DocuFi Knowledge Base
Getting Started Guide – ImageRamp 8

603-685-4033
www.docufi.com
info@docufi.com

Welcome to the ImageRamp Batch Getting Started Guide.

This guide will walk you through the basic operations of the ImageRamp Batch product with samples configured for several folder processing scenarios. You can start ImageRamp Batch by selecting the desktop icon or by selecting to launch it at the end of the installation process.

We recommend you use the Process button after selecting a document type (Profile). Each profile is pre-configured to perform specific tasks and loads from a dedicated folder. The watch button will remove the original files for future processing.

All documents used in this sample are found in the installation folder and the \IRampSampleFiles directory. You can run the executable ReinstallDemoFiles.msi to restore the source files after running watch.

The examples include the following

AutoRotateOCR – is an example of performing AutoRotation and OCR creation. applied to Social Security Numbers, Credit Card Numbers and other configured items. A folder is created for each day and the files are time stamped.

BarcodeSplitting – is an example of reading 1 D Barcodes from a 3-page document and using the first barcode to name a file and the second barcode to create folders.

Bookmarks – is an example of reading 1 D Barcodes from a 33 page document (CharlesDickins.pdf) and uses the barcodes to create bookmarks for easy file navigation. No file splitting is performed.

MedicalInvoices – shows medical invoices where 2 OCR zone regions are extracted. One for naming the file and one for creating folders. This includes a file that is rejected to the QA bin due to a pen marker through the number. Use the QA tool to review rejected documents.

PrivacyAuditRedact – is an example of the privacy redaction capabilities where several documents have privacy blackouts applied to Social Security Numbers, Credit Card Numbers, Bank ID"s and other sensitive text. Use the home page and Privacy item to review the results.

ZoneInvoices – is another example of reading zones and using the results for file and folder naming purposes.
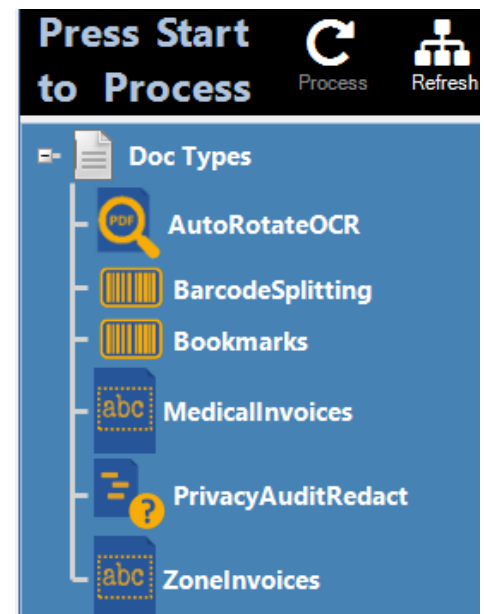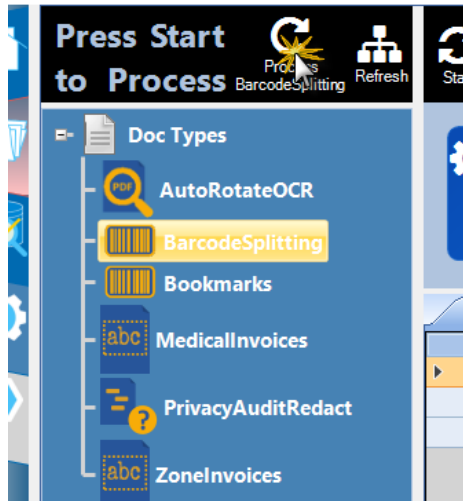
NOTE: Additional samples can be found by changing to a different configuration database. Select the drop down to change to a specific sample set including Barcode Mining, Zone OCR Mining, and Privacy Redaction and Auditing.
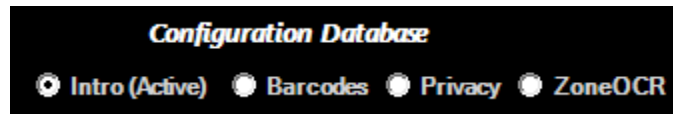

Getting Started

To get started with any of the provided demo document types, select (left click) the desired type of document (Doc Type) from the Doc Types tree. Then press the "Process" button. This will perform the tasks of the selected Doc Type and leave the file in its original folder. If you select Watch, all Doc Types configured will be processed and then removed from the source folder.

Once the processing begins, the logging display will provide you output of the data mining and processing tasks.
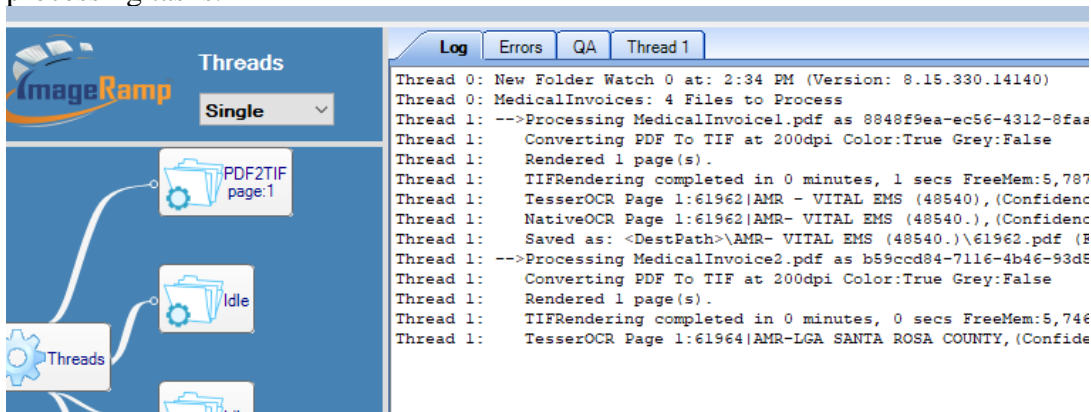
A database captures the relevant data extracted and records the steps performed.

# Configuring File Naming and Routing - BarcodeSplitting Example

Batch can split a multi-page file or document into multiple files and extract important content from the file or scan based on barcodes or mined text. This extracted information can be used for file splitting, and file naming and routing. How ImageRamp applies these functions is based on the **Barcode Extraction** and **File Naming** settings of the active Doc Type. In this example we will load a file of three invoices which we want to split based on the invoice number and place in a folder based on the invoice date. We are using the **"BarcodeSplitting"** doc type. *To follow along, make sure your selections match the selections in the images provided below.*
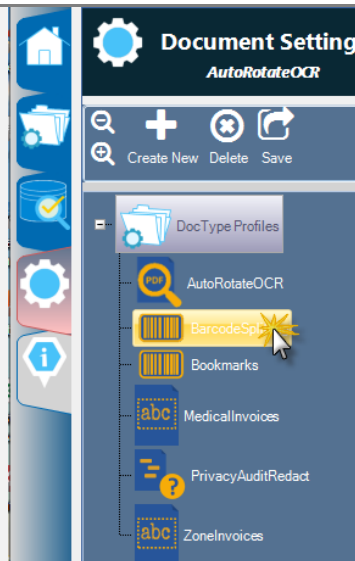
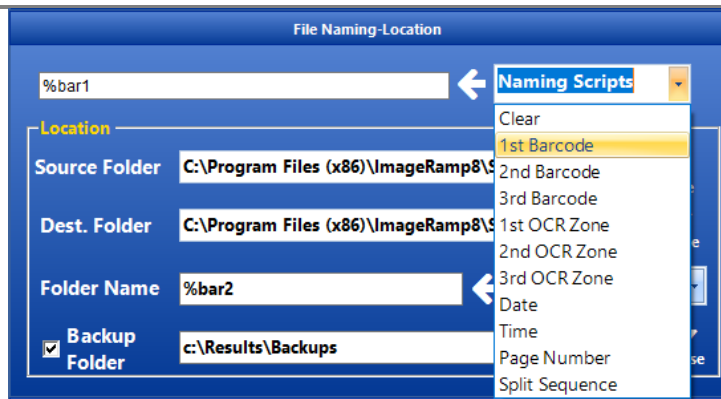Select the Settings tab in the left panel. This will display a tree of configured DocTypes.

Then click on the BarcodeSplitting profile or DocType. Each DocType is configured to a source and destination folder where specific workflow tasks are to be applied.
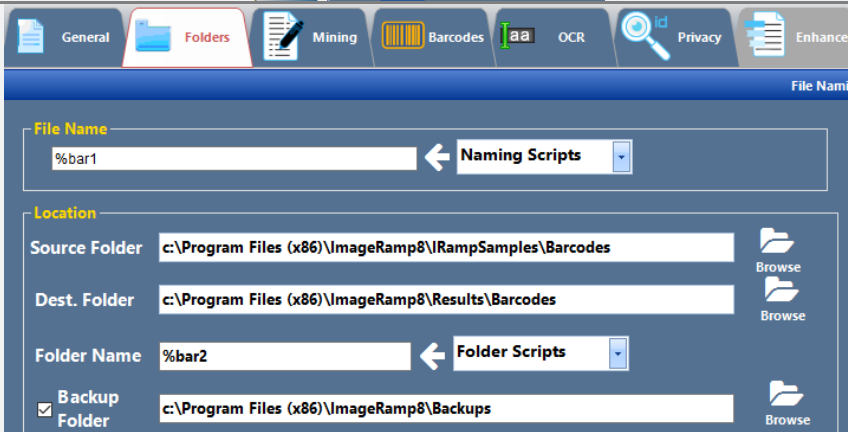
The default tab to load will identify the folders used for processing this profile. This includes a Source Folder, Destination Folder, and Backup Folder.

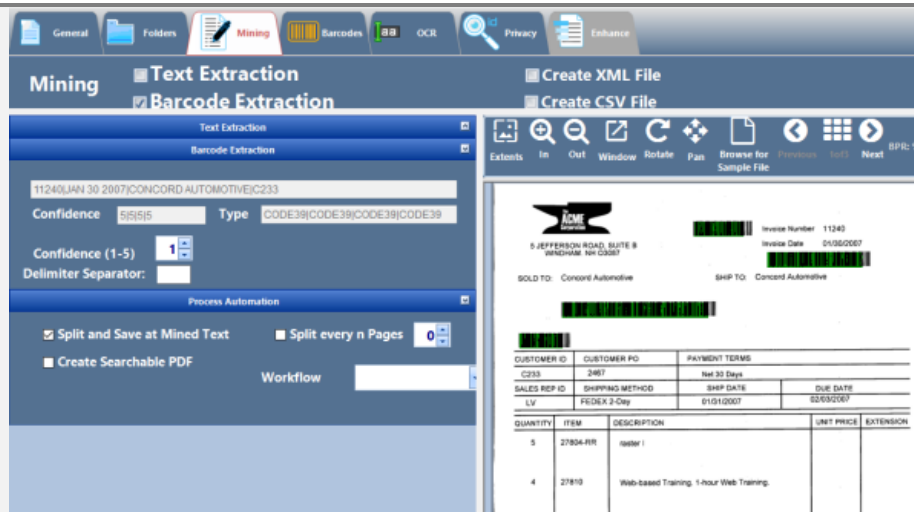You can also configure the script used to name the file or create folders.

File Naming-Location

%bar1 ← **Naming Scripts** ▼

Naming Scripts options
Clear
1st Barcode
2nd Barcode
3rd Barcode
1st OCR Zone
2nd OCR Zone
3rd OCR Zone
Date
Time
Page Number
Split Sequence

Location

Source Folder: C:\Program Files (x86)\ImageRamp8\\$

Dest. Folder: C:\Program Files (x86)\ImageRamp8\\$

Folder Name: %bar2 ←

☑ Backup Folder: c:\Results\Backups

A Few Words about Keywords

Keywords are available to name the file and subfolders using a date (%date), a barcode (%bar**n**) or time (%time) script along with any other characters. The pulldown menu provides a quick way to input the codes.

Keywords can be combined such as "%bar3-%time". This would name the file(s) as the values of the "third barcode-time".

Click here for more on keywords.

Mining

☐ Text Extraction   ☐ Create XML File
☑ Barcode Extraction   ☐ Create CSV File

Text Extraction

Barcode Extraction

11240|JAN 30 2007|CONCORD AUTOMOTIVE|C233

Confidence: 5|5|5|5   Type: CODE39|CODE39|CODE39|CODE39

Confidence (1-5): 1

Delimiter Separator:

Process Automation

☑ Split and Save at Mined Text   ☐ Split every n Pages: 0
☐ Create Searchable PDF
Workflow:

Select the Mining tab to review the extraction settings to use along with a Sample File. If one exists, the sample file will identify where barcodes (or ocr zones) are found in a color coded fashion. The confidence of barcodes is ranked 1-5 with 5 being the highest quality and green to 1 as the lowest quality and red.

The results are shown in the topmost text box with a pipe "|" delimiter separating barcodes. The confidence and type of barcode found is displayed as well.

In this example we found 4 Code39 barcodes all with a confidence of 5 out of 5.
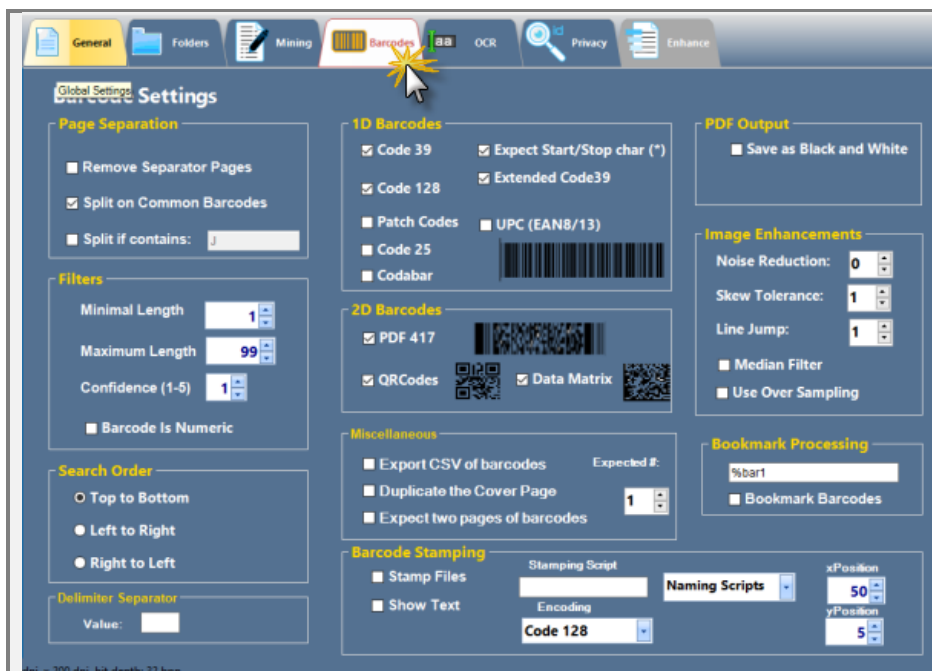
Click the **Barcodes** tab in the top most tree. This will load the barcode settings for the selected Document Type (**BarcodeSplitting**).

These settings identify the type of barcodes and actions to perform based on the barcodes. You can specify the types of **1D** or **2D Barcodes** to look for in this folder.

Filters can help isolate unwanted barcodes by length, confidence, or they must contain specific text. **Image Enhancement** tools allow you to fine-tune the image for accurate barcode reading.

In this example, Batch will split the files based on the located barcodes of the selected types with a search order of top to bottom. *Make your selections match those shown.*
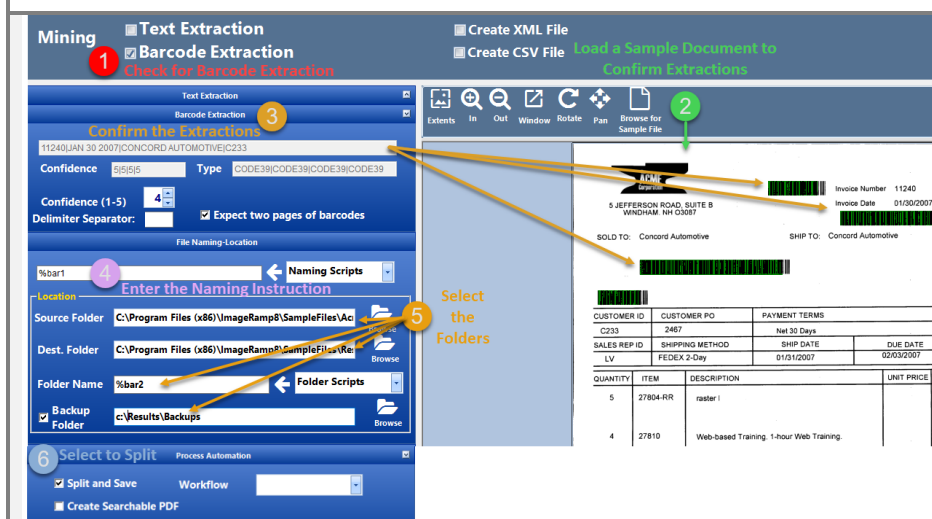

C > Local Disk (C:) > Program Files (x86) > ImageRamp8 > IRampSamples > Barcodes

ThreeBarcodedInvoices.pdf
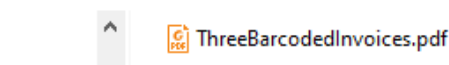
The sample file used in this example in located in

C:\Program Files (x86)\ImageRamp8\IRampSamples \Barcodes
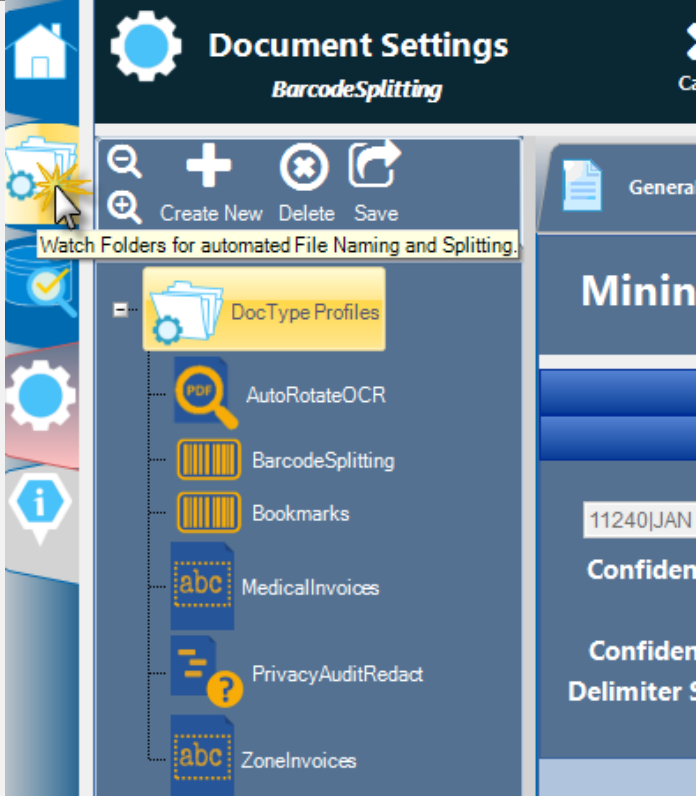
They are also located in the OriginalSamples for you to reuse when testing the application in watching mode.
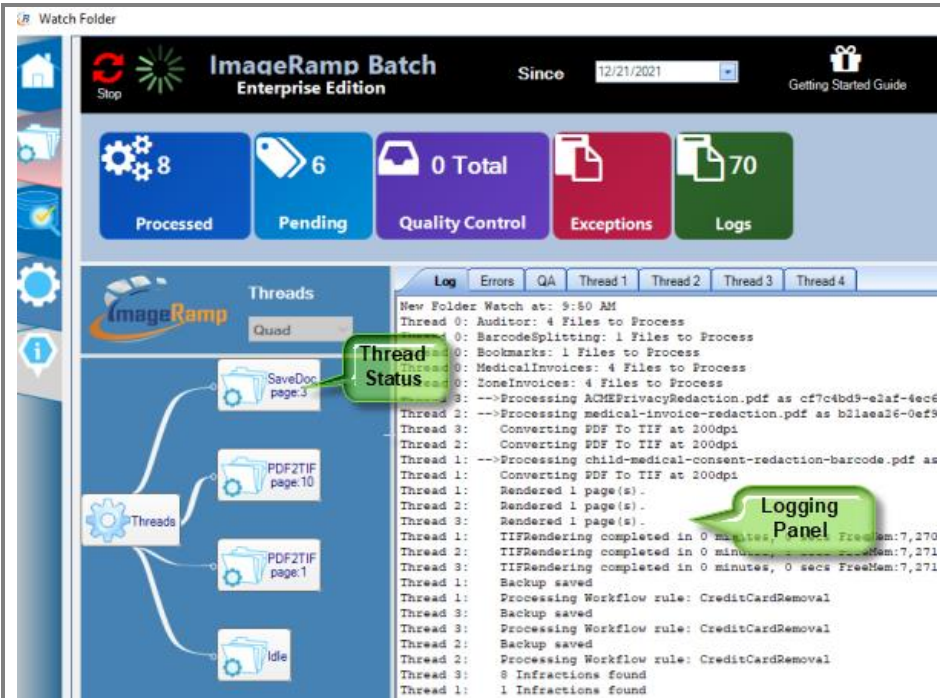

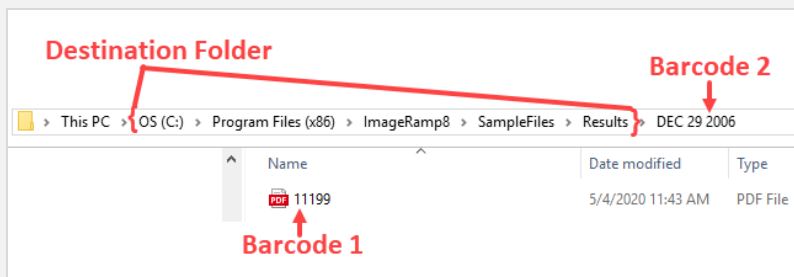Next, click on the **Mining** tab.

Note:

1. Check the checkbox to indicate you will be extracting barcode data.
2. Load a sample file to view and confirm the barcodes are found. They should be highlighted in green in the sample.
3. The extracted text, confidence score or barcode types are displayed on the left.
4. Enter the naming schema in the naming filed. Keyword codes can be used (ie %bar1). See the next entry for more on these codes.

	5. Name the **Source**, **Destination**, and **Backup** folders. The **Folder Name** item indicates if subfolders are to be created under the **Destination Folder** and how to name them. 6. Be sure to select to **Split and Save** option. Click the **Save** icon to save any changes in the doc type tree pane.
	Now we can process the files contained in the source folder path. Simply select the **Watch Folders** tab, click on the desired Doc Type and select **Process** to run that specific Doc Type. Processing a folder is different than Watching a folder. The source files will remain intact using this option. Depending on your licensing you can also choose to process the files as a single process, dual process or quad process. Using dual or quad processing allows your computer to perform multiple tasks at once, processing the files quicker. To run them as a folder watching application, select the **Start** button. This initiates folder "watching" of all folders defined in your doc Types. These folders will be watched and as new files are placed in the **Source Folder**, they are processed. Watching is stopped by clicking on the **Stop** icon or closing the application. For uninterrupted folder watching, see the manual to set up **Windows Services**.
	As files are processed, note the details of the processing are shown in the Logging panel. Each thread for multi-process operations, will display it's status in the Threads Panel (ie OCRing, Rendering PDF2TIF..) Each individual thread has it's own log or you can view the overall logging activity. You may also click on the Error and QA tabs here to see details about errors in processing or details about

files that have been place in the Quality Control Center. The tiles in the top of the right pane also give you information about the processing of the specified document type based on the date selection. Jump to Quality Control.
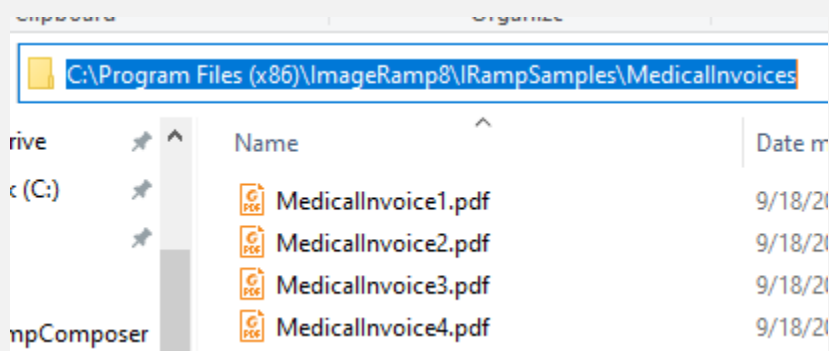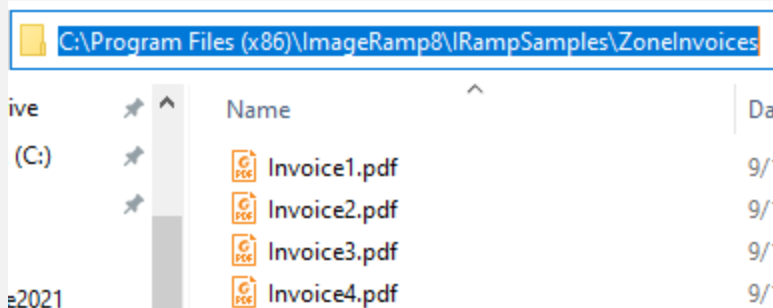

The net results are three subfolders named as the 2nd barcode (invoice date) in the destination folder named in the document type. Each folder has one file named for the 1st barcode (invoice date).

Note that all three files were processed. This brings us to what to do if batch finds an issue with the extraction. Jump to Quality Control.

File Naming and Routing – Extracted Text (Zone OCR) Example

Batch can extract text from within scanned documents and use the extracted text to name files and folders and route the file to a designated folder. Mining text works great for forms processing where the same information is located in the same zone. How ImageRamp applies these functions is based on the **File Naming** settings of the active Doc Type. In this example we will designate a source folder with three invoices which we want to name based on the invoice number of each invoice and place in a folder based on the invoice ship to name. We are using the **"ZoneInvoices"** and "**MedicalInvoices**" doc types. *To follow along, make sure your selections match the selections in the images provided below.*

C:\Program Files (x86)\ImageRamp8\IRampSamples\ZoneInvoices

Name	Da
Invoice1.pdf	9/1
Invoice2.pdf	9/1
Invoice3.pdf	9/1
Invoice4.pdf	9/1

C:\Program Files (x86)\ImageRamp8\IRampSamples\MedicalInvoices

| Name	Date m
MedicalInvoice1.pdf	9/18/2(
MedicalInvoice2.pdf	9/18/2(
MedicalInvoice3.pdf	9/18/2(
MedicalInvoice4.pdf	9/18/2(

All the sample files are found in the OriginalSamples folder. Whenever folder processing is run, the IRampSamples folders are emptied. To re-run the samples, copy them from the OriginalSamples folder into IrampSamples folder.
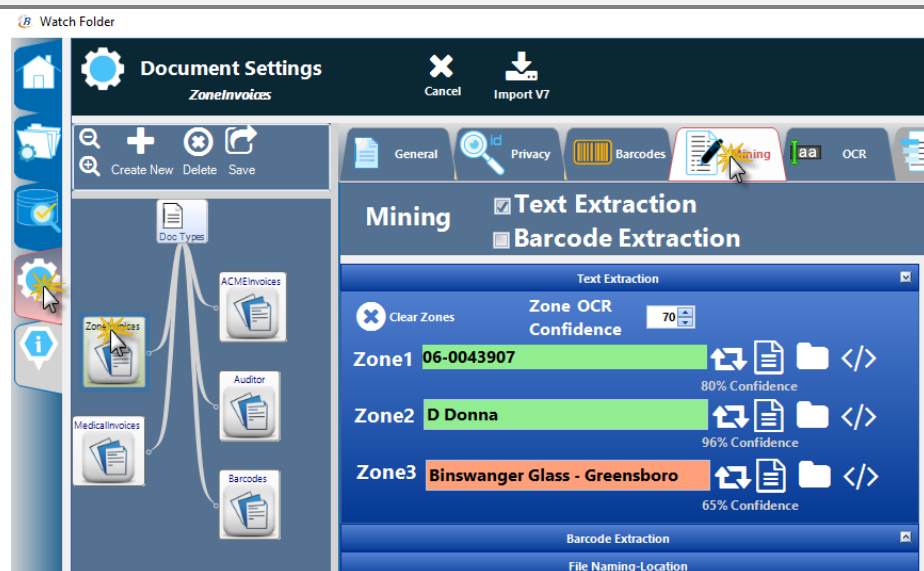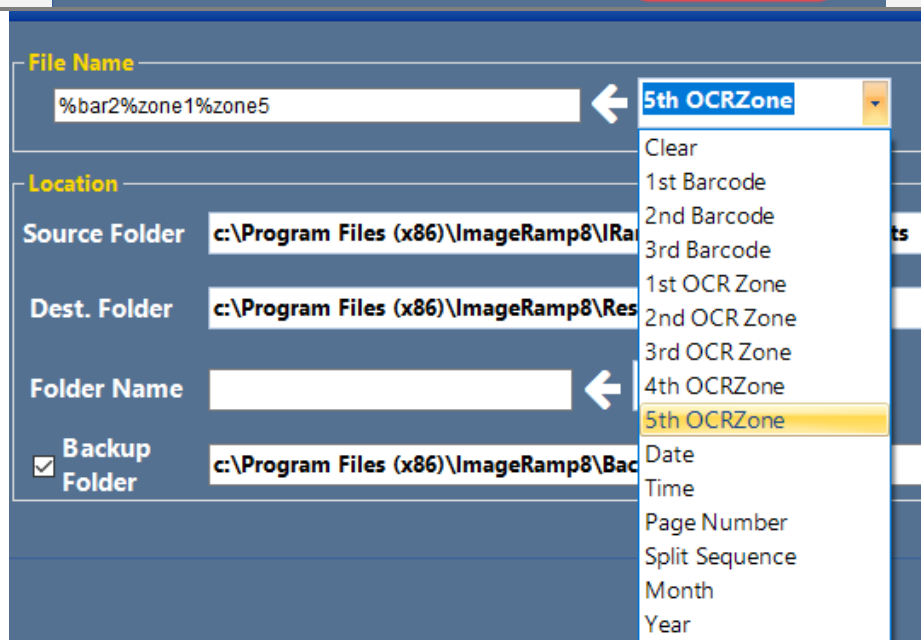

Click the **Settings** tab and select the **ZoneInvoices** doc type. This will load the Mining tab first along with any sample file.

Next, click on the **File Naming** settings tab.
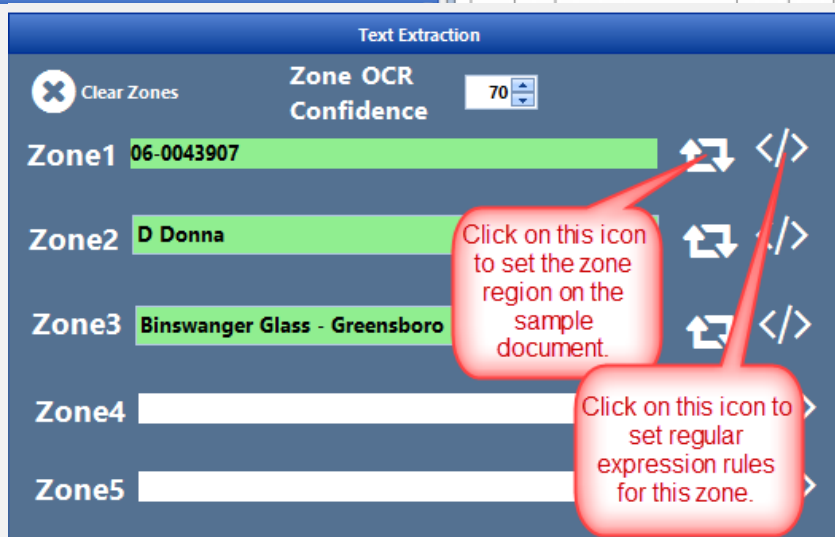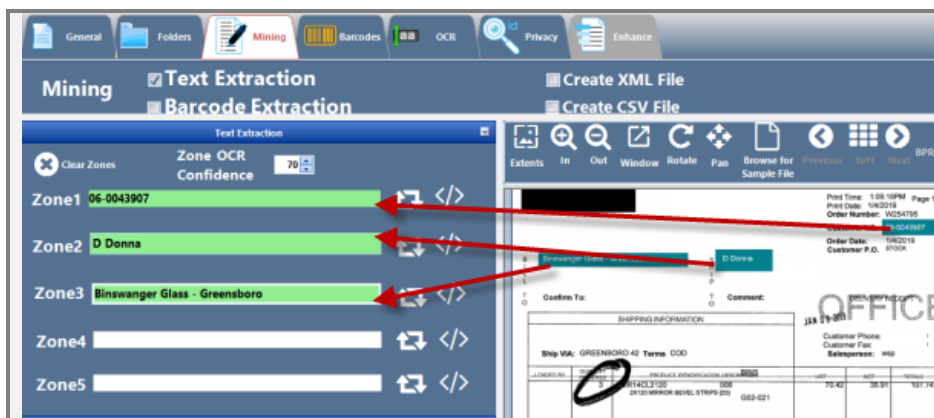
1. Check the checkbox to indicate you will be extracting text data.
2. Load a sample file to view and select the extraction zones. In our example three zones have been identified.
3. Use the select icons next to the zone fields to select the zones. The extracted text is displayed and you can set a confidence score from 20-99 for quality control purposes. See more on regex.
4. Enter the naming syntax in the naming fields. Keyword codes can used to select any OCR Zone along with other system variables.
5. Name the **Source**, **Destination**, **Subfolder**, and **Backup** folders. The **Folder Name** item indicates if subfolders are to be created under the **Destination Folder** and how to name them.

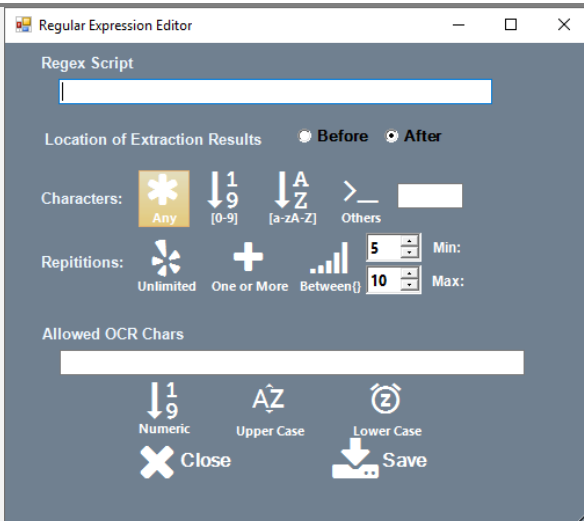Click the **Save** icon to save any changes in the doc type tree pane.

Click on this icon to set the zone region on the sample document.

Click on this icon to set regular expression rules for this zone.

A Few Words about Keywords

Keywords are available to name the file and subfolders using a date (%date), an OCR Zone (%field**n**) or time (%time) script along with any other characters. The pulldown menu provides a quick way to input the codes.

Keywords can be combined such as "%field3-%time". This would name the file(s) as the values of the "third OCR Zone-time".

Click here for more on keywords.

File Name

%bar2%zone1%zone5   5th OCRZone

Clear
1st Barcode
2nd Barcode
3rd Barcode
1st OCR Zone
2nd OCR Zone
3rd OCR Zone
4th OCRZone
5th OCRZone
Date
Time
Page Number
Split Sequence
Month
Year

Location

Source Folder   c:\Program Files (x86)\ImageRamp8\IRar...ts

Dest. Folder   c:\Program Files (x86)\ImageRamp8\Res...

Folder Name

☑ Backup Folder   c:\Program Files (x86)\ImageRamp8\Bac...

A Few Words about Regex

Clicking the regex icon **</>** brings up a regex editor where criteria describing the values of the extracted text can be defined. For example, patterns can be identified to only allow values that are numeric. This can be useful if all the invoices should be numeric.

Click here for more on regex.
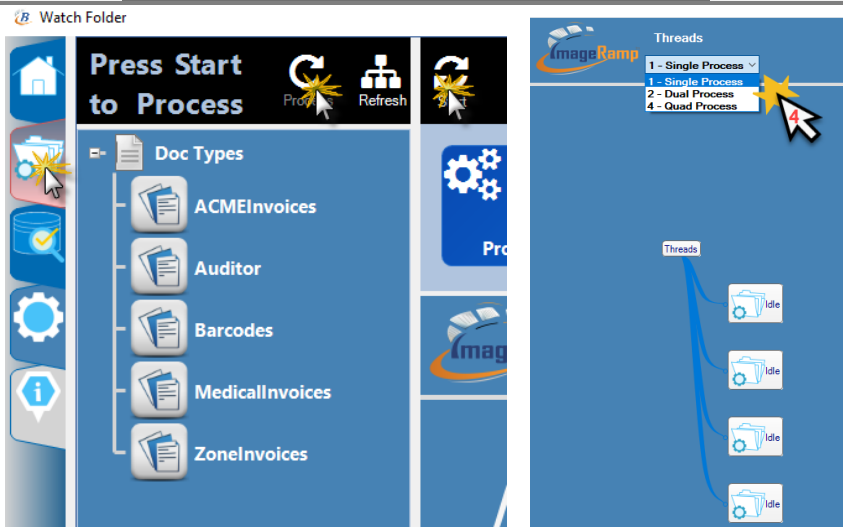
Now we can process the files contained in the source folder path. Simply select the **Watch Folders** tab, click on the **Unselect All** icon, check the box in the **MedicalInvoices** doc type from the document type tree in the left panel and start the processing by selecting the **Start** icon.

Depending on your licensing you can also choose to process the files as a single process, dual process or quad process. Choosing dual or quad processing allows your computer to perform multiple tasks at once, processing the files quicker.

Selecting the **Start** icon initiates folder "watching". The folder will be watched and as new files are placed in the **Source Folder**, they are processed. Watching is stopped by clicking on the **Stop** icon or closing the application. For uninterrupted folder watching, see the manual to set up **Windows Services**.

⚠️ Note: Once the **Start** icon has been clicked, Batch will process all files in the Source folder even if the **Stop** icon is then clicked. The **Stop** icon only stops the processing of any future files placed in the Source folder until the **Start** icon is clicked again.

```
Log    Errors    QA    Thread 1
New Folder Watch at: 1:23 PM
  Profile:PacificBio has 3 file(s).
Thread 1: -->Processing PacificBio1.pdf as f5e1dd3f-f23d-4c0a-ab29-6e42376e06bc
Thread 1:    Converting PDF To TIF
Thread 1:    Rendered 1 page(s).
Thread 1:    Page 1:61962|AMR- VITAL  EMS  (48540),(97|68),(0 degrees)
Thread 1:    Text Mining Below Confidence threshold 97|68
Thread 1: -->Processing PacificBio2.pdf as 762c11a1-02a1-40d3-9e82-6cbebb0b81eb
Thread 1:    Converting PDF To TIF
Thread 1:    Rendered 1 page(s).
Thread 1:    Page 1:61964|AMR-LGA SANTA ROSA COUNTY,(97|95),(0 degrees)
```

As files are processed, note the details of the processing are shown in the Log window in the lower right pane. You may also click on the Error and QA tabs here to see details about errors in processing or details about files that have been place in the **Quality Control Center**. The tiles in the top of the right pane also give you information about the processing of the specified document type based on the date selection. Jump to Quality Control.

The net results are two subfolders named as the 2nd zone (ship to) in the destination folder named in the document type. Each folder has one file named for the 1st zone (invoice no.).

Note there were three original files. One of them was not processed due to a quality issue. This brings us to what to do if batch finds an issue with the extraction. Jump to Quality Control.

Quality Control

When using barcode or text extraction for file naming, Batch offers confidence settings to ensure your extraction data is interpreted correctly. For barcodes, a confidence rating of 1-5 can be applied. With Zone OCR, the confidence rating can be set from 20-99.

For both ratings the higher the number, the more confident Batch must be to process the file. During processing, if a file's extraction data does not meet your selected confidence level, it is placed in the **Quality Control Center** for your review before it can be processed.

If we look back at the example above for extracting text via Zone OCR, there was one file that was not processed. Note that the log indicates the "Text Mining Below Confidence". The second zone only had a confidence of 68 which was below the setting we used of 70.


To process files that have been flagged for questionable extraction,

1. Click on the **Quality Control Center** icon
2. Click on the desired document in the left document tree to load it in the document viewing pane on the right.
3. Verify or edit the extraction information in the fields in the **File Naming** pane in the middle. Note the fields are color coded to indicate which field had questionable data. Green fields met the confidence level and those below the confidence level are indicated in red.
4. Once you have verified/edited the information in the middle pane, click **Release** to process it. Click **Reject** to not process the file, but add it to the backup folder. If the document should not have been in the Source folder, click **Remove** to delete it with our processing or placing it in the Backup folder.

Note, Barcode quality control functions similarly.
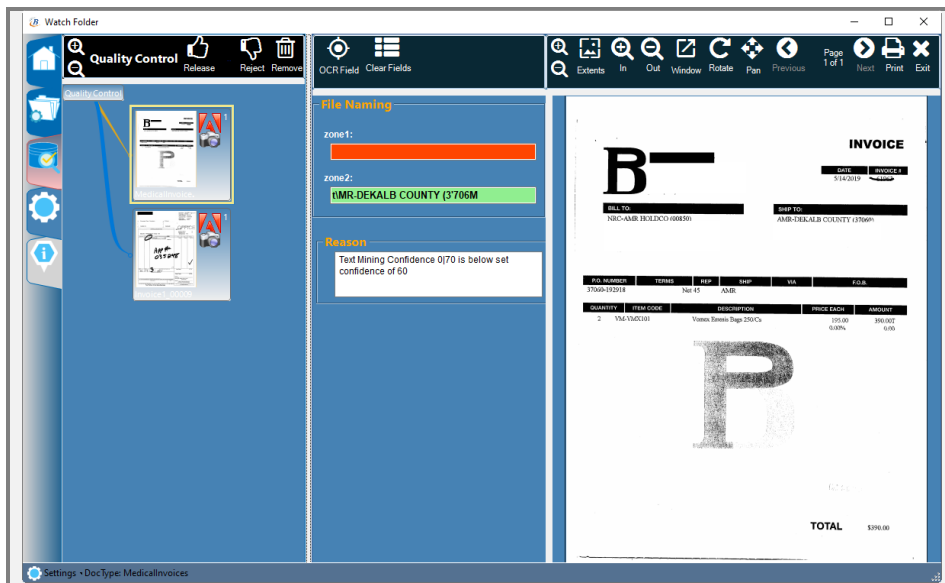
In our example the information is correct so the document can just be released. If it was not captured correctly, you can click in the field box for Zone 2 and enter the data. Alternatively, select the **OCR Field** icon which is a capture tool to drag and drop from the text in the document pane. Also, to assist is the **Clear Fields** icon to clear the fields before entering new data.

Setup a Live Web Demo

Like to see more? Call or email us for a personalized web demo. Contact DocuFi.