE1%	Not Captured
capture	settings
Use Barcode	Use Text

• Now select the zone at the top right corner



• Click on the "capture" button, 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.

С	Captured Text				
	(>o) Patent No.: US 9,348,580 B2				
	done				

• 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 and click on the "Apply Changes" button.

SETTINGS			
Change Test Match Ontinger			
<choose match="" options="" text=""></choose>			
<choose match="" options="" pattern=""></choose>			
Show Help Tips			
apply changes discard changes			

SETTINGS		
Select all text in zone after value Patent No.:		
<choose match="" options="" pattern=""></choose>		
Show Help Tips		
apply changes discard changes		

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

C	Captured Text				
	US 9,348,580 B2				
	done				

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

• Select the "Use Barcode" option associated with "%VALUE1%".



• Now select the zone at the top right corner



• Click on the "capture" button, you will see the text below

Captured Text			
US009348580B2			
done			

 After selecting the variables, click next to go to the "File Naming" tab, this allows you to specify 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.

Finish       %DATE%       Replaced By The Current Date         %TIME%       Replaced By The Current Time         %PAGENUMBER%       Replaced By The Page Number Where the Text/Barcode Was Found         %PAGESTART%       Replaced By The Start Page Number of a Split PDF         %PAGEEND%       Replaced By The End Page Number of a Split PDF         %INDEX%       Replaced By The Split or Extract Index of a file.         %BOOKMARK%       Replaced by the Bookmark title, Use this	s	
Job Definition Select Variables File Naming PDF Settings Finish Variable %VALUE1%_%VALUE2% Variable %VALUE1% %VALUE1% %VALUE1% %VALUE2% %VALUE2% %FILENAME% Replaced By the Name of the Original File %DATE% Replaced By The Current Date %TIME% Replaced By The Current Time %PAGENUMBER% Replaced By The Spint Page Number of a Split PDF %PAGEEND% Replaced By The Split or Extract Index of a file. %BOOKMARK% Replaced by the Bookmark title, Use this	troduction	
Variable     Description       PDF Settings     %VALUE1%       %VALUE2%     %VALUE2%       %FILENAME%     Replaced By the Name of the Original File       %DATE%     Replaced By The Current Date       %TIME%     Replaced By The Current Time       %PAGENUMBER%     Replaced By The Page Number Where the Text/Barcode Was Found       %PAGESTART%     Replaced By The Start Page Number of a Split PDF       %INDEX%     Replaced By The End Page Number of a Split PDF       %BOOKMARK%     Replaced by the Split or Extract Index of a file.		
Variable         Description           PDF Settings         %VALUE1%           %VALUE2%         %VALUE2%           %FILENAME%         Replaced By the Name of the Original File           %DATE%         Replaced By The Current Date           %TIME%         Replaced By The Current Time           %PAGENUMBER%         Replaced By The Page Number Where the Text/Barcode Was Found           %PAGEENDMBER%         Replaced By The Start Page Number of a Split PDF           %PAGEEND%         Replaced By The End Page Number of a Split PDF           %INDEX%         Replaced By The Split or Extract Index of a file.           %BOOKIMARK%         Replaced by the Solit or Extract Index of a file.	lect Variables	
PDF Settings %VALUE2% Sinish %VALUE2% %VALUE2% %VALUE2% %VALUE2% %EleNAME% Replaced By the Name of the Original File %DATE% Replaced By The Current Date %DATE% Replaced By The Current Time %PAGENUMBER% Replaced By The Page Number Where the Text/Barcode Was Found %PAGESTART% Replaced By The Start Page Number of a Split PDF %PAGEEND% Replaced By The Split or Extract Index of a file. %BOOKMARK% Replaced by the Bookmark title, Use this	Variable	Description
%FILENAME%       Replaced By the Name of the Original File         %GATE%       Replaced By The Current Date         %GATE%       Replaced By The Current Time         %FILENAME%       Replaced By The Current Time         %FILENAME%       Replaced By The Start Page Number Where the Text/Barcode Was Found         %PAGEENDMBER%       Replaced By The Start Page Number of a Split PDF         %PAGEEND%       Replaced By The End Page Number of a Split PDF         %INDEX%       Replaced By The Split or Extract Index of a file.         %BOOKMARK%       Replaced by the Bookmark title, Use this	e Naming %value1%	Ì
%FILENAME%         Replaced By the Name of the Original File           %DATE%         Replaced By The Current Date           %DATE%         Replaced By The Current Time           %FAGENUIMBER%         Replaced By The Page Number Where the Text/Barcode Was Found           %PAGESTART%         Replaced By The Start Page Number of a Split PDF           %PAGEEND%         Replaced By The End Page Number of a Split PDF           %INDEX%         Replaced By The Split or Extract Index of a file.           %BOOKMARK%         Replaced by the Bookmark title, Use this	PE Settings %VALUE2%	
%FIME%         Replaced By The Current Time           %FIME%         Replaced By The Page Number Where the Text/Barcode Was Found           %PAGESTART%         Replaced By The Start Page Number of a Split PDF           %PAGEEND%         Replaced By The End Page Number of a Split PDF           %INDEX%         Replaced By The Split or Extract Index of a file.           %BOOKMARK%         Replaced by the Bookmark title, Use this	%FILENAME%	Replaced By the Name of the Original File
%PAGENUMBER%         Replaced By The Page Number Where the Text/Barcode Was Found           %PAGESTART%         Replaced By The Start Page Number of a Split PDF           %PAGEEND%         Replaced By The End Page Number of a Split PDF           %INDEX%         Replaced By The Split or Extract Index of a file.           %BOOKMARK%         Replaced by the Bookmark title, Use this	nish %Date%	Replaced By The Current Date
Text/Barcode Was Found           %PAGESTART%         Replaced By The Start Page Number of a Split PDF           %PAGEEND%         Replaced By The End Page Number of a Split PDF           %INDEX%         Replaced By The Split or Extract Index of a file.           %BOOKMARK%         Replaced by the Bookmark title, Use this	%TIME%	Replaced By The Current Time
Split PDF           %PAGEEND%         Replaced By The End Page Number of a Split PDF           %INDEX%         Replaced By The Split or Extract Index of a file.           %BOOKMARK%         Replaced by the Bookmark title, Use this	%PAGENUMBER%	Replaced By The Page Number Where the Text/Barcode Was Found
Split PDF %INDEX% Replaced By The Split or Extract Index of a file. %BOOKMARK% Replaced by the Bookmark title, Use this	%PAGESTART%	
file. %BOOKMARK% Replaced by the Bookmark title, Use this	%PAGEEND%	
	96INDEX36	Replaced By The Split or Extract Index of a file.
with split by bookmark.	%BOOKMARK%	Replaced by the Bookmark title, Use this with split by bookmark.

4. The "PDF 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.

teps					
ntroduction	Security Properties	OCR Settings Othe	r Settings		
b Definition	Document Security				
lect Variables					
e Naming	Set Security				
lvanced Settings					
nish	Source Password		Target User Password	Target Owner	Password
	Target Document Perm	issions			
	larget Document Pen	lissions			
	Retain Permissions	Assembly Off	Degraded Printing	Modify Annotations	Copy Off
	<b>U</b>			<b>U</b> OII	<b>U</b> OII
	Allow All Permissions	Form Filling	Screen Readers	Modify Contents	Printing
	Off	Off	Off	Off	Off

5. The "Finish" tab will show you a summary of the job you set up

Steps		
Introduction		
Job Definition	Job Definition Job Name: Job 10006 Back to Pr	eview
Select Variables	Select Operation: Rename By Text & Barcode Create Folders: Yes	
File Naming	Max Cores = 1 Output PDF Page Rotation: -1	
Advanced Settings	Locations	
Finish	Source Type : file Source : C:Aquaforest/Kingfisher/QA/input/Text Match/postcode.pdf Destination : C:Aquaforest/Kingfisher/Ugfiles/.Job 10006/%TIMESTAMP%.txt Selected Variables(s) %VALUE1%: Captured Zone : 289.557.06.430.62 Select all text in zone after value : Patent No: Use Barcode : Ves Output File Naming(s) Name Template : %VALUE1% OCR Settings OCR Language: English Despeckie Remove Lines: True Deskew: True	
2		0
Help	Previous Next Preview Save Job	-

6. To confirm that you are happy with the job, click the preview button to see what

Steps	
Introduction	
Job Definition	Back to Summary
Select Variables	Expires On Wednesday, 22 Mar, 2017 (Days Remaining 13)
File Naming	Multi-core License With 8 cores Kingfisher Advanced Edition
Advanced Settings	Running Command File = US0064946781.pdf, Folder = C:\Aquaforest\Kingfisher Exit Code : 0
Finish	Time taken = 6 seconds
1	

7. 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.

Dashboard Job Settings Options Help & Support

+ Add New Job					
OPERATION	JOB ID	JOB NAME	LAST RUN	RUN STATUS	
Rename based on Text Match	10000	Job 10000	24 Feb 2017 15: 35:02	Finished	
Rename based on Barcode	10001	Job 10001	24 Feb 2017 15: 35:22	Finished	
Rename based on Text Match and Barcode	10002	Job 10002	24 Feb 2017 15: 36:06	Finished	
-   - Split File	10003	Job 10003	24 Feb 2017 15: 37:49	Finished	
- Split File Edit Job	10004	<del>split 1 -</del> Click the buttons to - edit the job.	24 Feb 2017 15: 38:02	Finished	
- I- Split File based on Page Ranges	10006	Split 3	15 Feb 2017 17: 37:31	Finished	
😴 Extract Pages from File based on Text Match	10007	Extarct 1	15 Feb 2017 17: 49:12	Finished	
💎 Extract Pages from File based on Barcode	10008	Extract 2	15 Feb 2017 17: 50:44	Finished	
T PDF to Text	10009	Job 10009	15 Feb 2017 17: 53:46	Finished	

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.

### Dashboard Job Settings Options Help & Support

Job Definition Select Variables File Naming Advanced Settings Running Job

Enter the Name of the Job and Select Operation	Location Details			
Job ID	Single File	O Folder	Tree	
10003	Source			
Job Name	C:\Kingfisher QA\Input\	\Text Match\postcode.pd	f	<b></b>
Job 10003	Destination			Same as Source
Select Operation	C:\Kingfisher QA\outpu	t\Text Match		-
Split PDF File	<ul> <li>Log File</li> </ul>			
	C:\Aquaforest\Kingfishe	er\logfiles\ Job 10003\%1	IMESTAMP%.txt	<b>_</b>
Split File based on Text Match	▼ Job Options			
Output File Options	Case Sensitive	Debug		Continue on Error
Page That Matches Text on First Page	Off	Off		Off
	Create Folders	OCR		Overwrite Existing Files
	On	Off		Off
	Max Cores			
	1	•		

# 5 EXTRACT TABULAR DATA FROM PDF

Select Operation			
Export PDF Content			•
PDF to CSV/XLSX			•
Output File		Append	Data to Existing File
.xlsx	•	True	-
			If you want to extract all the files
Append As WorkSheet			



Screen Field / Button	Description
Output file	<ul> <li>csv: Produces a simple csv file.</li> <li>xlsx: Produces and Excel File</li> </ul>
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 excel 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.

Coom In/Out Pages Reload Tables				
	Aquaforest Software	Keywords	Latest Release	•
Use First Table To Derive All Table Off	Autobahn DX	High performance	4.02	
Use Table Lines To Detect Tables	Kingfisher	Extraction	1.1	
Extract Only Grouped Tables	Searchlight	SharePoint	1.2	
● Off Use Word Cordinates To Detect Tables	PDF Junction	PDF	3.4	
On Minimum Table Gap	TIFF Junction	OCR Table ID	: Ó·Ġroup ID : 1	
50	Tabula DX	Search Engine	1.4	
Number of Pages To Check				
Number of Pages To Check	Tabula DX TIFF Server	Search Engine TIFF Viewer	4.0	

Iob Definition Select Variables File Naming Advanced Settings Running Job

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 King fisher 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)	Once the space between two adjacent lines exceed the value given here, the first line will be interpreted as the end of the table.
Table Space (pt)	Once the space between two words exceed 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	Only tables with number of rows greater than or equals to this value will be picked up
Minimum Number of Columns	Only tables with number of columns greater than or equals to this value will be picked up

### 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.

#### Job Definition Select Variables File Naming Advanced Settings Running Job

Document View Tab	le View			
Tables Grouped Tal	oles	Column 1	Column 2	Column 3
1 /3		Aquaforest Software	Keywords	Latest Release
	Tables	Autobahn DX	High performance	4.02
711.00.0		Kingfisher	Extraction	1.1
Table ID: 0		Searchlight	SharePoint	1.2
Table ID: 1		PDF Junction	PDF	3.4
Table ID: 2		TIFF Junction	OCR	4.1
		Tabula DX	Search Engine	1.4
		TIFF Server	TIFF Viewer	4.0
		TIFF Archive	TIFF to PDF	1.0
		OCR SDK	OCR, IRIS	2.2
Tab contain Table found results more	is controls to refine e.			
			f the Selected Ta n in a datagrid.	ble/Grouped
		$\bigcirc$		

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	This the control at the top of the "Tables" tab, it is used to traverse through the tables displayed in the List below. Clicking The up or down button changes the currently selected table which in loads the data grid on the right
The table list	This is the second control; it contains a list of all the table recognized in the document together with their IDs'. Clicking on any one of them will load the selected table in the Data grid on the right.

### 5.2.2 Grouped Tables

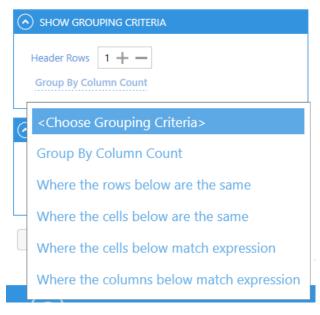
By default, Kingfisher can extract all structured data in a PDF file and write it to a ".csv" or ".xlsx" file. We also understand that users might not want to extract all the tables found in a document or the user might want to group some similar tables into one group with a common header.

Because of the issues raise in the preceding paragraph, 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.

Tables Grouped Tables		Column 1 🧯	Column 2 👔	Column 3 🖋
		Autobahn DX	High performance	4.02
1 /1		Kingfisher	Extraction	1.1
Conversion 10 or 1		Searchlight	SharePoint	1.2
Group ID: 1	Table ID : 0	PDF Junction	PDF	3.4
SHOW GROUPING CRITERI	Table ID : 0	TIFF Junction	OCR	4.1
SHOW GROUPING CRITERI.	Table ID : 1	Tabula DX	Search Engine	1.4
Header Rows 1 🕂 —	Table ID : 2	TIFF Server	TIFF Viewer	4.0
Group By Column Count	lable ID . 2	TIFF Archive	TIFF to PDF	1.0
croup by commit count		OCR SDK	OCR, IRIS	2.2
		Autobahn DX	High performance	4.02
SHOW EXCLUSION CRITERIA		Kingfisher	Extraction	1.1
		Searchlight	SharePoint	1.2
<choose criteria="" exclusion=""></choose>		PDF Junction	PDF	3.4
		TIFF Junction	OCR	4.1
		Tabula DX	Search Engine	1.4
GROUP TA	ABLES	TIFF Server	TIFF Viewer	4.0
	$\nabla$	TIFE Archive	TIFE to PDE	10

### 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 simply 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).

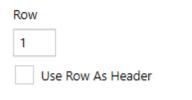
#### 5.2.2.1.2 Where the rows below are the same

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.

#### Where the rows below are the same



#### 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.

#### Where the cells below are the same

Row		Column
2		5
Us	e Row As Header	

#### 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.

Where the ce	lls below match	expression +
Row	Column	Expression
2	5	[0-9]{9}
Use Row /	As Header	

#### 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

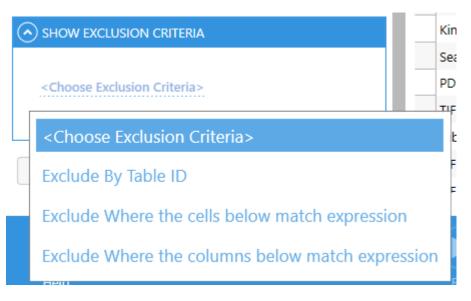
This option will also allow to change the provide the header rows (See 1.2.3).

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

Where the columns below match expr					
Column	Expression				
5	)-9]{9}				
Use Row As Header	1				
Column	Expression				
1	Regex cannot	be Empt			
Use Row As Header	1				

#### 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



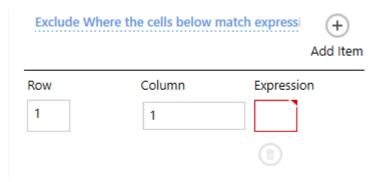
#### 5.2.2.1 Exclude by Table ID

This option simple excludes any table with the ID provided below.



#### 5.2.2.2.2 Exclude cells below match expression

This option excludes all the tables where the cells 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.

Exclude Where	the columns below match exp + Add Item	
Column	Expression	н
1	Regex cannot be Emp	oty

### 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.

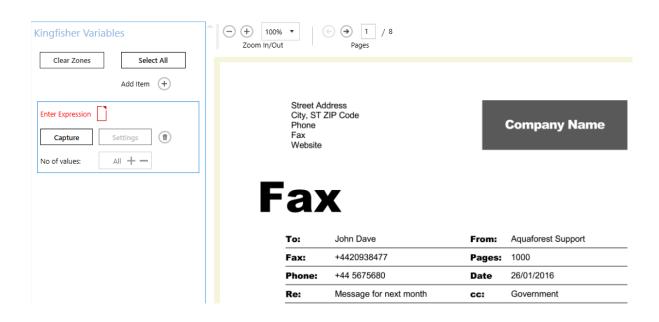
AQUAFOREST SOFTWARE	KEYWORDS	Column 3	1
Autobahn DX	High performance	4.02	
Kingfisher	Extraction	1.1	
Searchlight	SharePoint	1.2	

## **6 ADVANCED EXPORT TO CSV/XLSX**

This is step extract 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 "Smart Text Extraction" step then go to the "Select Variable" tab.



Value	Description
Clear Zone	This clears all the selected zones and values in the tab.
Select All	This button selects the whole page; it is just a quicker and accurate way of selectin a whole page.
Add Item	Click this button to add another expression, this means a new column in the csv output file.
Capture	Clicking the capture button will first assign the selected zone to that column, then it will extract the text based on the setting provided by the user.
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.
Settings	This button allows the user to refine their selection more. It is explained further below.
Delete Button	Clicking this button will delete the expression associated to it.
Zoom In/Out	This will help the user maximize and minimize the loaded document.
Pages	This allows the user browse through the loaded pages.

 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 "John Dave", enter "To:". For this example, I used the "Select All" to select the whole Page.

Kingfisher Variables	- + 100% Zoom In/Out	•	→ 1 / 8 Pages		
Clear Zones Select All Add Item +		Street Add City, ST Z			Common Norra
Capture     appears in the could with with appears in the could with with appears in the could with a point of the could w		Phone Fax Website			Company Name
		<b>a)</b>			
		То:	John Dave	From:	Aquaforest Support
		Fax:	+4420938477	Pages:	1000

#### 3. After you enter the Expression, click on the capture button.

Kingfisher Variables	$\begin{array}{c c} \hline & \hline \\ \hline & \hline \\ Zoom ln/Out \end{array} \qquad $
Clear Zones Select All Add Item (+)	CAPTURED TEXT
Captured To:	John Selected Text Matched Pattern.
Capture Settings	
No of values: All + -	
	done

- 4. As you can see by default Kingfisher will extract the word that appears after the Expression provided, if you want to modify the value extracted some more, click on the settings button.
- 5. Clicking the settings button will show the window below.

Select all text in zone after value DATE
Where text matches any pattern
Use This Value For All Rows Select Multiple Lines In Zone
Show Help Tips Off
apply changes discard changes

Value	Description
Text Position	This is the first Combo Box, see step 6 for more details on this control
Pattern	This is the second Combo Box, see step 7 for more details on this control
Use This Value for all Rows	Switch this on if you want a value like "Invoice Number" that appears only once in a page to be copied across all rows extracted from that page in the csv file.
Select Multiple Lines In Zone	This option is only available when you are selecting all text the text in a zone that appears before/after the text expression. If set to on, each line in the zone will be extracted as a different value. When set to off all the lines will be extracted as a single value.

#### 6. Below is a description of the text selection combo box

# Select all text in line after value **To**:

Select word after value

Select word before value

Select all text in line after value

Select all text in line before value

Select all text in zone after value

Select all text in zone before value

Value	Description
Select word after value	This picks the word that appears on the right side of the expression.
Select word before value	This picks the word that appears on the left side of the expression.
Select all text in line after value	This picks all the words that appears in the line on the right side of the expression.
Select all text in line before value	This picks all the words that appears in the line on the left side of the expression.
Select all text in zone after value	This picks all the words that appears in the zone selected after the expression.
Select all text in zone before value	This picks all the words that appears in the zone selected before the expression.

Note: If you have more than one Expression added, the selected string will be truncated at the first occurrence of any of the chosen expressions.

7. Kingfisher also allow users to provide regular expressions, the first string the matches that expression will be returned as the value. This is the ultimate way users can define exactly what they want. We even provided a "Show Help Tips" switch to show the user the basic about regular expressions, you can find out more online. For tis example we will not be using regular expressions.

Where text matches pattern
Where text matches any pattern
Where text matches pattern

#### Example Patterns

Patterns are used to match parts of the document. Some useful examples are below.

PATTERN	MATCHES	EXAMPLE
invoice		The word "invoice"
invoice no:		The phrase "invoice no:"
	А	
	9	Any 1 character
	%	
	10-h	

#### 8. You can add several Expressions before you run the job.

	Add Item (+)	
Captured To:		
Capture	Settings (1)	
No of values:	All 🕂 💳	
Captured From:		
Capture	Settings	
No of values:	All +	
Captured Fax:		
Capture	Settings (1)	

Fax	K						
To:	John Dave		Fi	rom:	Aquaforest	t Support	
Fax:	+4420938477		Pa	ages:	1000		
Phone:	+44 5675680		D	ate	26/01/2016	6	
Re:	Message for ne	xt month	C		Governme	nt	
Urgent	□ For Review	Please	Comment	🗆 Pl	ease Reply	□ Please	e Recycle
Comments							

Comments:

To get started right away, just tap any placeholder text (such as this) and start typing to replace it with your own.

Want to insert a nicture from your files or add a shape, text box, or table? You not

#### 9. Below is the expected output

Clip	Clipboard 🕞 Font		G Align	ment 🗔
A	L *	: × ✓ fx	To:	
	А	В	с	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
	Sheet 1 (+)			

## **7 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:

	A sample job me is shown here.
	<pdfsplitterjob></pdfsplitterjob>
	<operation>splitbarcode</operation>
	<conditionname> Job 10000</conditionname>
	<jobspecialoptions></jobspecialoptions>
	<joboptions> -language 0</joboptions>
	<pre><pre>conting of the second s</pre></pre>
	<splitrange> </splitrange>
	<frompage> </frompage>
	<topage> </topage>
	<splitmode> -SplitMode=0</splitmode>
	<splitrepeatingrange> </splitrepeatingrange>
	<maxcores>1</maxcores>
	<condition>barcode</condition>
	<useregex3>False</useregex3>
	<usenametemplate>False</usenametemplate>
	<useregex1>False</useregex1>
	<useregex2>False</useregex2>
	<useregex3>False</useregex3>
	<useregex4>False</useregex4>
	<outputnamemask>%VALUE1%_%VALUE2%</outputnamemask>
	<sourcefiles>C:\Users\dauda\Desktop\kingfisher\patents.pdf</sourcefiles>
	<password> </password>
	<target>C:\Users\dauda\Desktop\kingfisher\split</target>
	<sourcetype>file</sourcetype>
	<casesensitive>False</casesensitive>
	<advancedflags> </advancedflags>
	<metadata> </metadata>
	<security> </security>
	<docoptions> </docoptions>
	<value1location>all</value1location>
	<value1expression> </value1expression>
	<value2location>601.76:111.55:121.33:26.96</value2location>
	<value2expression> </value2expression>
	<value3location> </value3location>
	<value3expression> </value3expression>
	<value4location> </value4location>
	<value4expression> </value4expression>
	<value1format> </value1format>
	<value1position> </value1position>
	<value2format> </value2format>
	<value2position> </value2position>
	<value3format> </value3format>
	<value3position> </value3position>
	<value4format> </value4format>
ļ	<value4position> </value4position>

<barcodeformat1> </barcodeformat1>
<jobid>10000</jobid>
<barcodeformat2> </barcodeformat2>
<barcodeformat3> </barcodeformat3>
<barcodeformat4> </barcodeformat4>
<usebarcode1>True</usebarcode1>
<usebarcode2>False</usebarcode2>
<usebarcode3>True</usebarcode3>
<usebarcode4>True</usebarcode4>
<outputscript> </outputscript>
<onerrorcontinue>false</onerrorcontinue>
<errors> </errors>
<mergeextarctedpagesmatch>False</mergeextarctedpagesmatch>
<mergeextarctedpages> </mergeextarctedpages>
<logfile>C:\Aquaforest\Kingfisher\logfiles\ Job 10000\%TIMESTAMP%.txt</logfile>
<lastlogfile>C:\Aquaforest\Kingfisher\logfiles\ Job 10000\12-08-22 PM.txt</lastlogfile>
<runstatus>C:\Aquaforest\Kingfisher\status\10000.txt</runstatus>
<lastrundate>10 Jun 2016 12: 08:22</lastrundate>
<overwriteexisting>False</overwriteexisting>
<overwriteexistingcurrent>False</overwriteexistingcurrent>
<debug>False</debug>
<doocr>True</doocr>
<createfolders>True</createfolders>
<continueonerror>False</continueonerror>

</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.		
Value1location         The zones that correspond to the "%VALUE1%", "%VALUE		
Value2location "%VALUE3%" and "%VALUE4%" variables respectively. Descr		
Value3location	detail in Section 6.2.	
Value4location		

value1expression	The Text Expression that correspond to the "%VALUE1%",
value1expression	"%VALUE2%", "%VALUE3%" and "%VALUE4%" variables
value1expression	respectively. Described in detail in Section 6.2.
value1expression	
value1format	The Text Format that correspond to the "%VALUE1%",
value1format	"%VALUE2%", "%VALUE3%" and "%VALUE4%" variables
value1format	respectively. Described in detail in Section 6.2.
value1format	
Value1position	The Relative Position that correspond to the "%VALUE1%",
Value1position	"%VALUE2%", "%VALUE3%" and "%VALUE4%" variables
Value1position	respectively. Described in detail in Section 6.2.
Value1position	
usebarcode2	Boolean values that instruct Kingfisher to use search for barcodes
usebarcode3	instead of text for "%VALUE2%", "%VALUE3%" and "%VALUE4%"
usebarcode4	variables respectively. Described in detail in Section 6.2.
useregex1	Boolean values that instruct Kingfisher to use Regular Expressions
useregex2	instead of Text Format for validating "%VALUE1%", "%VALUE2%",
useregex3	"%VALUE3%" and "%VALUE4%" variables respectively. Described in
useregex4	detail in Section 6.2.
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

### **8** 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"

### 8.1 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.

Usage:

Kingfisher.Commandline.exe -jobfile "..\\conditions\\1001.xml"

### 8.2 Command Line Syntax

Kingfisher can also be ran using command line arguments, this is suitable for users who want to dynamically generate commands at runtime.

Usage:

#### Kingfisher.Commandline.exe [Execargs]

Parameter	Notes		
-source	Path to the source file or folder		
-SourceType	Specifies if the source is a file, folder or tree. The values are:		
	• file		
	• folder		
	• tree		
-SourceFolder	The original source folder, this is important for tree based operations		
-t	The target Folder		
-outputnamemask	The Name Template discussed in section 4.1.5		
-operation	Tells Kingfisher the operation to be executed.		
	renametext		
	renamebarcode		
	renamemix		
	extract		
	extractbarcode		
	• splittext		
	splitchange		

<ul> <li>splitrange         <ul> <li>extracttext</li> <li>case</li> <li>Kingfisher acts in a case insensitive manner when matching patterns and expressions. Setting this flag to true will make it case sensitive.</li> <li>-debug (bool)</li> <li>Set this to true to see extra debug information</li> <li>condition</li> <li>Specifies the entity that should be extracted, the options are                 <ul> <li>text</li> <li>barcode</li> <li>angle</li></ul></li></ul></li></ul>	[			
<ul> <li>extracttext</li> <li>ccase</li> <li>Kingfisher acts in a case insensitive manner when matching patterns and expressions. Setting this flag to true will make it case sensitive.</li> <li>debug (bool)</li> <li>Set this to true to see extra debug information</li> <li>condition</li> <li>Specifies the entity that should be extracted, the options are         <ul> <li>text</li> <li>barcode</li> </ul> </li> <li>angle</li> <li>All the pages in the output PDF files will have this angle.</li> <li>logFile</li> <li>Path to the log file for the operation.</li> <li>obfile</li> <li>The tocation of the job definition xml file that was created in the kingfisher GUI</li> <li>ocr (bool)</li> <li>If set to true, kingfisher will search for text on image PDFs too.</li> <li>overwrite(bool)</li> <li>If set to true, kingfisher will straighten the image.</li> <li>despeckle (int)</li> <li>Removes dots below the specified pixel size options are from 1-9</li> <li>temp</li> <li>A path where kingfisher can perform all its temporary operations.</li> <li>language</li> <li>OCR Language dictionary to be used. The default is English.</li> <li>LANGUAGE</li> <li><i>language Flag Value</i></li> <li>English</li> <li>German</li> <li>French</li> <li>Russian</li> <li>Swedish</li> <li>Spanish</li> <li>Serbian</li> <li>Croatian</li> <li>Portuguese</li> <li>13</li> <li>Dutch</li> <li>14</li> <li>Czech</li> <li>Roman</li> <li>I6</li> <li>Hungar</li> <li>Hungar</li> <li>Hungar</li> <li>I1</li> <li>Danish</li> <li>I2</li> <li>Portuguese</li> <li>Blugar</li> <li>Slovenian</li> <li>Ia</li> <li>Bugar</li> <li>Slovenian</li> <li>Latvian</li> <li>Estonian</li> <li>Latvian</li> <li>Estonian</li> <li>Latvian</li> <li>Estonian</li> <li>Intrkish</li> </ul>		splitbarcode		
-case Kingfisher acts in a case insensitive manner when matching patterns and expressions. Setting this flag to true will make it case sensitive. -debug (bool) Set this to true to see extra debug information -condition Specifies the entity that should be extracted, the options are • text • barcode -angle All the pages in the output PDF files will have this angle. -logFile Path to the log file for the operation. -logFile Path to the log file for the operation. -logFile The Location of the job definition xml file that was created in the kingfisher GUI -ocr (bool) If set to true, kingfisher will search for text on image PDFs too. -overwrite(bool) If set to true, kingfisher shill earch for text on image PDFs too. -overwrite(bool) If set to true, kingfisher can perform all its temporary operations. -despeckle (int) Removes dots below the specified pixel size options are from 1-9 -temp A path where kingfisher can perform all its temporary operations. -language OCR Language dictionary to be used. The default is English. <b>LANGUAGE</b> /language flag Value English 0 German 1 French 2 Russian 3 Swedish 4 Spanish 5 Italian 6 Russian 10 Protatian 9 Croatian 10 Polish 11 Danish 12 Portuguese 13 Dutch 14 Czech 15 Roman 16 Hungar 17 Bulgar 18 Slovenian 19 Latvian 20 Lithuanian 21 Estonian 22 Turkish 23 - Retain Bookmark/s				
and expressions. Setting this flag to true will make it case sensitive.         -debug (bool)       Set this to true to see extra debug information         -condition       Specifies the entity that should be extracted, the options are <ul> <li>text</li> <li>barcode</li> </ul> <li>angle</li> <li>All the pages in the output PDF files will have this angle.</li> <li>logFile</li> <li>Path to the log file for the operation.</li> <li>lobFile</li> <li>The Location of the job definition xml file that was created in the kingfisher GUI</li> <li>ocr (bool)</li> <li>If set to true, kingfisher will overwrite any file with the same name as the output file, else the job will error out.</li> <li>deskew (bool)</li> <li>If set to true, kingfisher will straighten the image.</li> <li>despeckle (int)</li> <li>Removes dots below the specified pixel size options are from 1-9</li> <li>temp</li> <li>A path where kingfisher can perform all its temporary operations.</li> <li>OCR Language dictionary to be used. The default is English.</li> <li>LANGUAGE</li> <li>/language Flag Value</li> <li>English</li> <li>German</li> <li>French</li> <li>Russian</li> <li>Swedish</li> <li>4</li> <li>Spanish</li> <li>Statian English</li> <li>Russian</li> <li>Serbian</li> <li>Portuguese</li> <li>13</li> <li>Dutch</li> <li>Roman</li> <li>Ide</li> <li>Roman</li> <li>Ide</li> <li>Roman</li> <li>Extonian</li> <li>Ide</li> <li>Roman</li> <li>Ide</li> <li>Sovenian</li> <li>Ige</li> <li>Sovenian</li> <li>Ige</li> <li>Latvian</li> <li>Ide</li> <li>Roman</li> <li>German</li> <li>Extonian</li> <li>Craotian</li> <li>Ide</li> <li>Roman</li> <li>Ide</li> <li>Roman</li> <li>Ide</li> <li>Roman</li> <li>Ide</li> <li>Ro</li>		extracttext		
and expressions. Setting this flag to true will make it case sensitive.         -debug (bool)       Set this to true to see extra debug information         -condition       Specifies the entity that should be extracted, the options are <ul> <li>text</li> <li>barcode</li> </ul> <li>angle</li> <li>All the pages in the output PDF files will have this angle.</li> <li>logFile</li> <li>Path to the log file for the operation.</li> <li>lobFile</li> <li>The Location of the job definition xml file that was created in the kingfisher GUI</li> <li>ocr (bool)</li> <li>If set to true, kingfisher will overwrite any file with the same name as the output file, else the job will error out.</li> <li>deskew (bool)</li> <li>If set to true, kingfisher will straighten the image.</li> <li>despeckle (int)</li> <li>Removes dots below the specified pixel size options are from 1-9</li> <li>temp</li> <li>A path where kingfisher can perform all its temporary operations.</li> <li>OCR Language dictionary to be used. The default is English.</li> <li>LANGUAGE</li> <li>/language Flag Value</li> <li>English</li> <li>German</li> <li>French</li> <li>Russian</li> <li>Swedish</li> <li>4</li> <li>Spanish</li> <li>Statian English</li> <li>Russian</li> <li>Serbian</li> <li>Portuguese</li> <li>13</li> <li>Dutch</li> <li>Roman</li> <li>Ide</li> <li>Roman</li> <li>Ide</li> <li>Roman</li> <li>Extonian</li> <li>Ide</li> <li>Roman</li> <li>Ide</li> <li>Sovenian</li> <li>Ige</li> <li>Sovenian</li> <li>Ige</li> <li>Latvian</li> <li>Ide</li> <li>Roman</li> <li>German</li> <li>Extonian</li> <li>Craotian</li> <li>Ide</li> <li>Roman</li> <li>Ide</li> <li>Roman</li> <li>Ide</li> <li>Roman</li> <li>Ide</li> <li>Ro</li>	-0360	Kingfisher acts in a c	ase insensitive manner when matching patterns	
-debug (bool) -condition Sectifies the entity that should be extracted, the options are - text - barcode -angle -angle -All the pages in the output PDF files will have this anglelogFile Path to the log file for the operationlopFile The Location of the job definition xml file that was created in the kingfisher GUI -ocr (bool) If set to true, kingfisher will search for text on image PDFs toooverwrite(bool) If set to true, kingfisher will search for text on image PDFs toooverwrite(bool) If set to true, kingfisher will straighten the imagedespeckle (int) Removes dots below the specified pixel size options are from 1-9 -despeckle (int) Removes dots below the specified pixel size options are from 1-9 -temp A path where kingfisher can perform all its temporary operationslanguage OCR Language dictionary to be used. The default is English.  IANGUAGE Inguish I Cancenter Removes dots below the specified pixel size options are from 1-9 -temp A path where kingfisher an perform all its temporary operationslanguage OCR Language file Value English O German 1 French Z Russian Swedish Spanish S Italian 6 Russian Swedish 4 Spanish Serbian 9 Croatian 10 Polish 11 Danish 12 Portuguese 13 Dutch 144 Czech 15 Roman 16 Hungar 17 Bulgar 18 Slovenian 19 Latvian 20 Lithuanian 21 Estonian 22 Turkish 23	-case	-		
-condition Specifies the entity that should be extracted, the options are • text • barcode -angle All the pages in the output PDF files will have this anglelogFile Path to the log file for the operationlobFile The Location of the job definition xml file that was created in the kingfisher GUI -ocr (bool) If set to true, kingfisher will search for text on image PDFs toooverwrite(bool) If set to true, kingfisher will overwrite any file with the same name as the output file, else the job will error outdeskew (bool) If set to true, kingfisher will straighten the imagedespeckle (int) Removes dots below the specified pixel size options are from 1-9 -temp A path where kingfisher can perform all its temporary operationslanguage OCR Language dictionary to be used. The default is EnglishIANGUAGE //anguage Flag Value English 0 German 1 French 2 Russian 3 Swedish 4 Spanish 5 Italian 6 Russian English 7 Ukrainian 8 Serbian 9 Croatian 10 Polish 11 Danish 12 Portuguese 13 Dutch 14 Czech 15 Roman 16 Hungar 17 Bulgar 18 Slovenian 19 Latvian 20 Lithuanian 21 Estonian 22 Turkish 23 Retain Bookmarks retainbookmark(boo				
-condition Specifies the entity that should be extracted, the options are • text • barcode -angle All the pages in the output PDF files will have this anglelogFile Path to the log file for the operationlobFile The Location of the job definition xml file that was created in the kingfisher GUI -ocr (bool) If set to true, kingfisher will search for text on image PDFs toooverwrite(bool) If set to true, kingfisher will overwrite any file with the same name as the output file, else the job will error outdeskew (bool) If set to true, kingfisher will straighten the imagedespeckle (int) Removes dots below the specified pixel size options are from 1-9 -temp A path where kingfisher can perform all its temporary operationslanguage OCR Language dictionary to be used. The default is EnglishIANGUAGE //anguage Flag Value English 0 German 1 French 2 Russian 3 Swedish 4 Spanish 5 Italian 6 Russian English 7 Ukrainian 8 Serbian 9 Croatian 10 Polish 11 Danish 12 Portuguese 13 Dutch 14 Czech 15 Roman 16 Hungar 17 Bulgar 18 Slovenian 19 Latvian 20 Lithuanian 21 Estonian 22 Turkish 23 Retain Bookmarks retainbookmark(boo	-debug (bool)	Set this to true to se	e extra debug information	
<ul> <li>text         <ul> <li>barcode</li> </ul> </li> <li>-angle             <ul> <li>All the pages in the output PDF files will have this angle.</li> <li>-logFile             <ul></ul></li></ul></li></ul>	<b>.</b>		-	
-angle       All the pages in the output PDF files will have this angle.         -logFile       Path to the log file for the operation.         -JobFile       The Location of the job definition mul file that was created in the kingfisher GUI         -ocr (bool)       If set to true, kingfisher will search for text on image PDFs too.         -overwrite(bool)       If set to true, kingfisher will overwrite any file with the same name as the output file, else the job will error out.         -deskew (bool)       If set to true, kingfisher will Straighten the image.         -deskew (bool)       Removes dots below the specified pixel size options are from 1-9         -temp       A path where kingfisher can perform all its temporary operations.         -language       OCR Language dictionary to be used. The default is English.         Image       OCR Language dictionary to be used. The default is English.         Italian       6         Russian       3         Swedish       4         Spanish       5         Italian       6         Russian English       7         Ukrainian       8         Serbian       9         Croatian       10         Portuguese       13         Dutch       14         Czech       15         Roman       16				
-logFile       Path to the log file for the operation.         -JobFile       The Location of the job definition xml file that was created in the kingfisher GUI         -ocr (bool)       If set to true, kingfisher will search for text on image PDFs too.         -overwrite(bool)       If set to true, kingfisher will overwrite any file with the same name as the output file, else the job will error out.         -deskew (bool)       If set to true, kingfisher will Straighten the image.         -deskew (bool)       If set to true, kingfisher can perform all its temporary operations.         -despeckle (int)       Removes dots below the specified pixel size options are from 1-9         -temp       A path where kingfisher can perform all its temporary operations.         -language       OCR Language dictionary to be used. The default is English.         URANGUAGE       /language Flag Value         English       0         German       1         French       2         Russian English       7         Ukrainian       8         Serbian       9         Croatian       10         Polish       11         Danish       12         Portuguese       13         Dutch       14         Czech       15         Roman       16		<ul> <li>barcode</li> </ul>		
-logFile       Path to the log file for the operation.         -JobFile       The Location of the job definition xml file that was created in the kingfisher GUI         -ocr (bool)       If set to true, kingfisher will search for text on image PDFs too.         -overwrite(bool)       If set to true, kingfisher will overwrite any file with the same name as the output file, else the job will error out.         -deskew (bool)       If set to true, kingfisher will Straighten the image.         -deskew (bool)       If set to true, kingfisher can perform all its temporary operations.         -despeckle (int)       Removes dots below the specified pixel size options are from 1-9         -temp       A path where kingfisher can perform all its temporary operations.         -language       OCR Language dictionary to be used. The default is English.         -language       OCR Language files Value         English       0         German       1         French       2         Russian English       7         Ukrainian       8         Serbian       9         Croatian       10         Portiguese       13         Dutch       14         Czech       15         Roman       16         Hungar       17         Bulgar       18 <td>-angle</td> <td>All the pages in the o</td> <td>output PDF files will have this angle.</td>	-angle	All the pages in the o	output PDF files will have this angle.	
-JobFile The Location of the job definition xml file that was created in the kingfisher GUI -ocr (bool) If set to true, kingfisher will search for text on image PDFs toooverwrite(bool) If set to true, kingfisher will overwrite any file with the same name as the output file, else the job will error outdeskew (bool) If set to true, kingfisher will Straighten the imagedespeckle (int) Removes dots below the specified pixel size options are from 1-9 -temp A path where kingfisher can perform all its temporary operationsdeskew (bool) CR Language dictionary to be used. The default is English.           LANGUAGE       /language Flag Value         English       0         German       1         French       2         Russian       3         Swedish       4         Spanish       5         Italian       6         Russian English       7         Ukrainian       8         Serbian       9         Croatian       10         Polish       11         Danish       12         Portuguese       13         Dutch       14         Czech       15         Roman       16         Hungar       17         Bulgar       18         Slovenian       19         Latvian       20         Lithuanian       21         Estonian	-			
-ocr (bool)       If set to true, kingfisher will search for text on image PDFs too.         -overwrite(bool)       If set to true, kingfisher will overwrite any file with the same name as the output file, else the job will error out.         -deskew (bool)       If set to true, kingfisher will Straighten the image.         -despeckle (int)       Removes dots below the specified pixel size options are from 1-9         -temp       A path where kingfisher can perform all its temporary operations.         -language       OCR Language dictionary to be used. The default is English.         -language       OCR Language dictionary to be used. The default is English.         -kussian       1         French       2         Russian       3         Swedish       4         Spanish       5         Italian       6         Russian English       7         Ukrainian       8         Serbian       9         Croatian       10         Polish       11         Danish       12         Portuguese       13         Dutch       14         Czech       15         Roman       16         Hungar       17         Bulgar       18         Slovenian	-JobFile	_	•	
-overwrite(bool) If set to true, kingfisher will overwrite any file with the same name as the output file, else the job will error out. -deskew (bool) If set to true, kingfisher will Straighten the image. -despeckle (int) Removes dots below the specified pixel size options are from 1-9 -temp A path where kingfisher can perform all its temporary operations. -language OCR Language dictionary to be used. The default is English. <b>LANGUAGE</b> /language Flag Value English 0 German 1 French 2 Russian 3 Swedish 4 Spanish 5 Italian 6 Russian English 7 Ukrainian 8 Serbian 9 Croatian 10 Polish 11 Danish 12 Portuguese 13 Dutch 14 Czech 15 Roman 16 Hungar 17 Bulgar 18 Slovenian 19 Latvian 20 Lithuanian 21 Estonian 21 Estonian 21 Estonian 22 Turkish 23				
-overwrite(bool) If set to true, kingfisher will overwrite any file with the same name as the output file, else the job will error out. -deskew (bool) If set to true, kingfisher will Straighten the image. -despeckle (int) Removes dots below the specified pixel size options are from 1-9 -temp A path where kingfisher can perform all its temporary operations. -language OCR Language dictionary to be used. The default is English. <b>LANGUAGE</b> /language Flag Value English 0 German 1 French 2 Russian 3 Swedish 4 Spanish 5 Italian 6 Russian English 7 Ukrainian 8 Serbian 9 Croatian 10 Polish 11 Danish 12 Portuguese 13 Dutch 14 Czech 15 Roman 16 Hungar 17 Bulgar 18 Slovenian 19 Latvian 20 Lithuanian 21 Estonian 21 Estonian 21 Estonian 22 Turkish 23	-ocr (bool)	If set to true, kingfis	ner will search for text on image PDFs too.	
<ul> <li>the output file, else the job will error out.</li> <li>-deskew (bool)</li> <li>If set to true, kingfisher will Straighten the image.</li> <li>-despeckle (int)</li> <li>Removes dots below the specified pixel size options are from 1-9</li> <li>A path where kingfisher can perform all its temporary operations.</li> <li>OCR Language dictionary to be used. The default is English.</li> <li>UANGUAGE //anguage Flag Value</li> <li>English</li> <li>OC German</li> <li>French</li> <li>Russian</li> <li>Swedish</li> <li>Spanish</li> <li>Italian</li> <li>G</li> <li>Russian English</li> <li>Ukrainian</li> <li>Serbian</li> <li>Croatian</li> <li>Polish</li> <li>11</li> <li>Danish</li> <li>12</li> <li>Portuguese</li> <li>13</li> <li>Dutch</li> <li>14</li> <li>Czech</li> <li>Roman</li> <li>16</li> <li>Hungar</li> <li>17</li> <li>Bulgar</li> <li>Slovenian</li> <li>19</li> <li>Latvian</li> <li>20</li> <li>Lithuanian</li> <li>21</li> <li>Turkish</li> <li>23</li> </ul>		-	-	
-deskew (bool) If set to true, kingfisher will Straighten the image. -despeckle (int) Removes dots below the specified pixel size options are from 1-9 -temp A path where kingfisher can perform all its temporary operations. -language OCR Language dictionary to be used. The default is English. LANGUAGE       /language Flag Value         English       0         German       1         French       2         Russian       3         Swedish       4         Spanish       5         Italian       6         Russian       9         Croatian       10         Polish       11         Danish       12         Portuguese       13         Dutch       14         Czech       15         Roman       16         Hungar       17         Bulgar       18         Slovenian       19         Latvian       20         Lithuanian       21         Estonian       22         Turkish       23			•	
-temp A path where kingfisher can perform all its temporary operations. -language OCR Language dictionary to be used. The default is English. OCR Language dictionary to be used. The default is English. CANGUAGE /language Flag Value English 0 German 1 French 2 Russian 3 Swedish 4 Spanish 5 Italian 6 Russian English 7 Ukrainian 8 Serbian 9 Croatian 10 Polish 11 Danish 12 Portuguese 13 Dutch 14 Czech 15 Roman 16 Hungar 17 Bulgar 18 Slovenian 19 Latvian 20 Lithuanian 21 Estonian 22 Turkish 23 - Retain Bookmarks	-deskew (bool)			
-language OCR Language dictionary to be used. The default is English.          LANGUAGE       /language Flag Value         English       0         German       1         French       2         Russian       3         Swedish       4         Spanish       5         Italian       6         Russian English       7         Ukrainian       8         Serbian       9         Croatian       10         Polish       11         Danish       12         Portuguese       13         Dutch       14         Czech       15         Roman       16         Hungar       17         Bulgar       18         Slovenian       19         Latvian       20         Lithuanian       21         Estonian       22         Turkish       23	-despeckle (int)	Removes dots below	the specified pixel size options are from 1-9	
- retainbookmarks	-temp	A path where kingfis	her can perform all its temporary operations.	
English0German1French2Russian3Swedish4Spanish5Italian6Russian English7Ukrainian8Serbian9Croatian10Polish11Danish12Portuguese13Dutch14Czech15Roman16Hungar17Bulgar18Slovenian19Latvian20Lithuanian21Estonian22Turkish23	-language	OCR Language dictio	nary to be used. The default is English.	
English0German1French2Russian3Swedish4Spanish5Italian6Russian English7Ukrainian8Serbian9Croatian10Polish11Danish12Portuguese13Dutch14Czech15Roman16Hungar17Bulgar18Slovenian19Latvian20Lithuanian21Estonian22Turkish23				
German1French2Russian3Swedish4Spanish5Italian6Russian English7Ukrainian8Serbian9Croatian10Polish11Danish12Portuguese13Dutch14Czech15Roman16Hungar17Bulgar18Slovenian19Latvian20Lithuanian21Estonian22Turkish23			/language Flag Value	
- rench 2 French 2 Russian 3 Swedish 4 Spanish 5 Italian 6 Russian English 7 Ukrainian 8 Serbian 9 Croatian 10 Polish 11 Danish 12 Portuguese 13 Dutch 14 Czech 15 Roman 16 Hungar 17 Bulgar 18 Slovenian 19 Latvian 20 Lithuanian 21 Estonian 22 Turkish 23		English	0	
Russian3Swedish4Spanish5Italian6Russian English7Ukrainian8Serbian9Croatian10Polish11Danish12Portuguese13Dutch14Czech15Roman16Hungar17Bulgar18Slovenian19Latvian20Lithuanian21Estonian22Turkish23		German	1	
Swedish 4 Spanish 5 Italian 6 Russian English 7 Ukrainian 8 Serbian 9 Croatian 10 Polish 11 Danish 12 Portuguese 13 Dutch 14 Czech 15 Roman 16 Hungar 17 Bulgar 18 Slovenian 19 Latvian 20 Lithuanian 21 Estonian 22 Turkish 23 Parture Retain Bookmarks		French	2	
Spanish5Italian6Russian English7Ukrainian8Serbian9Croatian10Polish11Danish12Portuguese13Dutch14Czech15Roman16Hungar17Bulgar18Slovenian19Latvian20Lithuanian21Estonian23		Russian	3	
Italian6Russian English7Ukrainian8Serbian9Croatian10Polish11Danish12Portuguese13Dutch14Czech15Roman16Hungar17Bulgar18Slovenian19Latvian20Lithuanian21Estonian23		Swedish	4	
Russian English       7         Ukrainian       8         Serbian       9         Croatian       10         Polish       11         Danish       12         Portuguese       13         Dutch       14         Czech       15         Roman       16         Hungar       17         Bulgar       18         Slovenian       19         Latvian       20         Lithuanian       21         Estonian       23		Spanish	5	
- Retain Bookmarks		Italian	6	
- Retain Bookmarks		Russian English	7	
- Retain Bookmarks		Ukrainian	8	
Polish 11 Danish 12 Portuguese 13 Dutch 14 Czech 15 Roman 16 Hungar 17 Bulgar 18 Slovenian 19 Latvian 20 Lithuanian 21 Estonian 22 Turkish 23		Serbian	9	
- Retain Bookmarks		Croatian	10	
- Retain Bookmarks		Polish	11	
- Retain Bookmarks		Danish	12	
- Retain Bookmarks		Portuguese	13	
Roman16Hungar17Bulgar18Slovenian19Latvian20Lithuanian21Estonian22Turkish23		Dutch	14	
Hungar       17         Bulgar       18         Slovenian       19         Latvian       20         Lithuanian       21         Estonian       22         Turkish       23		Czech	15	
- retainbookmark(boo		Roman	16	
- retainbookmark(boo		Hungar	17	
- retainbookmark(boo		Bulgar	18	
- retainbookmark(boo		Slovenian	19	
- retainbookmark(boo		Latvian	20	
- Retain Bookmarks		Lithuanian	21	
- Retain Bookmarks retainbookmark(boo			22	
retainbookmark(boo		Turkish	23	
retainbookmark(boo				
retainbookmark(boo				
<u>1)</u>				
	1)			

-	Retain Metadata	
retainmetadata(bool		
)		
-createfolders(bool)	Create Folders	
- maxcores (int)	Specifies the number files to be processed in parallel, note this will	
	be limited by the license type and number of processors on your	
-1	machine.	
-z1	Represent the zone for the variable "%VALUE1%" discussed in section 2.2, an example is "-z1=354:15:366:49" where:	
	<ul> <li>354 = The X coordinate of the TopLeft Corner</li> </ul>	
	<ul> <li>15 = The Y coordinate of the TopLeft Corner</li> </ul>	
	<ul> <li>366 = Width</li> </ul>	
	• 49 = Height	
-z2	Represent the zone for the variable "%VALUE2%" discussed in	
	section 2.2	
-z3	Represent the zone for the variable "%VALUE3%" discussed in	
	section 2.2	
-z4	Represent the zone for the variable "%VALUE4%" discussed in	
h . 4 /l 1)	section 2.2	
-ubc1 (bool)	If set to true, kingfisher will check for barcodes instead of text values	
ubc2 (bool)	when extracting text for "%VALUE1%" discussed in section 2.2	
-ubc2 (bool)	If set to true, kingfisher will check for barcodes instead of text values when extracting text for "%VALUE2%" discussed in section 2.2	
-ubc3 (bool)	If set to true, kingfisher will check for barcodes instead of text values	
	when extracting text for "%VALUE3%" discussed in section 2.2	
-ubc4 (bool)	If set to true, kingfisher will check for barcodes instead of text values	
	when extracting text for "%VALUE4%" discussed in section 2.2	
-te1	This string value represents the text value used for the Text Match of	
	"%VALUE1%" discussed in section 2.2	
-te2	This string value represents the text value used for the Text Match of	
	"%VALUE2%" discussed in section 2.2	
-te3	This string value represents the text value used for the Text Match of	
	"%VALUE3%" discussed in section 2.2	
-te4	This string value represents the text value used for the Text Match of "%VALUE4%" discussed in section 2.2	
tn1	This string value represents the position used for the Text Match for	
-tp1	"%VALUE1%" discussed section 2.2. Below is a list of accepted	
	values:	
	O Select word after value	
	1 Select word before value	
	• 2 Select all text in line after value	
	3 Select all text in line before value	
	4 Select all text in zone after value	
	<ul> <li>5 Select all text in zone before value</li> </ul>	
	6 Select text in zone	
-tp2	This string value represents the position used for the Text Match for	
t=2	"%VALUE2%" discussed section 2.2	
-tp3	This string value represents the position used for the Text Match for "%VALUE3%" discussed section 2.2	
-tp4	This string value represents the position used for the Text Match for	
יף <del>י</del> 	"%VALUE4%" discussed section 2.2	
-tf1	This string value represents the Regex Pattern for "%VALUE1%"	
L	55	

	discussed in section 2.2
-tf2	This string value represents the Regex Pattern for "%VALUE2%"
	discussed in section 2.2
-tf3	This string value represents the Regex Pattern for "%VALUE3%"
	discussed in section 2.2
-tf4	This string value represents the Regex Pattern for "%VALUE4%"
	discussed in section 2.2
-barcodeformat1	This is an integer representing the barcode type you want kingfisher
	to look for when finding %VALUE1%, you can include multiple
	formats separated by commas. The formats available a listed below
	with their integer id
	• 1 AZTEC
	• 2 CODABAR
	• 4 CODE_39
	• 8 CODE_93
	• 16 CODE_128
	32 DATA_MATRIX
	• 256 ITF
	• 512 MAXICODE
	• 1024 PDF_417
	• 2048 QR_CODE
	<ul> <li>4096 RSS_14</li> </ul>
	• 8192 RSS_EXPANDED
-barcodeformat2	This is an integer representing the barcode type you want kingfisher
	to look for when finding %VALUE2%.
-barcodeformat3	This is an integer representing the barcode type you want kingfisher
	to look for when finding %VALUE3%.
-barcodeformat4	This is an integer representing the barcode type you want kingfisher
	to look for when finding %VALUE4%.
-usescript (bool)	If set to true, it tells kingfisher to use the script to generate filenames
	instead of the Name Template.
-outputScript	A string value representing the Script used to generate filenames.
-temp	A string value representing the Kingfisher temp location.

### 8.2.1 Split Operation Parameters

When performing a split operation, you can use the parameters below to further describe the job.

Parameter	Notes	
-retainbookmark(bool)	Retain Bookmarks	
-splitmode (int)	<ul> <li>Tells Kingfisher to how to split the PDF document, options are:</li> <li>0 (Text/Barcode on first page)</li> <li>1 (Text/Barcode on last page)</li> <li>2 (Remove Text/Barcode page)</li> </ul>	
-r range{,range}	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 : 1 Specifies a single page	

	1-3	Specifies a range of pages	
	10-20even	Specifies even pages within the range	
	10-20odd	Specifies odd pages within the range	
	20-lastpage	Note that <i>lastpage</i> can be used to represe	nt
		the last page in the document.	
-g grouppages	Apply the page range to each set of <i>grouppages</i> pages within the document. For example, if –r 2-4 and –e 5 is specified, 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.		
-k	Split by top-level bookmark		
-S	Split by single pages (The Default)		

### 8.2.2 Extract Job Parameters

Parameter	Notes
-mergeextract (bool)	Merge all the pages extracted
- mergeextractmatch(bool )	Merge pages that have the same "%VALUE1%"

### 8.2.3 Extract Text Job Parameters

Parameter	Notes	
-frompage (int)	The page that the text extraction starts from	
-topage (int)	The page where the text extraction ends	

### 8.2.4 Metadata Specification

To set the Metadata, append the following command to the command line arguments.

#### -1="Author=Tony Bush; Title=A Special Document"

The attributes that may be set are:

- Author
- Title
- Subject
- Keywords
- Creator

### 8.2.5 Security Specification

To set the Security, append the following command to the command line arguments.

-3="userpassword=Secret;ownerpassword=secret;permissions=allowprinting,allowmodifyannotations"

Parameter	Description		
Userpassword	A password that will be required to open the document.		
Ownerpassword	A password that will be required to change the document permissions.		
Permissions	A comma separated list of document permissions		
	Permission	Description	
	allowprinting	Allow high-quality printing	
	allowmodifycontents	Allow assembly (see below) and other document medications	
	allowcopy	Allow text and graphic copying and extraction	
	allowmodifyannotatio ns	Allow modification of annotations	
	allowfillin	Allow filling of form fields	
	allowscreenreaders	Allow extraction of text and graphics in support of accessibility.	
	allowassembly	Allow rotation, insertion or deletion of pages.	
	allowdegradedprinting	Allow low-quality printing	
	allowall	Allow all the permissions	
strength	Unless this is set to some other value (such as 40), the default of 128 is used.		

### 8.2.6 Document Options Specification

To set the Document Options, append the following command to the command line arguments.

-2 ="hidetoolbar,pagemodeuseoutlines"

Property	Description
pagelayoutsinglepage	Single Page (SinglePage)
pagelayoutonecolumn	Continuous(OneColumn)

pagelayouttwocolumnleft	ContinuousFacing [odd pages left] (TwoColumnLeft)
pagelayouttwocolumnright	ContinuousFacing [odd pages right] (TwoColumnRight)
pagemodeusenone	No Thumbnails or Bookmarks Visible (UseNone)
pagemodeuseoutlines	Bookmarks Visible (Use Outlines)
pagemodeusethumbs	Thumbnails Visible (UseThumbs)
pagemodefullscreen	Full Screen Mode (FullScreen)
hidetoolbar	The viewer's tool bar will be hidden
hidemenubar	The viewer's menu bar will be hidden
hidewindowui	The viewer's UI elements (scrollbars etc) will be hidden
fitwindow	The viewer will resize the document's window to fit the
	size of the first displayed page.
centerwindow	The document window will be positioned in the center of
	the screen.
nonfullscreenpagemodeusenone	On exiting full screen mode, neither outline nor
	thumbnail images are shown.
nonfullscreenpagemodeusenoutline	On exiting full screen mode outlines are visible.
nonfullscreenpagemodeusethumbs	On exiting full screen mode thumbnails are visible.

### 8.3 Command Line Examples

This section will show you a few command line samples.

### 8.3.1 Command line to split all PDFs in a folder tree by Barcode

Kingfisher.Commandline.exe -createfolders -debug -condition=barcode -outputnamemask=%VALUE1% -language 0 -logfile="...\logfiles\%TIMESTAMP%.txt" -operation splitbarcode -SplitMode=0 t="...\Samples\Output" -ubc1 -SplitMode=0 -source="...\Samples\Input" -SourceType=tree sourcefolder="...\Samples\Input"

### 8.3.2 Command line to extract text from all the PDF files in a folder

Kingfisher.Commandline.exe -createfolders -debug -condition=barcode -outputnamemask=%Filename% -language 0 -logfile="...\logfiles\%TIMESTAMP%.txt" -operation extracttext -SplitMode=0 t="...\Samples\Output" -SplitMode=0 -source="...\Samples\Input" -SourceType=folder sourcefolder="...\Samples\Input"

### 8.3.3 Command line to extract pages from a PDF file based on text match

Kingfisher.Commandline.exe -createfolders -condition=text outputnamemask=%VALUE1%\\%FILENAME%\_%VALUE2% 0 -logfile="..\logfiles\%TIMESTAMP%.txt" operation extract -t="..\Samples\Output" -z1=9.97:10.96:99.56:20.74 -z2=535.75:77.33:152.44:19.7 tp1=6 -tp2=6 -SourceType=file -source="..\Samples\Input\postcode.pdf" sourcefolder="..\Samples\Input"

### 8.3.4 Command line to rename a PDF file Based on text match

Kingfisher.Commandline.exe -createfolders -condition=text outputnamemask=%VALUE1%\\%FILENAME%\_%VALUE2% 0 -logfile="..\logfiles\%TIMESTAMP%.txt" operation renametext -t="..\Samples\Output " -z1=9.97:10.96:99.56:20.74 z2=535.75:77.33:152.44:19.7 -tp1=6 -tp2=6 -SourceType=file -source="..\Samples\Input\postcode.pdf" -sourcefolder="..\Samples\Input"

# 8.3.5 Command line to rename all the PDF files in a folder tree to the value of the first barcode encountered in the file.

Kingfisher.Commandline.exe -createfolders -outputnamemask=%VALUE1% logfile="..\logfiles\%TIMESTAMP%.txt" -operation renamebarcode -t="..\Samples\Output" -SourceType=tree -source="..\Samples\Input" -sourcefolder="..\Samples\Input"

# 8.3.6 Command line to extract all the PDF pages with barcodes in a PDF file with debug information

Kingfisher.Commandline.exe -createfolders -debug -condition=barcode -outputnamemask=%VALUE1% -language 0 -logfile="...\logfiles\%TIMESTAMP%.txt" -operation extractbarcode -SplitMode=0 t="...\Samples\Output" -ubc1 -source="...\Samples\Input\barcode.pdf" -SourceType=file sourcefolder="...\Samples\Input"

### 8.3.7 Command line to split all the PDF files in a folder by barcode

Kingfisher.Commandline.exe -createfolders -condition=text -outputnamemask=%VALUE1% logfile="...\logfiles\%TIMESTAMP%.txt" -operation splitbarcode -t="...\Samples\Output" -SourceType=folder -source="...\Samples\Input\barcode files" -sourcefolder="...\Samples\Input\barcode files"

### 8.3.8 Command line to rename a PDF file based on Text and barcode

Kingfisher.Commandline.exe -createfolders -debug -outputnamemask=%value1%\_%value2% -language 0 -logfile="...\logfiles\%TIMESTAMP%.txt" -operation renamemix -t="...\Samples\Output" -ubc2 - z1=2.71:5.77:117.19:44.59 -z2=337.68:15.1:358.81:52.89 -tp1=6 - source="...\Samples\Input\postcode.pdf" -SourceType=file -sourcefolder="...\Samples\Input"

# 8.3.9 Command line to split a PDF file by top level bookmarks and rename the PDF files to the bookmark titles

Kingfisher.Commandline.exe -createfolders -debug -condition=barcode -outputnamemask=%bookmark% -language 0 -logfile="..\logfiles\%TIMESTAMP%.txt" -operation splitrange -k -t="..\Samples\Output" -SplitMode=0 -source="..\Samples\Input\RefGuide.pdf" -SourceType=file sourcefolder="..\Samples\Input"

### 8.3.10 Command line to split a PDF file by a change in the text in a on the page.

Kingfisher.Commandline.exe -createfolders -debug -outputnamemask=%value1% -language 0 logfile="...\logfiles\%TIMESTAMP%.txt" -operation splitchange -tf1="US [0-9],[0-9]{3},[0-9]{3} [A-Z][0-9]" -z1=559.07:84.59:149.33:32.15 -t="...\Samples\Output" -ocr -SplitMode=0 source="...\Samples\Input\Patents.pdf" -SourceType=file -sourcefolder="...\Samples\Input"

# **9** 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.	

## **10 SUPPORT**

Any problems running this application should give evident error messages, but if you are unable to resolve a problem please contact <a href="mailto:support@aquaforest.com">support@aquaforest.com</a>

### **11 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	<u>Homepage</u>