



PSIGEN software inc.
the **science** of document capture

QUICKstart

03

Modified: 12.5.19
Screenshots: 7.5.0.11



PSI**capture**

Classification

Quick Start Guide

Getting Started

Welcome to PSICapture!

This Quick Start Guide is designed to help you through the process of creating a Capture Profile that uses Classification to process multiple varieties of documents within the same batch. This guide includes sample images that will be used to configure and test the Capture Profile. For additional information on features not covered in this guide please refer to the PSIGEN wiki at help.psigen.com. PSIGEN offers detailed hands on training with the PSICapture software as well.

Make sure you have completed the **00 - Installation Quick Start Guide** and used it to get PSICapture installed, licensed and running on your computer before starting with this Guide. You will need to be running version 7.6.x or newer for this guide to be fully effective.



Requirements:

- Completed **00 - Installation Quick Start Guide**
- Ensure your computer meets the **min. requirements**
- Running version 7.6.x of PSICapture or newer
- PSICapture License with Classification module

What is Classification?

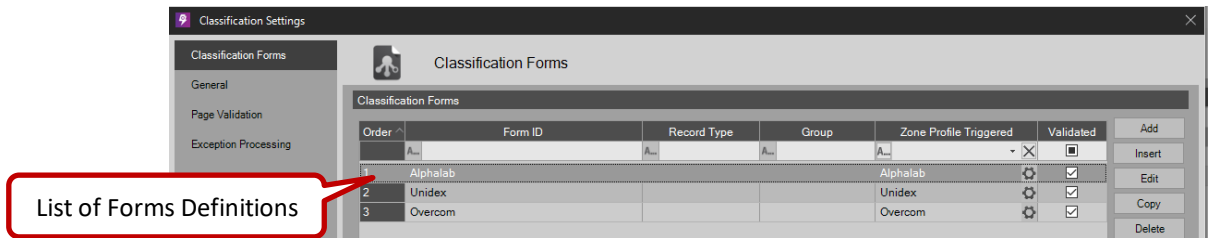
Classification is a PSICapture module that uses rule-based logic to distinguish between disparate varieties of documents. This allows you to process dissimilar documents using a single Capture Profile, rather than requiring you to design a separate Capture Profile for each.

☆ Note:
Each type of document being classified is referred to as a **Form ID**.

You can associate one or more Regular Expression rules with each **Form ID** that you need to classify. During the Classification workflow module, the OCR engine will read the text contained on the documents. Depending on the rule matches that are found, PSICapture will determine the proper Form ID under which to classify each document.

CLASSIFICATION QUICK START GUIDE

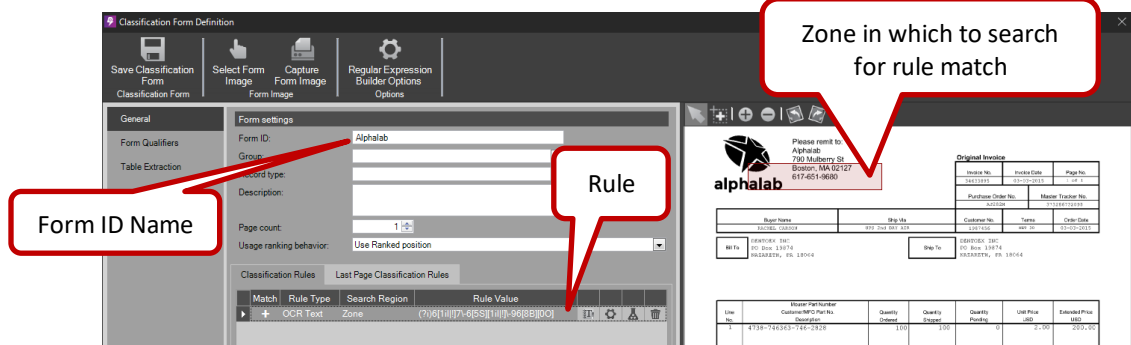
The rules associated with each Form ID will be configured within a separate Classification Form Definition.



☆ Note:

When using multiple rules, you can specify each rule as either positive (*must match*), or negative (*must NOT match*). You can also specify whether you require *all* rules to match in order to classify, or only one of them.

When writing rules, look for a phrase or pattern that appears uniquely and consistently on the documents that you are trying to classify. Firstly, you will add one or more rules that you expect to find on the **first** page of the document. You can choose to search for the rule match anywhere on the page, or within the header or footer. You can also choose to narrow the search region down to a specific zone.

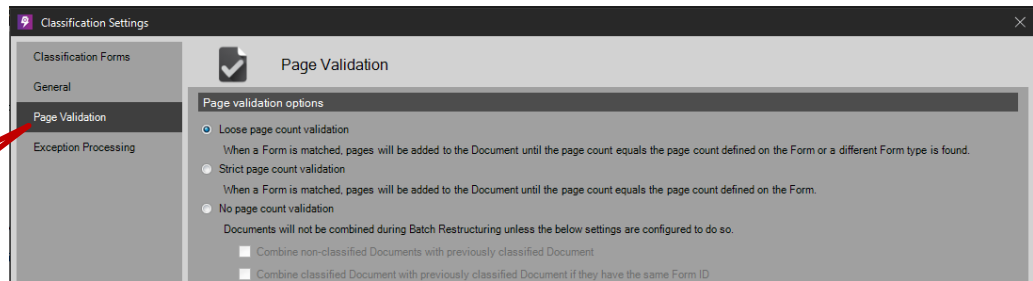


Classification can also be used for document separation. When using Classification for separation, you will initially separate the batch into single-page documents. Then, the Classification module will be used to merge those documents together as necessary.

If you expect all documents belonging to a given Form ID to contain a consistent number of pages, you can specify that number. Alternatively, if each document may consist of a variable number of pages, then you can specify one or more **Last Page Classification Rules**. When using Last Page Rules, PSImage will continue merging all pages into the same document until a match to the Last Page Rule(s) is found.

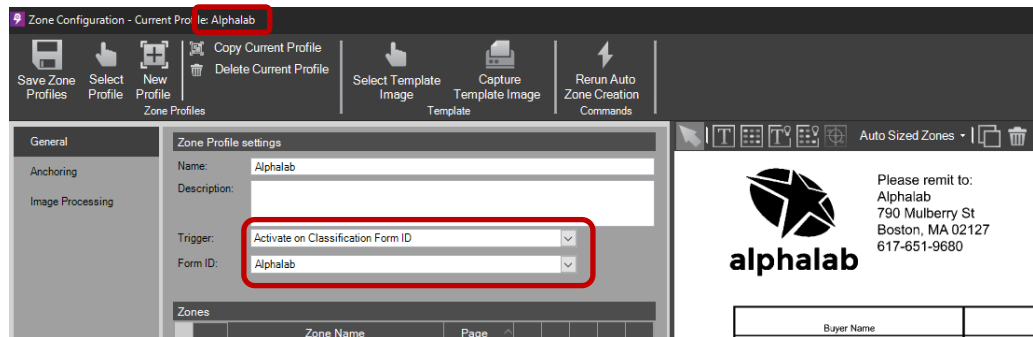


CLASSIFICATION QUICK START GUIDE

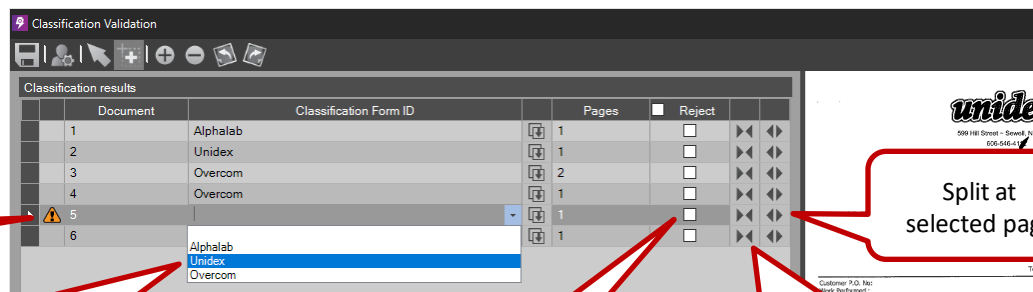


You can trigger different functions on each Form ID. For example, you can use separate Zone Profiles to look for the same index data in different places and formats on each type of document.

☆ Note:
Optionally, you can associate a Form ID with a specific **Record Type**, each of which may have a different set of associated index fields.



The Classification workflow module is typically added immediately after the Capture/Import module. Classification comes before the Index module, because each document must first be classified in order to know which set of Indexing rules to apply. If a document cannot be classified automatically, the user may be presented with a **Classification Validation** window, from which they can manually assign the proper Form ID and merge, split, or reject unclassified documents.

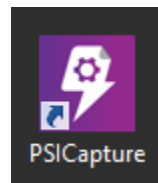


How will the Capture Profile function?

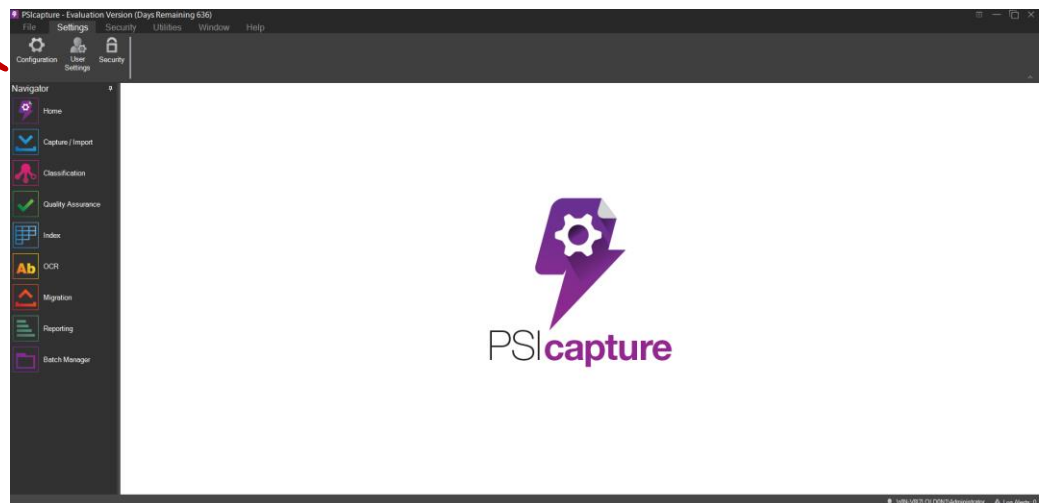
You will configure your Capture Profile to process invoices from 3 different vendors. You will be looking for the same 4 Index Data Fields on all the documents (Vendor Name, Invoice Number, Invoice Date and Invoice Total). However, the required data will be found in a different place on each vendor's invoice. In order to account for this, you will first set up Classification to determine which vendor (Form ID) each document belongs to. In the case of multi-page documents, you will use Last Page Rules to merge pages together. Then you will build 3 unique zone profiles (one per vendor) to extract the required data using Smart Zones. You will trigger each zone profile on its corresponding Form ID.

Creating your Capture Profile

First, you'll need to get a basic Capture Profile started. Open PSICapture.



Click to open
Configuration



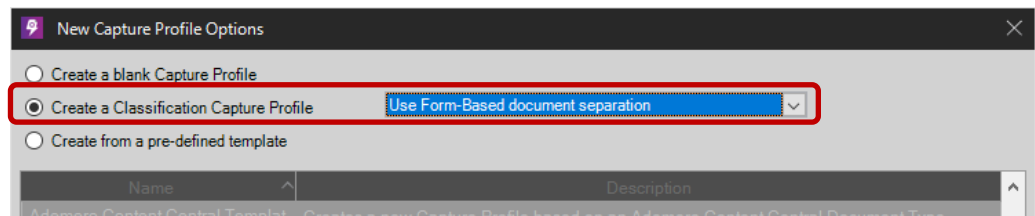
After PSICapture opens you will start in the Configuration module to create your Capture Profile.

CLASSIFICATION QUICK START GUIDE

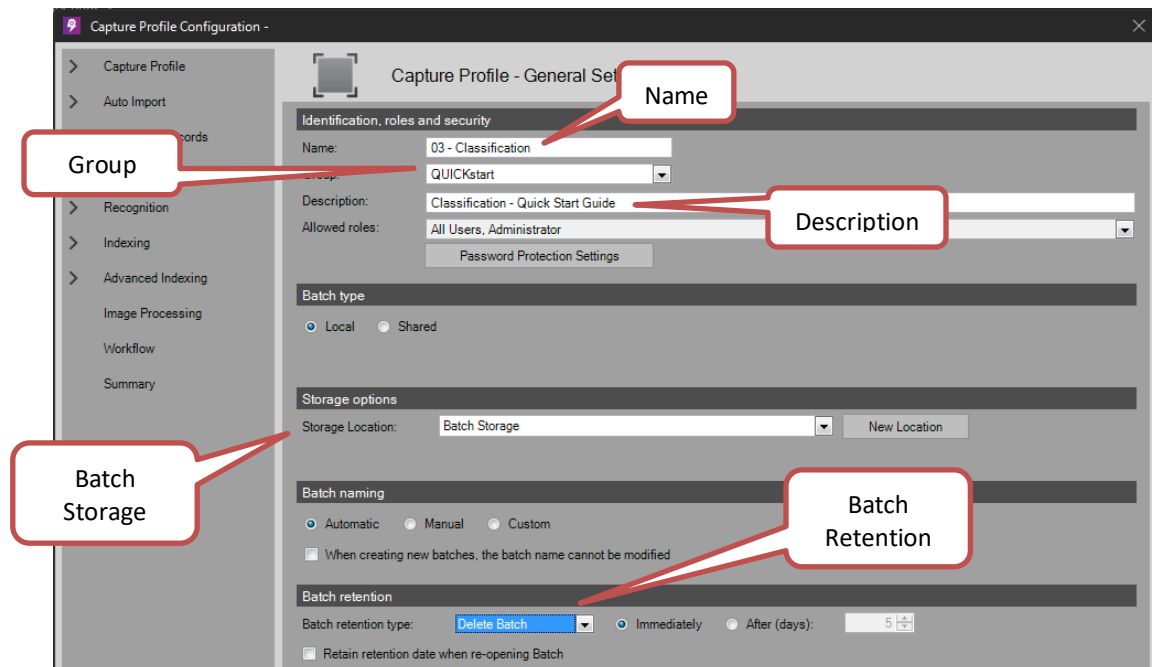
Once the Configuration module is open you can begin work on your Capture Profile. Please click the **Add** button on the right side of the window to proceed.



Change selection to **Create a Classification Capture Profile** and click **Create**.

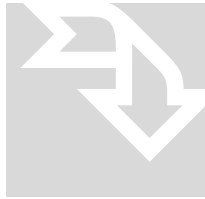


Capture Profile Tab



Give your new Capture Profile a name, group and description. Confirm your Batch Storage Location and set your Batch Retention.

Auto Import Tab



For the purpose of this Quick Start Guide you will not be making any changes to Auto Import or its sub tabs. Many of the features within PSIsapture Capture Profiles can be left at their default settings. For additional information about features available on this tab please refer to the PSIGEN [Help](#).

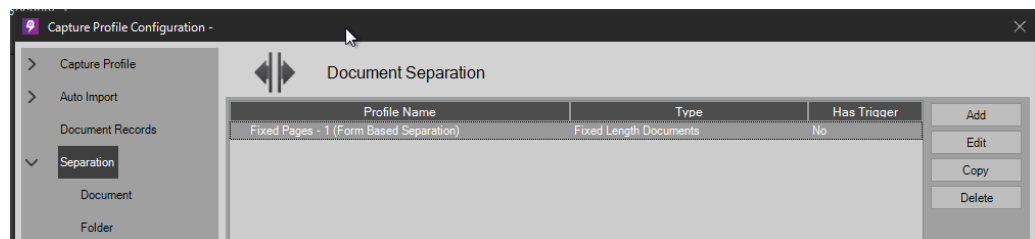
Document Records Tab

Again, leave all the default settings for Document Records, as they are not needed for this Capture Profile Configuration.

Separation Tab

When you first began creating this Capture Profile (see Page 5), you selected the option to **Create a Classification Capture Profile**. This change automatically configured the needed separation profile as Fixed Pages – 1 (Form Based Separation).

☆ Note:
Separation profiles are unique to each Capture Profile.



Even though some of the documents consist of multiple pages, those documents will not be merged together until the Classification workflow module. At that point, the Last Page Rules will be used to determine which documents need to be merged.

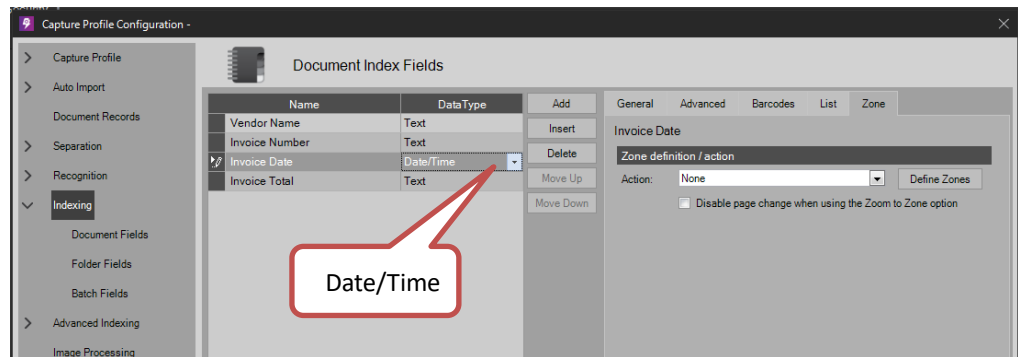
Recognition Tab

No barcode or other recognition engines are being used in this guide, so you can skip this tab.

Indexing Tab

Click the **Add** button in the middle of the window 4 times to create 4 blank index fields. Name those index fields as follows:

- Vendor Name
- Invoice Number
- Invoice Date
- Invoice Total



Change the **Data Type** dropdown to the right of the Invoice Date field to **Date/Time**.

Now you've got your index fields defined. You will return to the Indexing tab shortly to set up the indexing logic, but first you are going to move ahead to the Workflow tab so that you can create the Classification Form Definitions.

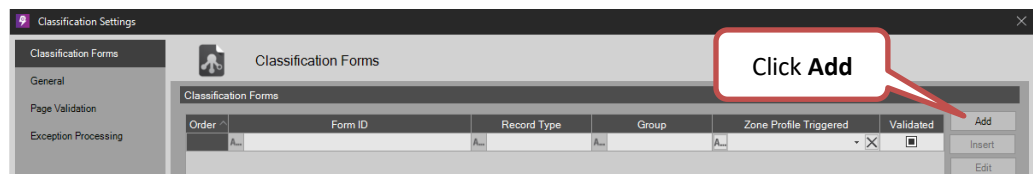
Workflow Tab

Proceed to the Workflow tab, where you will configure the Classification workflow module to recognize the different types of documents.

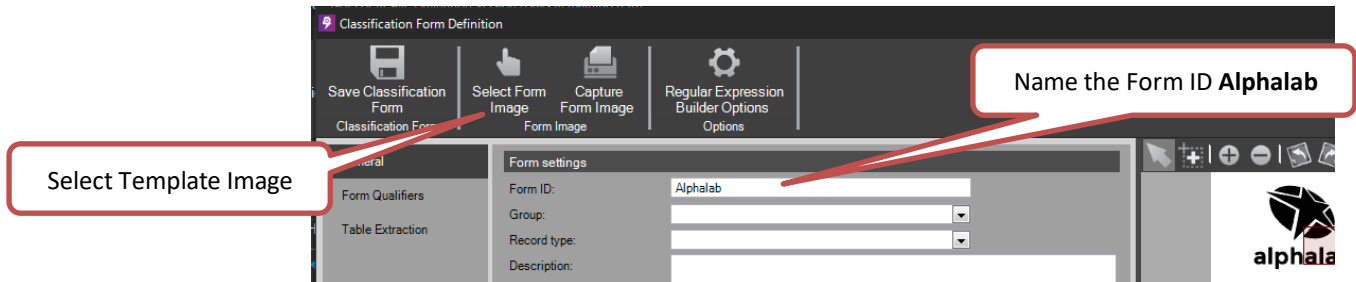
Select the **Classification** workflow module from the list on the *right* and click the **Configure** button to edit this module.

*The change you made when first creating the Capture Profile (see page 5), to **Create a Classification Capture Profile** automatically added the Classification Module to this workflow.*

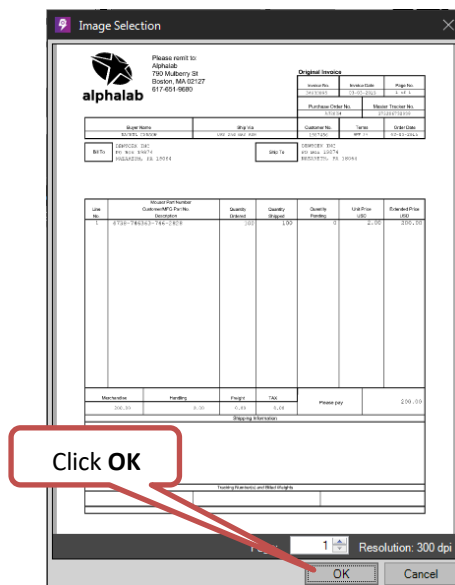
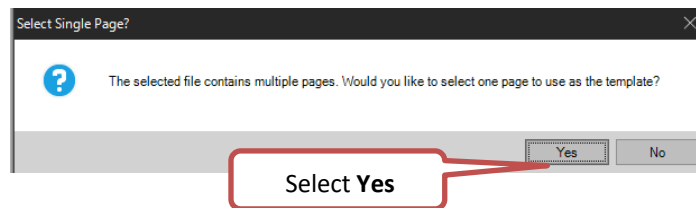
You will create one Form Definition at a time for each of the 3 vendors' invoices that you will be processing. From the **Classification Settings** window, click the **Add** button to create your first Form Definition.



Name your first Form ID **Alphalab**. Click the button to load in a sample of your first image type so that you can write a rule to identify it. Navigate to the folder where you saved your files and load page 1 of **03 - Classification Guide Samples.tif** as your template image.



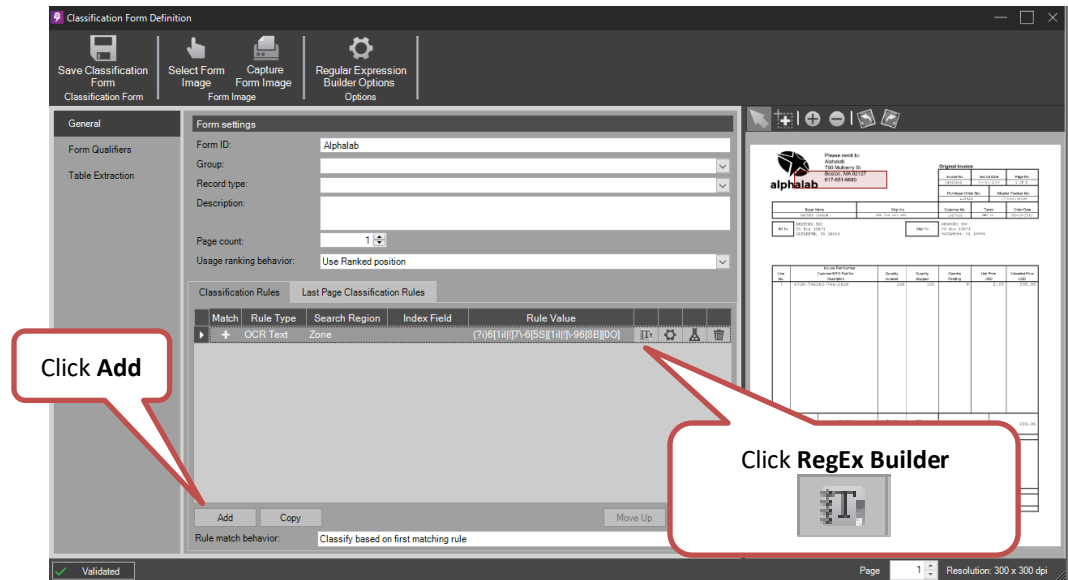
When prompted, select **Yes** to load only one page of the multi-page document as your template.



This first Form Definition will be used to classify the Alphalab invoice.

Because the Alphalab document is the first page of the PDF, you will choose to load Page 1 as your template image.

With the template image loaded, you will now add a rule to recognize it. Click **Add**, and then click the **Regular Expression Builder** button.



PSIcapture will run OCR against the sample image, and then display the image again in a separate pop-up window from which you will have the opportunity to use **point-and-click** to select any string of text to use as the Classification Rule.

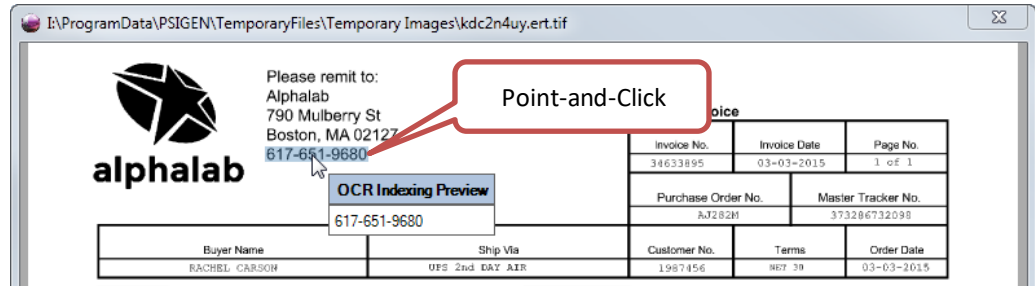
This literal text will be converted to a Regular Expression that will allow for the possibility that certain characters may be misread as other characters that are similar in appearance. For example, even if the OCR engine misreads the number zero as the letter “O”, the expression will be robust enough so that the rule will still match successfully.

☆ Note:
When building Capture Profiles to be used in production, it is considered a best practice to use *multiple* Classification Rules instead of just one. This will help avoid failures if the first rule fails to read correctly for any reason.

For this Capture Profile, you will only be creating one single rule for each Form Definition. You will choose a string of text that you expect to *always* appear on this vendor’s document, and *never* appear on any other vendor’s document. For this Capture Profile, choose the vendor’s **phone number** as the Classification Rule.

{This area is intentionally left blank, please continue to the next page}

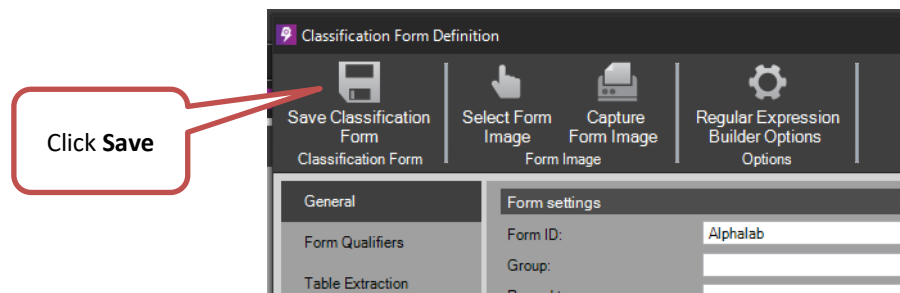
From the window that appears after clicking the RegEx Builder button, move the cursor over the **phone number**, and ensure that the value in the tooltip popup matches what you see on the page. If so, click the mouse once to create the rule.



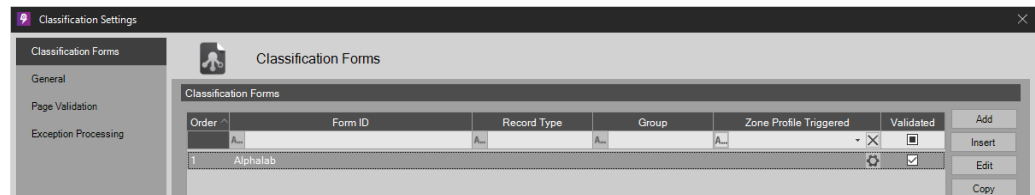
Notice that the Regular Expression Rule is now written for you. Test the Rule by clicking the beaker icon.

Match	Rule Type	Search Region	Index Field	Rule Value				
▶	+	OCR Text	Zone	(?)6[1il][7]-6[5S][1il][]-96[8B][00]				

Once the rule has been added, click **Save** to complete this Form Definition.



You have now completed the Form Definition for the Alphalab invoice.



You will repeat the same steps listed above to configure PSIsapture to classify the next vendor, **Unidex**.



1. From the Classification Form Definitions tab, click **Add**.
2. Enter the name of the Form ID, **Unidex**.
3. Click the **Select Template Image** button from the top-left.
4. Select the same **03 - Classification Guide Samples.pdf** that you loaded last time.
 - a. Select **Yes** to only load one page as your template.
 - b. Choose **Page 2**, which is the Unidex invoice.



5. Click **Add** near the lower-left to add a Classification Rule.



6. Click the RegEx Builder Tool button.
7. Point-and-click the Unidex phone number on the template pop-up.
8. Confirm the rule with the beaker icon.



9. Click the **Save** button to complete the Form Definition.

☆ Note:

You are choosing to load Page 5 as your template image for Overcom because it is a single-page invoice, meaning the same page is *both* the first *and* the last page of the document.

Now you will repeat the same steps for the invoice from the third vendor, **Overcom**. It is important that you load **Page 5** as the template image. There will be one additional step to be completed for the Overcom invoice before you save the Form Definition.

Like the other two vendors' invoices, some Overcom invoices consist of only one page. However, in other cases, depending on the number of line items that pertain to the document, Overcom invoices may span *multiple* pages. This means that, in order to recognize the last page of an Overcom invoice, you will need to include a **Last Page Classification Rule** as well.

☆ Note:

You did not need to set up Last Page rules for the other two vendors because it is known that all invoices belonging to those vendors consist of only one page.

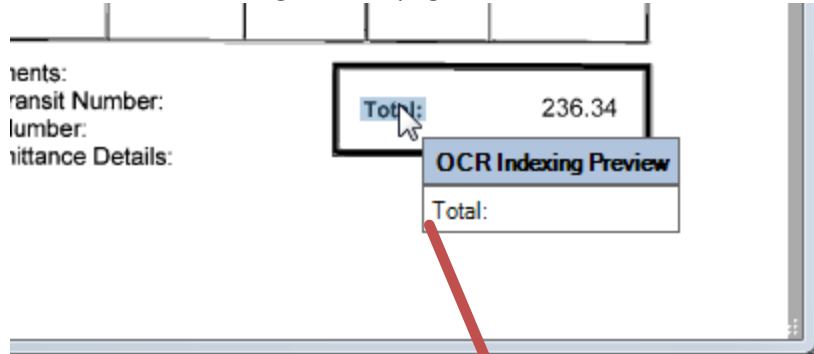
After you finish adding your first page rule to the Classification Rules tab on the left, select the **Last Page Classification Rules** tab immediately to the right.

When choosing a string of text to use for a Last Page Rule, pick something that *always* appears on the *last* page of documents that belong to the current Form ID, and *never* appears on any previous pages.

☆ Note:

It was important to load one of the Last pages of an Overcom invoice as your template image so that the word **Total** would be available for you to point-and-click.

If you evaluate the Overcom sample documents in the folder, you will notice that the word **Total** *always* appears on the last page, and *only* on the last page. Based on this observation, you will choose to use the word **Total** as the Last Page Rule. Click the **Add** button to create a new rule, then use the **RegEx Builder Tool** to point-and-click the word **Total** on the lower-right of the page.



Classification Rules		Last Page Classification Rules	
Match	Rule Type	Search Region	Index Field
+	OCR Text	Zone	(?i)T[oD]ta[li1!];?

Click **Save**, and all 3 of your Classification Form Definitions will now be complete. Select the **General** tab. You will make one change on this tab.

Classification Settings

Classification Forms

General

Page Validation

Exception Processing

Form processing options

Form processing order: Defined Order

Calculate usage ranking using: All usage

Page processing options

Run Classification on: All Pages of Document

Custom page list: Enter pages and page ranges, separated by commas (ex. 1, 3, 5-7, 9-END, LAST)

☒ When classifying multi-page Documents, stop processing Document as soon as Classification succeeds on a page

Group processing options

Groups to process: All groups

Classification Forms will be processed for the selected Groups only, in the group order specified.

Select Process Groups...

Indexing options

Index field to populate with Form ID: [Dropdown]

Index field to populate with Group Name: Document.Vendor Name

Index field to populate with Record Type: Document.Invoice Number

Index field to populate with Description: Document.Invoice Date

Document.Invoice Total

OCR text classification options

Header/Footer method: Lines

Header/Footer line count: 25

Header/Footer area: 4.00 inches

OCR text viewing options

☐ Allow OCR text used for matching to be viewed on Page Information dialog

☐ Allow OCR text used for matching to be viewed on Classification Validation dialog

Default options

Default rule match behavior: Classify based on first matching rule

Save Cancel

☆ Note:

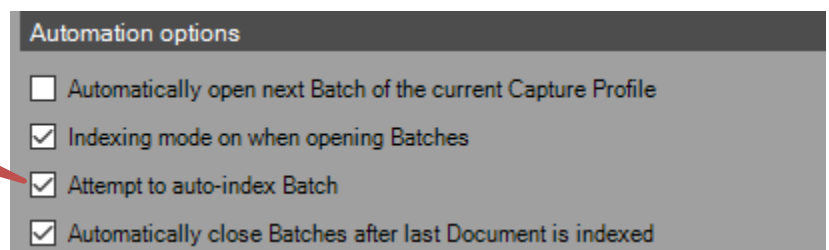
Because you are using this method to populate the field, you will not need to set up any other indexing logic from the Indexing tab to populate the Vendor Name index field.

In the **Indexing options** section, set the **Index Field to populate with Form ID** dropdown to be **Document.Vendor Name**. This will take the value of the Classification Form ID that is recognized for a given document (e.g. Alphalab) and use it to populate the Vendor Name index field for that document.

Now **Save** out of the Classification Settings.

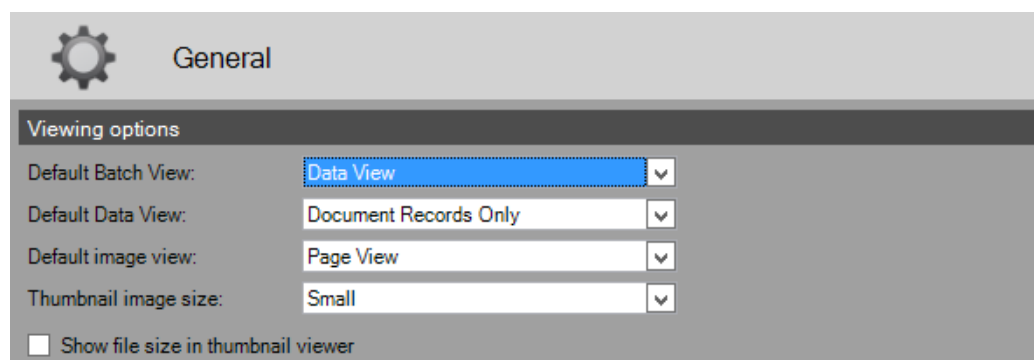
Before you set up Indexing to populate the other 3 fields, you will take this opportunity to finish building the settings in the Workflow tab by adding two additional steps. Select **Index** from the left and click **Add**. You will want to only make one change – check the **Attempt to Auto-Index Batch** checkbox from the left side in the Automation section. This will auto-index each document without requiring user interaction from the Index step. Click **Save**.

Check



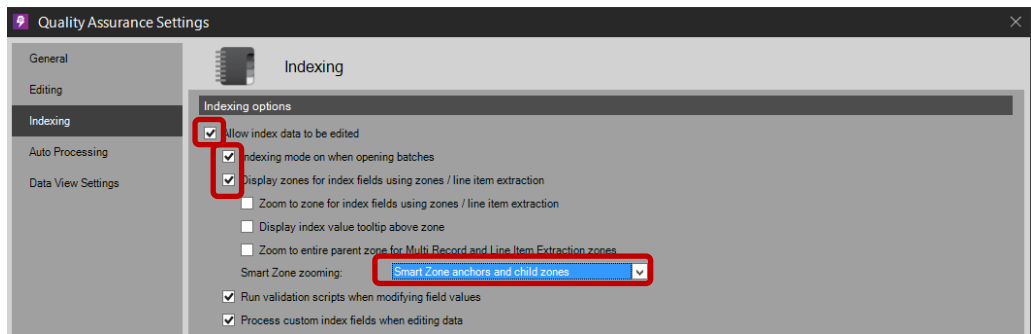
Automation options	
<input type="checkbox"/>	Automatically open next Batch of the current Capture Profile
<input checked="" type="checkbox"/>	Indexing mode on when opening Batches
<input checked="" type="checkbox"/>	Attempt to auto-index Batch
<input checked="" type="checkbox"/>	Automatically close Batches after last Document is indexed

Next, select **Quality Assurance** from the left and click **Add**. Change the Default Batch View from Tree View to **Data View**. This will cause the values that will have been extracted during Indexing to be displayed when you enter the Quality Assurance workflow module.

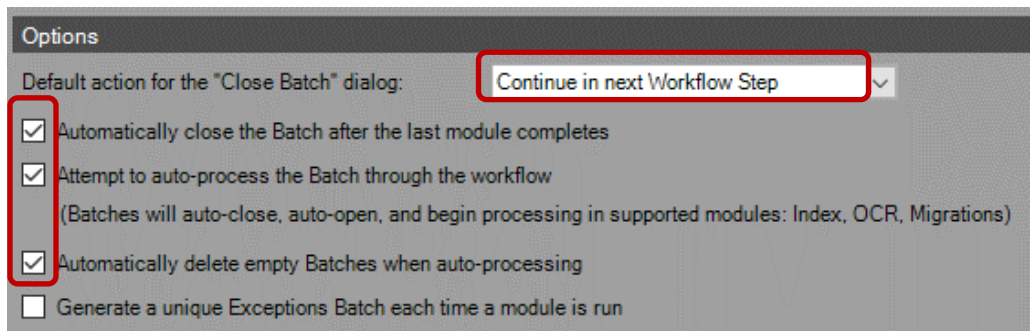


General	
Viewing options	
Default Batch View:	Data View
Default Data View:	Document Records Only
Default image view:	Page View
Thumbnail image size:	Small
<input type="checkbox"/>	Show file size in thumbnail viewer

On the Quality Assurance – Indexing sub tab, check the 3 checkboxes outlined in the below screenshot and choose **Smart Zone anchors and child zones** from the dropdown. This will display the exact location from which each index value was extracted as you verify the data from the Quality Assurance screen. Click **Save**.



Next, in the Workflow options (at the bottom of the tab), check the top 3 checkboxes and change the dropdown to match below. This will automate the workflow, so that the Capture Profile will process the documents without requiring user interaction after the Capture/Import module.

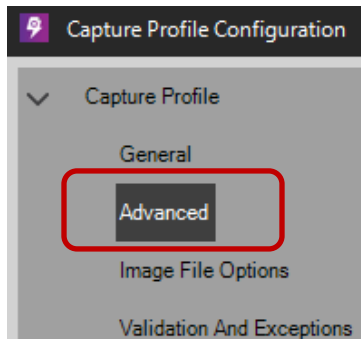


Your workflow will now consist of 4 modules.

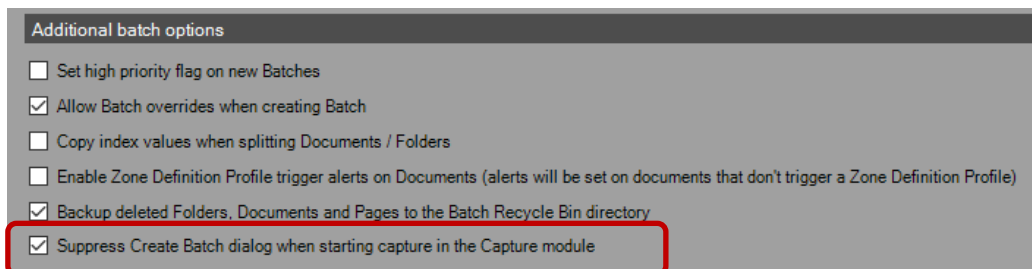
Workflow		
1	Capture / Import	
2	Classification	
3	Index	
4	Quality Assurance	

Suppress Create Batch Screen

When running a batch through PS!capture the Create Batch screen is displayed by default. We will not have a need to change any settings on that screen in this profile, so we are going to suppress that screen from being displayed. On the left-hand Navigator expand out Capture Profile and select the Advanced tab.



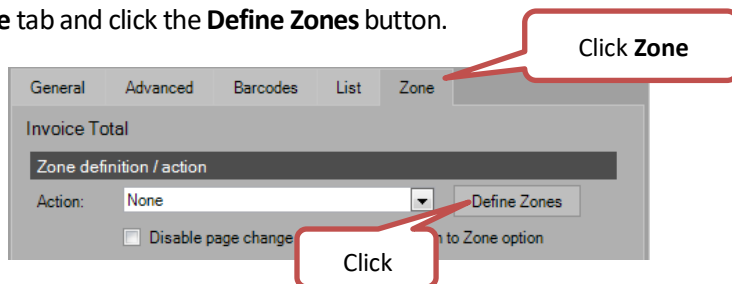
In the Additional batch options section, we want to check the 6th checkbox titled “Suppress Create Batch dialog when starting capture in the Capture module”



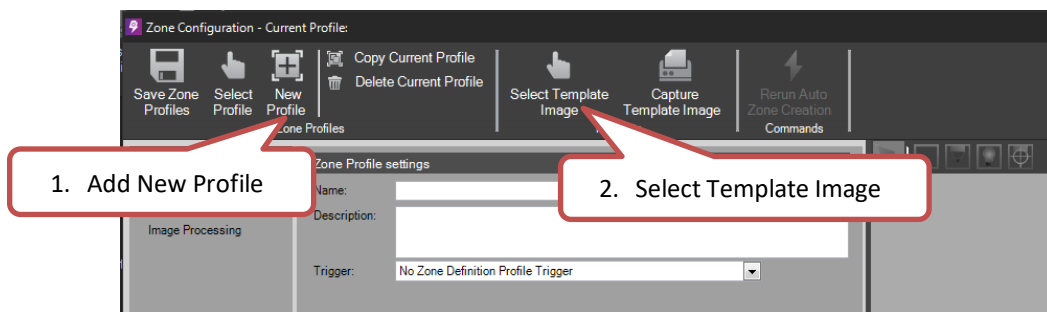
Now you will go back to the Indexing tab so that you can configure your Zone Profiles, which the Classification Forms you configured will use to extract the index data from each invoice.

Indexing tab (cont.)

Click on the **Zone** tab and click the **Define Zones** button.



You ***DO NOT*** want to create your zones on the ***Default*** profile. Instead, you will create 3 separate **Zone Profiles**, each of which is specifically customized to process documents that belong to only one of the three vendors. To add a new Zone Profile, click the **New Profile** button at the top of the Zone Configuration window.



Then click the toolbar icon to Select Template Image. Navigate to the folder where you saved your files and select **03 - Classification Guide Samples.tif** as your template image.

When prompted, select **Yes** to load only one page of the multi-page document as your template. The first Zone Profile you will configure will be for the Alphalab Invoice, which is **Page 1** of the sample file.

Now that your Zone Profile has been loaded, you will be entering some information pertaining to the profile in the **General** tab on the left side of the Zone Configuration window.

1. Give the Zone Profile a Name that matches the vendor and the Form ID. In this case, that name would be **Alphalab**.
2. In the Trigger dropdown, select **Activate on Classification Form ID**. That means the Zone Profile that you are about to create will *only* be used for documents that belong to a specific Form ID.
3. Then choose the Classification Form ID from the dropdown for **Alphalab**.

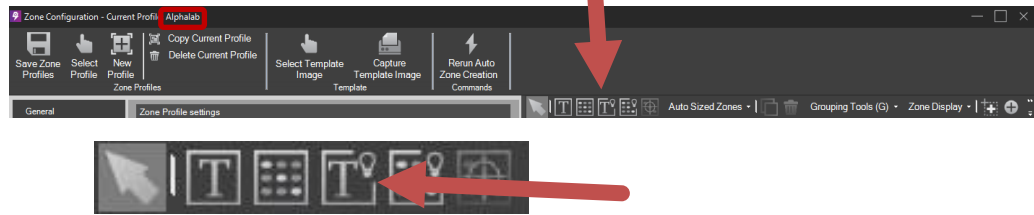
☆ Note:

You are not able to select the Classification Form ID from this list until the Form IDs have been set up. This is why you went to the Workflow tab to configure the Classification Form Definitions *before* configuring the Zone Profiles.

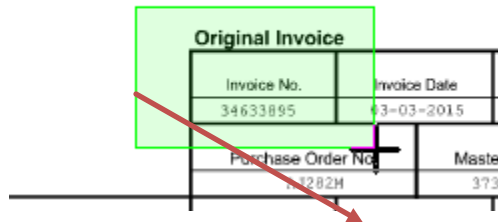
☆ Note:

Remember, you do not need a zone to extract a value to populate the Vendor Name field, because that field will have already been populated with the Classification Form ID (set up in the Classification Workflow module config).

Next you will create 3 **Smart Zones** to extract the required data. Click the Smart Zone tool to draw the zone.



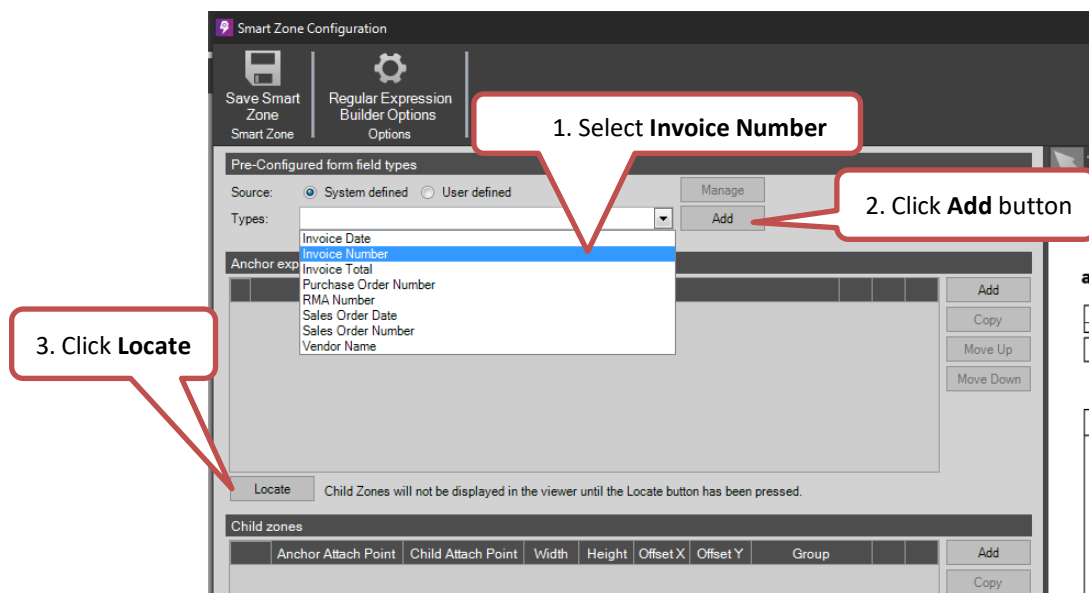
The first Smart Zone that you create will be designed to extract Invoice Number. Draw the zone around the label **Invoice No.** on the document. This label will serve as a good anchor for the Smart Zone, because you expect to find the Invoice Number itself immediately beneath this label on all Alphalab invoices.



☆ Note:

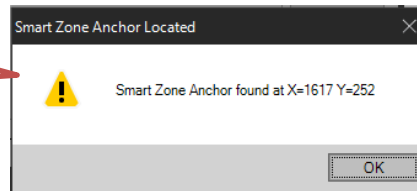
You are intentionally making the Smart Zone slightly oversized, in case the label shifts around the page somewhat on other documents from the same vendor. It is a best practice to center the **anchor** within the Smart Zone; the position of the Invoice Number itself is irrelevant at this point.

Upon releasing the mouse button, the Smart Zone Configuration window will appear. Select **Invoice Number** from the **Types** dropdown, then click **Add**, then **Locate**.



An alert notifies you that the anchor was found. Notice that the words **Invoice No.** are now highlighted in a green box. Based on your selection in the Form Field Types dropdown, PSICapture automatically selected this label as the anchor.

Alert notifies you that the anchor was located – click **OK**

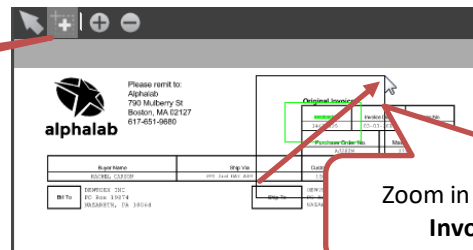


Original Invoice

invoice No.	
34633895	

In order to enhance precision while drawing your child zone, you will zoom in on the anchor using the **Selection Zoom** tool.

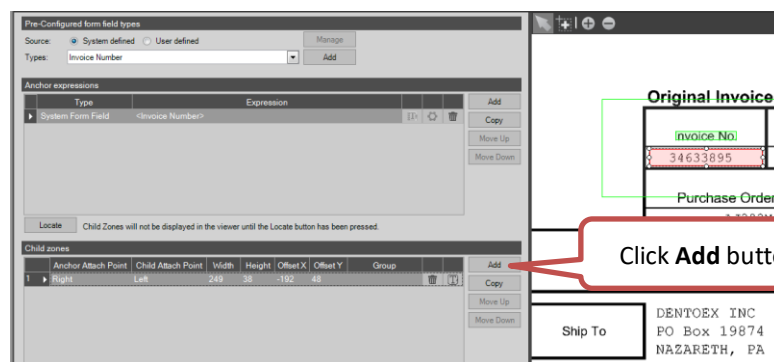
Click the **Selection Zoom** tool bar button



Zoom in around the region where the **Invoice No.** anchor is located

☆ Note:
It's a best practice to leave some extra space around the data in the child zone, but do not overlap any extraneous text.

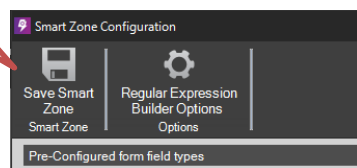
Next, you will add a child zone using the **Add** button in the lower Child Zones portion of the Smart Zone Configuration window. Manually resize the child zone so that it overlaps the entire Invoice Number itself, **34633895** in this case.



Click **Add** button

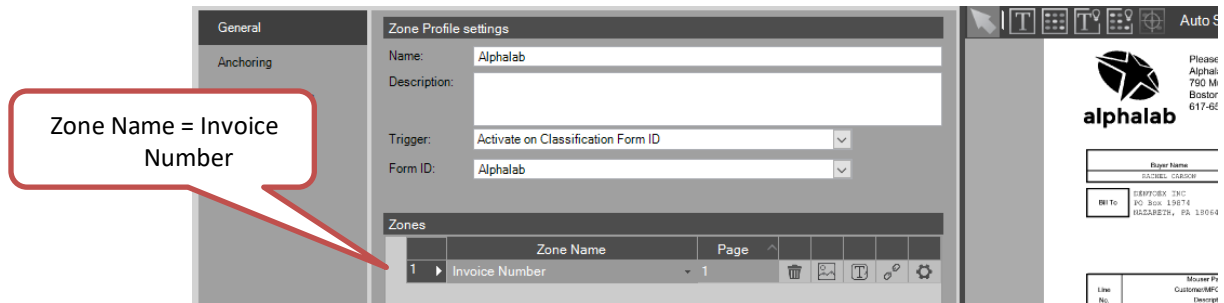
Click **Save** button

Now that you have finished building your first Smart Zone, click the Save Smart Zone button in the upper-left corner of the window.

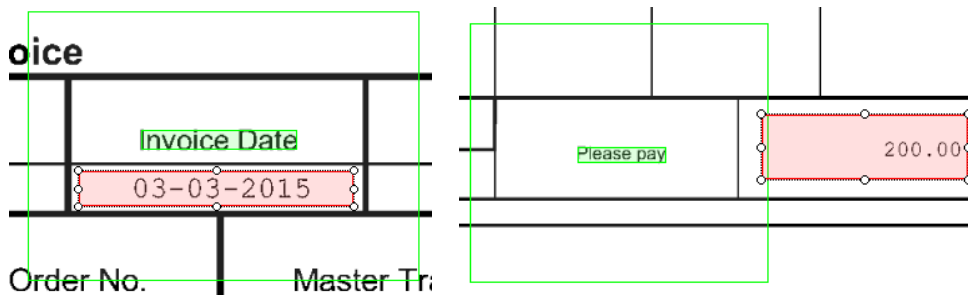


This will take you back to the Zone Configuration window.

Now you will assign a name to your newly created Smart Zone. Because this zone is designed to extract the Invoice Number, you will name the zone **Invoice Number**.



Now repeat the process for the other 2 Smart Zones (Invoice Date and Invoice Total) on the Alphalab invoice.



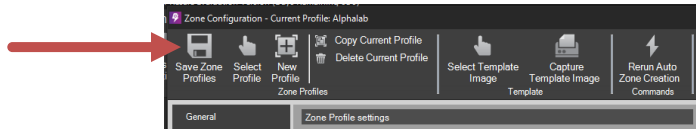
1. Draw the Smart Zone.
2. Select the proper Form Field Type and click Add.
3. Click the Locate button to find the anchor on the template image.
4. Use Selection Zoom around the anchor to enhance precision.
5. Click the lower Add button to create a child zone.
6. Re-position and re-size the child zone appropriately.
7. Save out of the Smart Zone Configuration window.
8. Name the zone according to the data it will extract.

Once all your Smart Zones have been created, the list of Zone names will look like this:

Zones								
	Zone Name	Page						
1	Invoice Number	1						
2	Invoice Date	1						
3	Invoice Total	1						

You will return to the Zone Configuration to build Zone Profiles for the other two vendors shortly, but first you will return to the indexing tab to map the index fields to the newly created zones. You will also Enable **Auto Zone Creation** for the index fields. This will cause the Smart Zones for the other two vendors to be positioned automatically, rather than requiring you to draw each zone and its child by hand.

Save out of the Zone Configuration by clicking the **Save** button on the toolbar.



☆ Note:
You do not need to set an action on the Zone tab for the Vendor Name field, because you are populating it using the Classification Form ID instead.

Select **Invoice Number** from the list of fields on the Indexing tab. Click the **Zone** tab on the right. From the **Action** dropdown, select **OCR**. In the **Zone** dropdown below, select the proper zone name if it has not already been selected automatically. Then check the box to **Auto Create Zones on new Zone Definition Profiles**. Click the **Setup** button in order to configure the type of zone to auto-create.

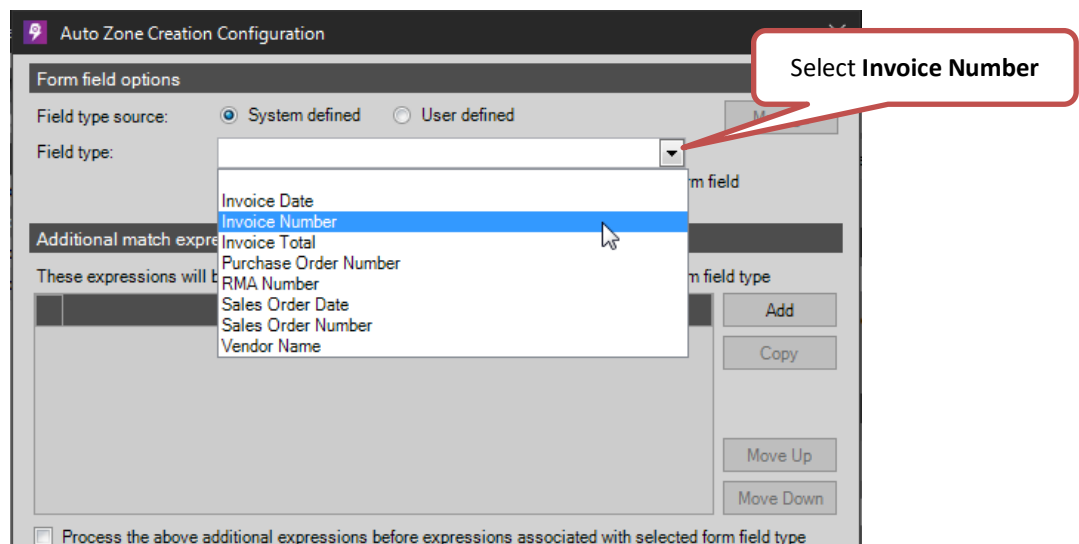
A screenshot of the 'Zone Configuration' interface with several numbered callouts in red boxes:

- 1. Click the index field**: Points to the 'Invoice Number' field in the 'Document Index Fields' table.
- 2. Select the Zone tab**: Points to the 'Zone' tab in the top right navigation bar.
- 3. Set the Action as OCR**: Points to the 'Action' dropdown menu, which has 'OCR' selected.
- 4. Choose matching zone name**: Points to the 'Zone' dropdown menu, which has 'Invoice Number' selected.
- 5. Enable Auto Zone Creation**: Points to the checked checkbox 'Auto create Zones for new Zone Definition Profiles'.
- 6. Click Setup to configure Auto Zone creation**: Points to the 'Setup' button.

☆ Note:
Index Fields will be automatically mapped to the proper zones if their names match exactly.

{This area is intentionally left blank, please continue to the next page}

From the **Auto Zone Creation Configuration** window, select the matching **Field Type** from the dropdown. This will cause other settings in the same window to default to the options that are most commonly associated with this type of field. The software will use a pre-defined list of Regular Expressions to locate the **Invoice Number** label anchor. When attempting to position the Child Zone, it will look near that anchor for Text with at least 1 Numeric Character.



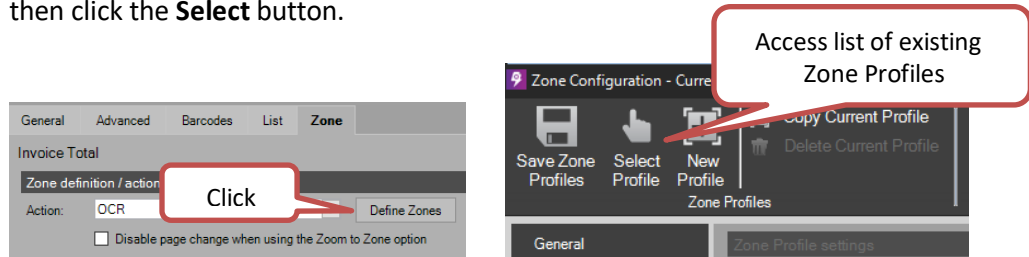
After choosing the Form Field Type, click **Save** at the bottom of the window to close the Auto Zone Creation Configuration window.

Repeat these same steps for the **Invoice Date** and **Invoice Total** index fields:

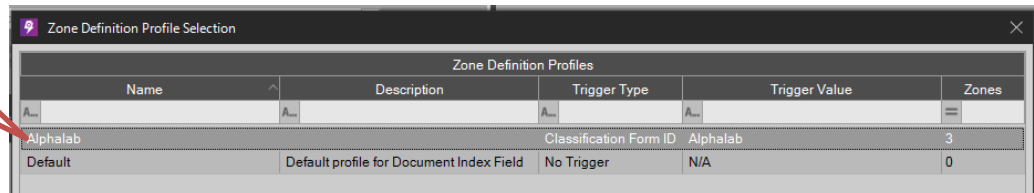
1. Click the field to select it.
2. Click the Zone tab.
3. Set the Action as OCR.
4. Ensure that the proper zone is mapped from the dropdown below.
5. Check the box to enable Auto Zone Creation.
6. Click the Setup button to open Auto Zone Creation Configuration.
7. Select the proper Field Type from the dropdown.
8. Click the Save button to finish configuring the Auto Zone Creation.

Now all the Index fields except Vendor Name are mapped to their corresponding zones. For the other two Zone Profiles that you are about to build, the software will attempt to position the 3 Smart Zones automatically.

Click the **Define Zones** button again to return to the Zone Configuration. Then, click the **Select Profile** button to select the Alphalab Profile that you created previously, then click the **Select** button.



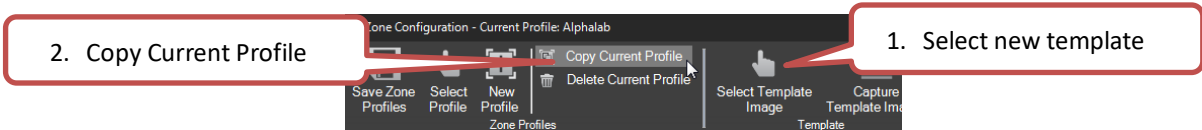
Select Alphalab profile



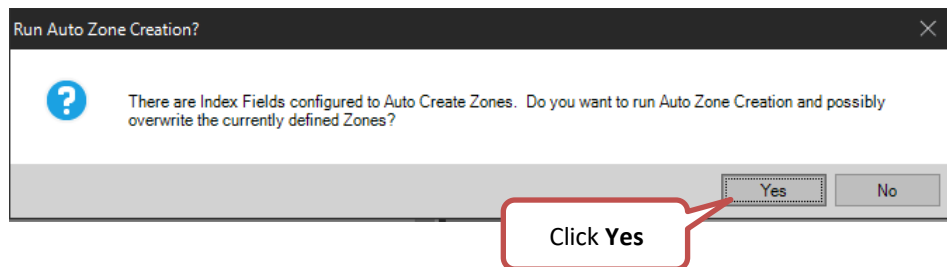
☆ Note:
Copying Zone Profiles does not remove the template image. You will need to manually overwrite it by selecting a new template image.

The profiles that you will create for the other two vendors will have the same list of zones, although their placement will differ. With the Alphalab profile loaded, click the button to **Copy Current Profile**.

Then overwrite the Template Image with the Unidex Invoice, which is **Page 2** of the same **03 - Classification Guide Samples.tif**.



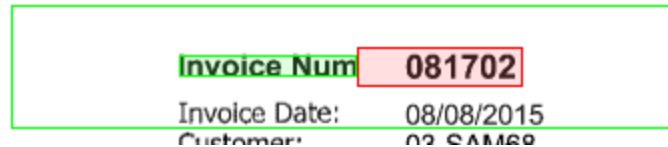
Because Auto Zone Creation is enabled, the message box shown below will pop up as soon as the template image loads.



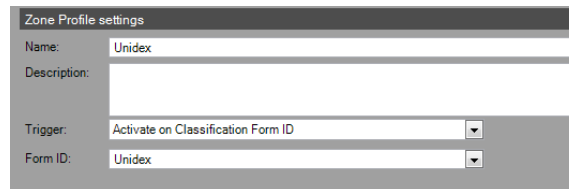
☆ Note:

If the automatic Smart Zone Creation failed to locate the correct data, you may need to make such manual corrections as dragging the zones elsewhere or adding additional custom anchor expressions.

Notice that the position of the Smart Zones automatically changes to accommodate the data on the newly loaded Unidex document. Verify the position of each green Smart Zone, then double-click it to open the Smart Zone Configuration and verify the position of its Child Zone.



If the zones are positioned correctly, then you do not need to make any changes.

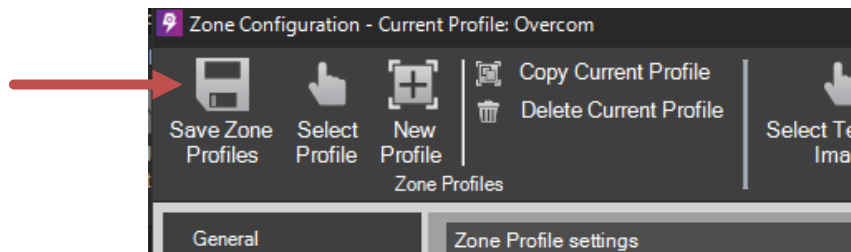


Fill out the **Zone Profile Settings** section using Unidex for the name and selecting the Unidex Form ID to trigger off of.

Now repeat the process for the third vendor, **Overcom**, using **Page 5** from the sample document.

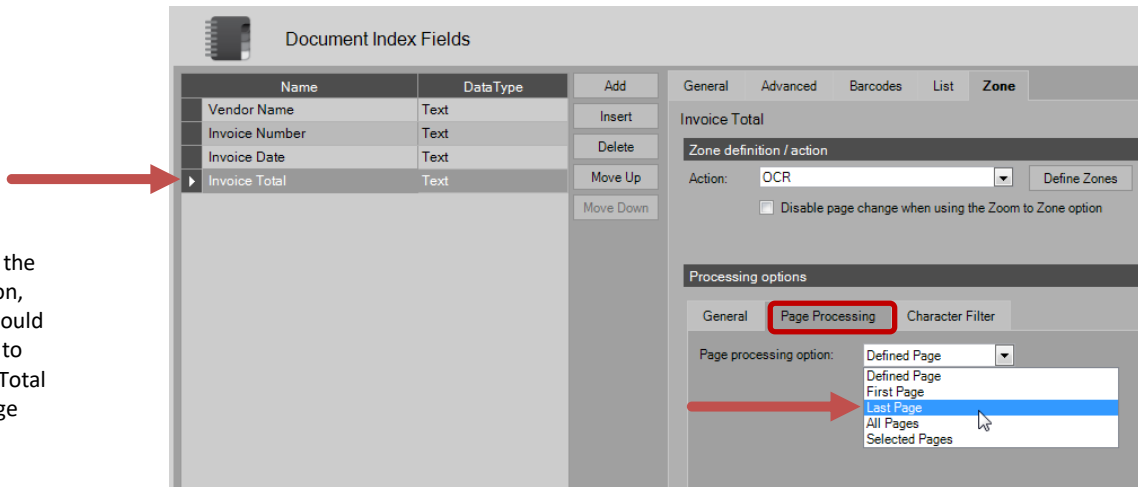
1. Copy Current Profile.
2. Load in the Template Image (Page 5 only).
3. Click Yes on the Run Auto Zone Creation alert.
4. Verify the position of all 3 Smart Zones and their Child Zones.
5. Fill out the proper Overcom information on the Zone Definition Profile tab.

Now that you've built all 3 Zone Profiles, **Save** out of the Zone Configuration.



For the **Invoice Total** field, you will make one additional change. On the Zone tab, click the **Page Processing** tab. Change the Page Processing Option dropdown to **Last Page**.

☆ Note:
If you did not change the Page Processing option, the Capture Profile would fail to extract a value to populate the Invoice Total field on any multi-page invoices.

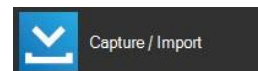


You are choosing this option because the Total always appears on the Last Page of the document (which would also be the First Page of the document in the case of single-page invoices).

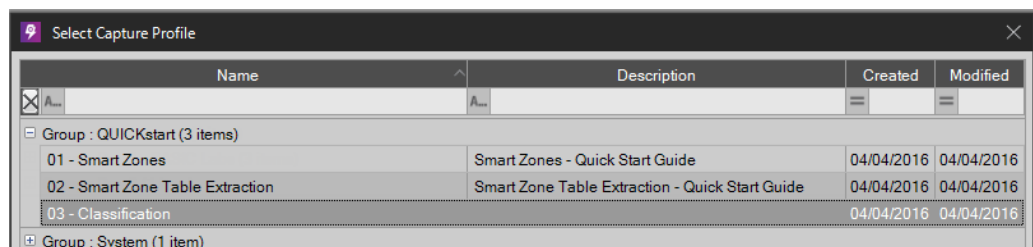
You can now **Save** the Capture Profile, and then **Save** out of the Configuration module to return to the Home Screen.

Run the Capture Profile

Run the Capture Profile by clicking the Capture / Import Module icon.

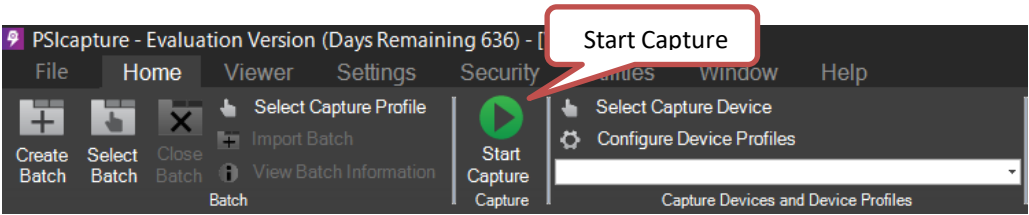


Select the **03 – Classification** Capture Profile, Click **OK**.

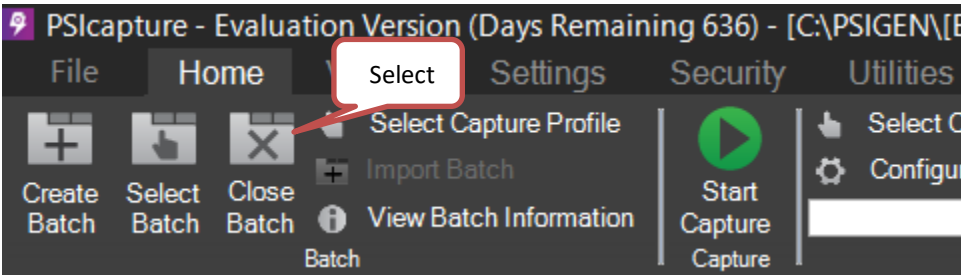


Using the File Browser window, navigate to where you saved the files for this guide. Select the folder on top and then select **03 – Classification Guide Samples.tif** from the lower pane. Highlight the file and select the green **Start Capture** button or Control+S to begin importing the documents into PSImapture.

☆ Note:
If you do not see the File Browser toward the left of the PSIsapture screen, click the **Select Capture Device...** icon to switch to the **Import** mode.



When the file has finished importing, you will click the **Close Batch** button.



Ensure the selection for **Close Batch** and **Continue in next Workflow step (Classification)** is selected. Click **OK**.

The Classification module will attempt to classify the invoices according to the rules that you created in the Form Definitions. One of the documents will fail to classify due to a mark over the phone number, causing the **Classification Validation** window to appear.

☆ Note:
Notice that one of the Overcom invoices consists of 2 pages that have been merged together to form a multi-page invoice.

Failed to auto-classify

Classification results						
	Document	Classification Form ID		Pages	<input type="checkbox"/> Reject	
▶	1	Alphalab		1	<input type="checkbox"/>	⏪ ⏩
	2	Unidex		1	<input type="checkbox"/>	⏪ ⏩
	3	Overcom	→	2	<input type="checkbox"/>	⏪ ⏩
	4	Overcom		1	<input type="checkbox"/>	⏪ ⏩
	5			1	<input type="checkbox"/>	⏪ ⏩
	6	Alphalab		1	<input type="checkbox"/>	⏪ ⏩

{This area is intentionally left blank, please continue to the next page}

Click on Document 5, which failed to classify. Look at the image of the Unidex invoice on the right. Notice the large mark obscuring the vendor's phone number. This mark has prevented the OCR engine from finding a successful match to the Classification rule, which was created based on the complete phone number, resulting in the Classification failure.

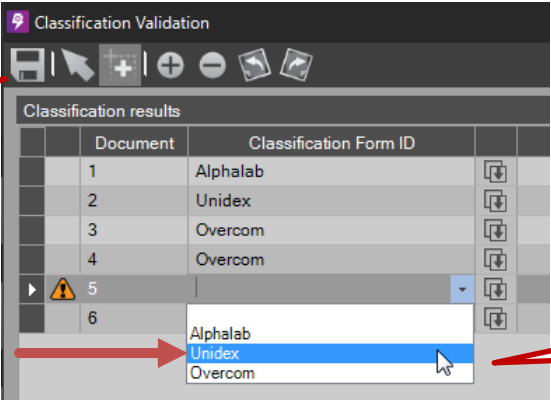
☆ Note:

In production, it is a best practice to use *multiple* Classification rules in order to avoid these types of problems. For example, you could have also added a second rule to look for the vendor's address.



Obscuring checkmark

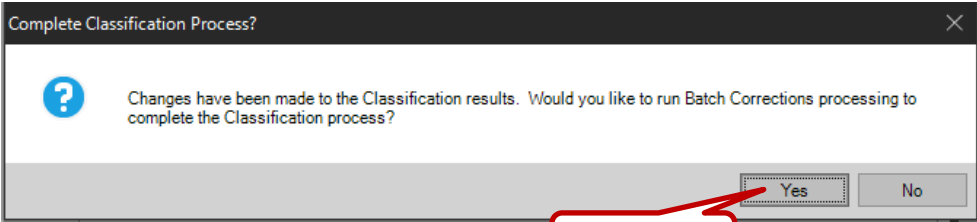
The document should have been classified as Unidex. Select **Unidex** from the dropdown or just begin typing the word Unidex in the box. Then click **Save**.



1. Click **Save**

2. Manually select Form ID

Click **Yes** to confirm your changes and Complete the Classification Process.



Click **Yes**

The Index module will auto-index the invoices and will then pass the batch directly to the Quality Assurance module, from which you will be able to review the extracted data.

Quality Assurance

valid Item

Tree View Data View

Batch 20191205.151437 (Edit Mode Enabled)

	Document Name	Vendor Name	Invoice Number	Invoice Date	Invoice Total
	00000001	Alphalab	34633895	03/03/2015	200.00
	00000002	Unidex	081702	08/08/2015	300.00
	00000003	Overcom	65165111	02/02/2013	2856.92
	00000005	Overcom	65161354	09/09/2013	236.34
	00000006	Unidex	081624	07/07/2015	1,000.00
	00000007	Alphalab	34634601	04/04/2015	2125.00

Click inside each individual cell of the Data View to view the exact zone on the page from which the data was extracted. PSIs capture knows which vendor each invoice belongs to, and where to find the data it's looking for on each vendor's document.

For Document 3, click back and forth between the Invoice Date field and the Invoice Total field to see that those values were extracted from two different pages of the same document.

1. Click

2. Click

Total: 2856.92

PSI capture - Not F

CONGRATULATIONS!!
 You've just configured and tested a
 Capture Profile using Classification.