

Table Extraction Algorithm & UI Revamp

Ephesoft offers you a various mechanism to extract tables. For this, you have to configure various table using one or the combination of below-mentioned extraction types:

- Column Regex Pattern
- Column Header Pattern
- Column Coordinates

Earlier, table configuration screen did not have provision to display the image along with the configuration, hence making it difficult to configure the extraction rules. The user had to manually open the image file while configuring table extraction. Moreover, all extraction rules were manually entered by the user, with no suggestions from the system.

Also, while testing the configured extraction rules, you could not view the area of the image from which the table rows were extracted. This could create accuracy concerns if wrongly extracted data skipped the operator's eye.

With Ephesoft Transact v4.1.0.0, you have the following improvements:

- Complete table extraction rule can be configured with image view in one screen.
- Regex suggestion utility is now included while configuring table extraction.
- Regex is suggested for Column pattern, Column Header Pattern, Between Left Pattern, Between Right Pattern, Start Coordinate, and End Coordinate so that user don't need to type them over.
- User can test configuration on the image file from the configuration page itself with result overlays for assistance.
- Support for uploading multipage image and configure rules accordingly.

The screenshot displays the Ephesoft Transact v4.1.0.0 interface. On the left, the 'Extraction Rule' configuration window is open, showing the 'Rule Name' as 'Invoice Table', 'Start Pattern' as 'ACME', and 'End Pattern' as '\$19.19'. Below this, the 'Table Extraction API' section shows 'Column Coordinates', 'Column Header', and 'Regex Extraction' options. The 'Column Configuration' section shows 'Table Column' as 'Part No', 'Column Header Pattern' as 'Part No', and 'Column Pattern' as '[vd]{13}'. The main area displays an 'INVOICE' image with extracted data. The invoice includes the company name 'ACME Company', address 'Belverly Hills Blvd, Irvine, CA 90210', and contact information. It also lists the 'Invoice No: 5432000', 'Invoice Date: 04/06/08', and 'PO Number: 2005012345'. A table of items is extracted from the invoice, showing columns for 'Part No', 'Quantity', 'Unit Price', 'Description', 'Discount', and 'Total'. The table contains two rows of data. A QR code is visible at the bottom left of the invoice image. The bottom right of the invoice shows the 'Sub Total: \$17.81', 'Tax 7.75%: \$1.38', and 'Total: \$19.19'.

Part No	Quantity	Unit Price	Description	Discount	Total
9981000000156	50	\$0.16	FS B&W Card Stock	0.00%	\$8.00
9981000001990	9	\$1.09	Cutting-Per Reim	0.00%	\$9.81

Sub Total: \$17.81
Tax 7.75%: \$1.38
Total: \$19.19

In addition of simplifying the table configuration, various improvements have been made at the backend to support number of table variations:

- Tables which are not in tabular form but more in horizontal manner, Ephesoft Transact calls them horizontal table.
- Tables containing a mixture of horizontal columns and vertically aligned columns.
- Tables containing wrapped header, for example Unit Price, where Unit is in one line and Price in next.
- Tables with columns spreading across multiples lines, for example, Order Number, Part Number and Description, while Unit Price are in first line whereas, Quantity, Unit Price and Total Price in next line.
- Support columns whose data is not aligned under the header. Earlier, column header extraction could not extract such columns. Column coordinate and regex based extraction could be used in such cases but they had their own limitations. They both used to fail if column ordering changed.

There are few more additional features added Ephesoft Transact v4110:

- Allow user to extract repeating tables. This is to cater tables which are present in multiple pages of a document. For example, similar format invoice details are present in first and third page of a document. In previous versions, user has to configure multiple tables to extract those tables. In 4.1.1.0 version, such tables can be extracted in one table only.
- Table extraction now accounts for OCR confidence too at the time of validation.
- User can configure default values for a column that will be set if no data could be extracted for this column.

Next section will describe in detail how to configure table extraction rule.

Setting up Table Extraction

Ephesoft Transact supports data extraction from various table formats. Using a combination of configuration settings provided by Ephesoft, you can easily and accurately extract data from the following table formats:

- header and data are horizontally aligned
- header is horizontal and vertically aligned
- data is not aligned under header
- header is wrapped in multiple rows
- header is spanned in multiple lines and so the data is spanned.

Extraction of data from tables involves the following tasks:

- Configuring tables
- Configuring table columns
- Configuring table extraction rules
- Configuring column extraction rules

- Testing table extraction

Configuring Tables

The first step in setting up table extraction is to define the table within Ephesoft.

To configure a table

1. From the DCMA Home page, click **ADMINISTRATOR** and select **BATCH CLASS MANAGEMENT**.

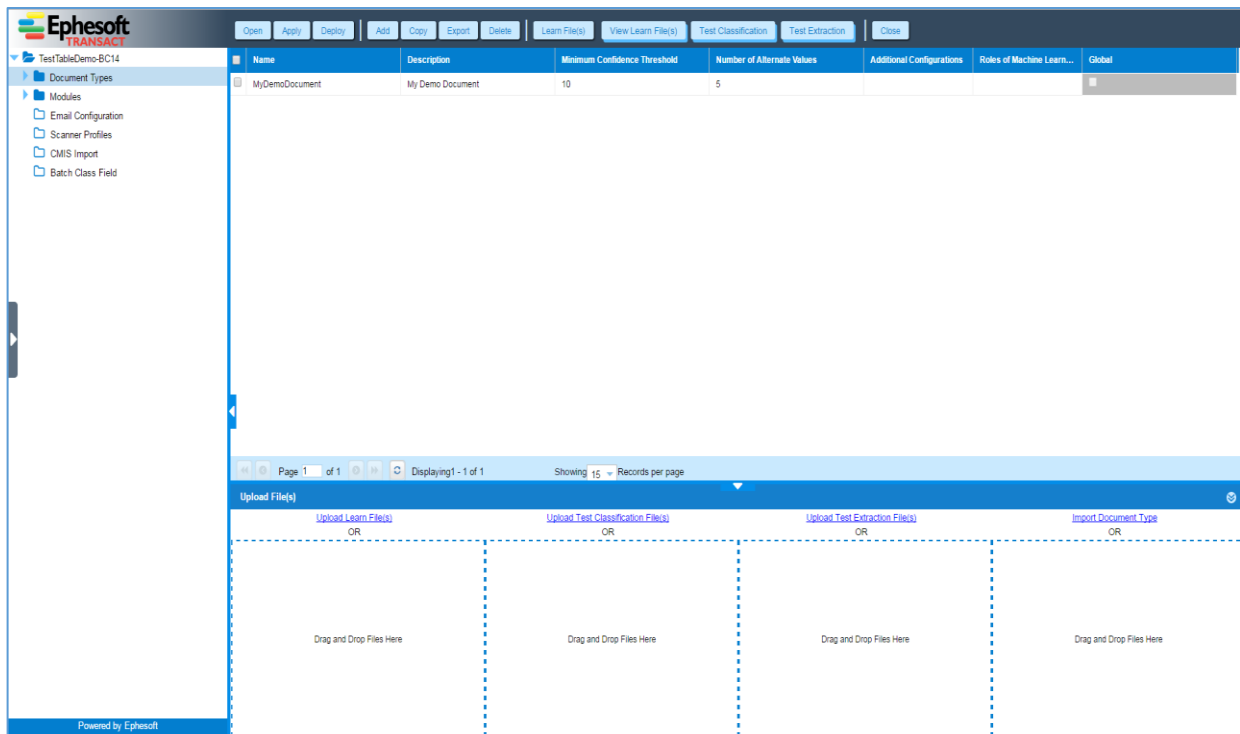
The Ephesoft Enterprise **Login** page displays.

2. Enter valid credentials to login.

The **Batch Class Management** screen displays.

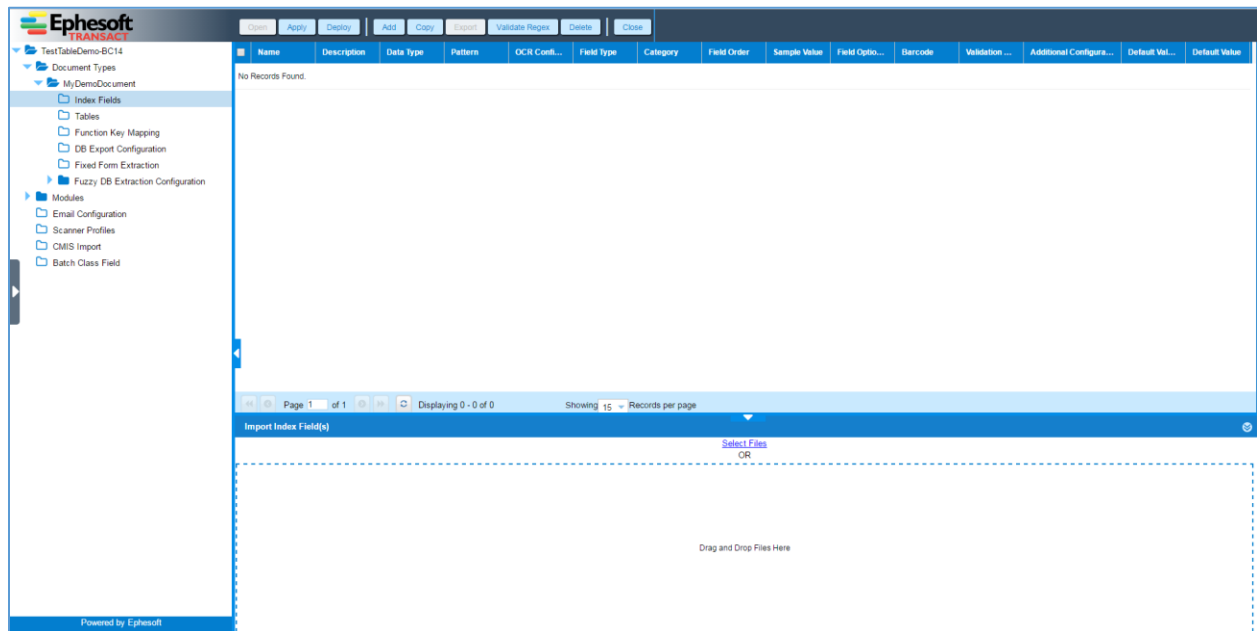
3. Select the batch class in question and click **OPEN** from the toolbar on top of the **Batch Class Management** screen.

The batch class opens with a list of document types.



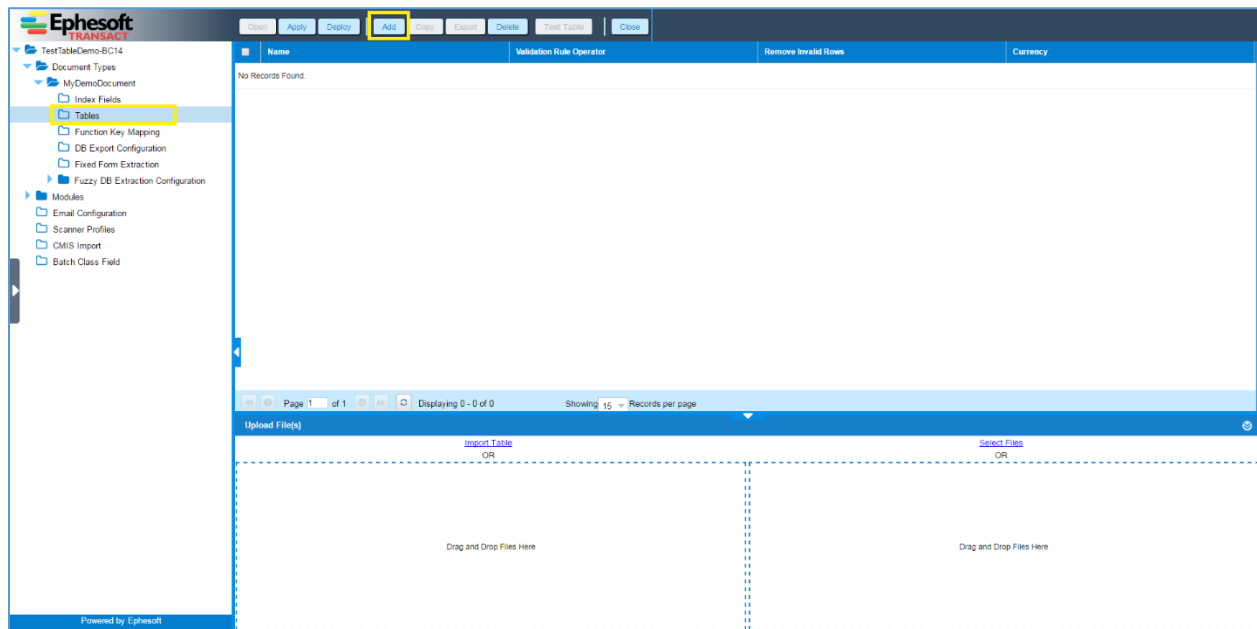
4. Select the document type from the list and click **OPEN** from the toolbar on top of the page.

The document type page displays with the **Index Fields** node selected by default in the left navigation pane.



4. From the left navigation pane select the **Tables** node.

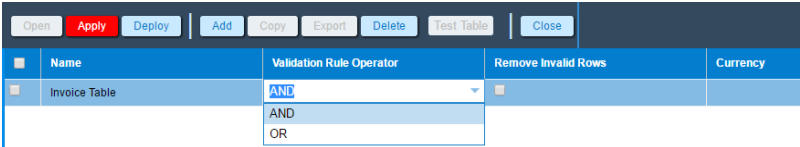
The following screen displays.

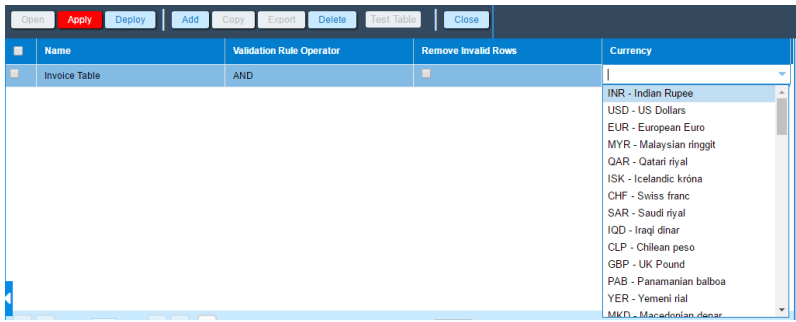


5. Click **ADD** from the toolbar on top of the page.

The following screen displays.

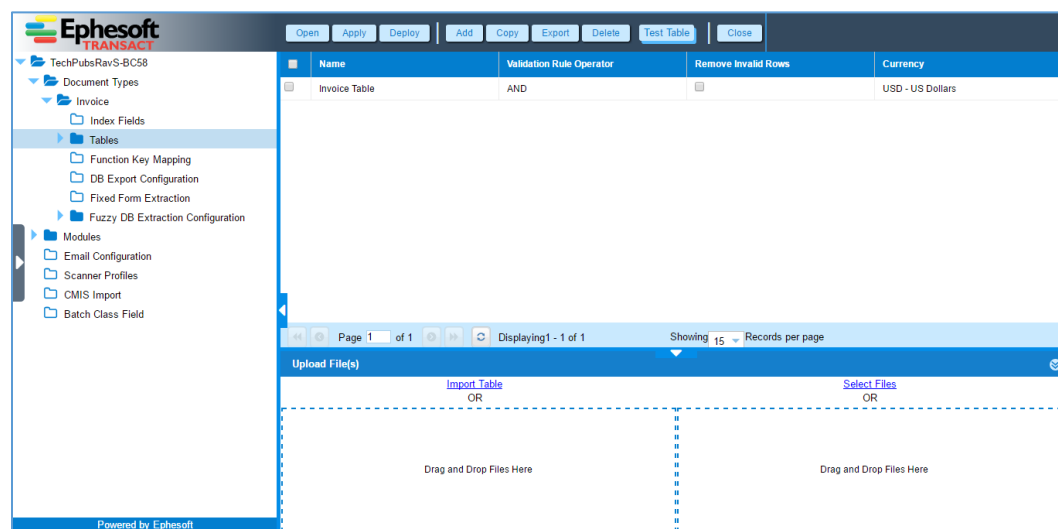
6. Enter the relevant information as described in the table below:

Item	Description
Name	Enter a name for the new table. Use a short name here, such as Invoice Table, Form1040, or HUD01.
Validation Rule Operator	<p>Select an operator for the validation rule. The available options are AND and OR. By default, AND operator is selected.</p>  <p>As we know user can configure multiple Rules to validate the Row of a table, this switch helps the algorithm to mark the Row validate based on switch.</p> <p>If “AND” is set: The row is valid if all the Row validation rules are satisfied.</p> <p>If “OR” is set: The row should validate at least one of the Row validation rule.</p>
Remove Invalid Rows	Select this checkbox if you want to eliminate any rows from the extracted data that do not match the validation rule.
Currency	Select the currency format that you want to use with the extracted data from the Currency drop-down list.

Item	Description
	

- Click **APPLY** and **DEPLOY** from the toolbar on top of the page to save the table details.

The **Tables** node page is updated with details of the newly added table.



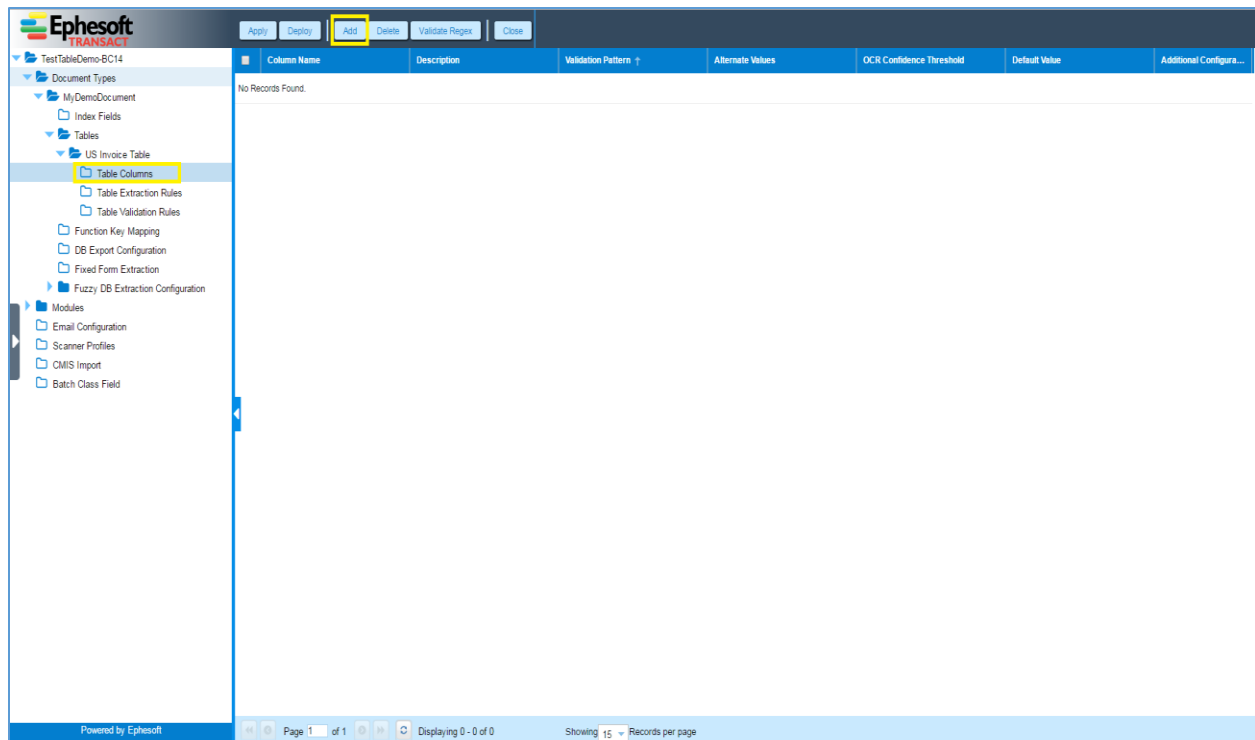
Configuring Table Columns

After defining the table, you must define the columns that constitute the table.

To configure a table column

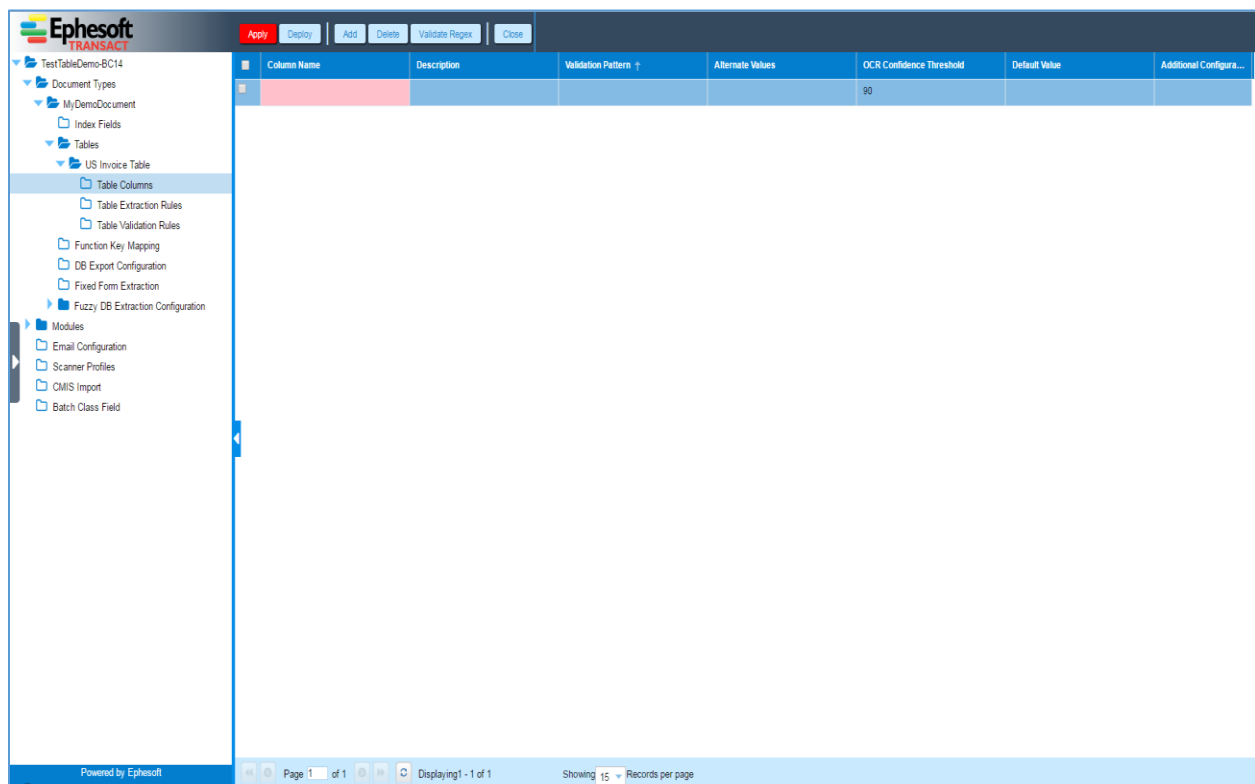
- On the **Tables** node page ([Step 12 above](#)), click **OPEN** from the toolbar on top of the page. Alternatively you can click on table node on left pane to open table and configure table columns.

The following screen displays with the **Table Columns** node selected in the left navigation pane.




2. Click **Add** from the toolbar on top of the page.

The following screen displays.



3. Enter the relevant information as described in the table below:

Item	Description	
Column Name	Enter a name for the new table column.	
Description	Enter a brief description about the new table column.	
Validation Pattern	<p>Enter a validation pattern for the new table column name using the Regex Pool or the Regex Builder widgets.</p> <p>This field is not mandatory. You can enter the validation pattern based on requirement.</p> <div>  <p>Click Validate Regex from the toolbar on top of the page to validate any regex patterns that you enter.</p> </div>	
Alternate Values	It provides the pre-populated list of values for a particular column to assist user while selecting data while validating tables.	
OCR Confidence Threshold	The default value is 90. You can edit this and enter a value between 0 to 100.	
Default Value	In case you want to set default value for any column when no value is extracted for particular column, you can enter the default value in this field.	
Additional Configuration	New Row Anchor	This switch is used to mark any column to indicate the start of new row if the value for particular column gets extracted.
	Required	This switch is used to mark any column mandatory for a given row. If the row does not have the value for the column marked as required, the row is discarded.
	Currency	This switch is used to mark any column as currency type.

4. Click **APPLY** from the toolbar on top of the page to save the column details.

The **Table Columns** node page is updated with details of the newly added table column(s).

You can keep adding columns as per requirement as shown below -

The screenshot shows the Ephesoft TRANSACT application. On the left, a navigation pane lists various configuration nodes. The main area displays a table with columns for configuration. The table has the following data:

Column Name	Description	Validation Pattern	Alternate Values	OCR Confidence Threshold	Default Value	Additional Configura...
Part No	Part No			90		
Quantity	Quantity			90		
Unit Price	Unit Price			90		
Description	Description			90		
Discount	Discount			90		
Total	Total			90		

At the bottom of the interface, there is a status bar indicating "Page 1 of 1", "Displaying 1 - 6 of 6", and "Showing 15 Records per page".

Configuring Table Extraction Rules

After defining the table columns, you must define the extraction rules for the table.

To configure a table extraction rule

1. From the left navigation pane, select the **Table Extraction Rules** node.

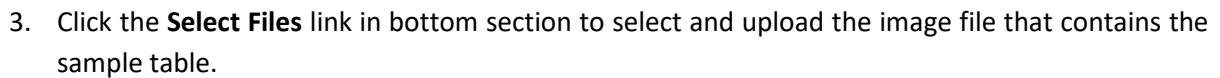
The following screen displays.

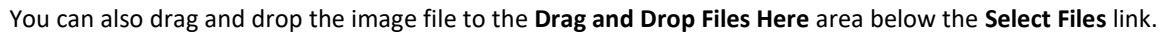
The screenshot shows the Ephesoft TRANSACT application with the 'Table Extraction Rules' node selected in the left navigation pane. The main area displays a table with columns for configuration. The table has the following data:

Rule Name	Start Pattern	End Pattern	Extract Repeating Table...	Overlapping Columns ...	Table Extraction API
No Records Found.					

At the bottom of the interface, there is a status bar indicating "Page 1 of 1", "Displaying 0 - 0 of 0", and "Showing 15 Records per page".

- The **Extraction Rules/Column Configuration** screen displays.

[illegible]



The screenshot displays the Ephasef TRANSCACT application window. The title bar includes the Ephasef logo and menu options like Apply, Reset Column, Reset All Columns, Cancel, Validate Rules, and Test Table. Below the title bar, there's a status bar indicating "BC14 : MyDemoDocument :: US Invoice Table".

Left Panel:

- Extraction Rule:** Contains fields for Rule Name, Start Pattern, End Pattern, Extract Repeating Tables, Overlapping Columns Table, and Table Extraction API (with checkboxes for Column Coordinates, OR, Column Header, OR, and Regex Extraction).
- Columns Configuration:** Includes Basic Configuration with Table Columns, Part No, Column Header Patterns, Column Patterns, Pattern Left, Pattern Right, Start Coordinate, End Coordinate, and Extract Data From Columns.

Main Area:

INVOICE

ACME Company Belverly Hills Blvd Irvine, CA 90210 Tel: 949-331-7500	Invoice No: 5432000 Invoice Date: 04/06/08 PO Number: 2005012345
---	---

Part No	Quantity	Unit Price	Description	Discount	Total
9981000000156	50	\$0.16	FS B&W Card Stock	0.00%	\$8.00
9981000001990	9	\$1.09	Cutting-Per Reim	0.00%	\$9.81





Sub Total: \$17.81

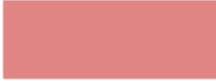

Tax 7.75%: \$1.38

Total: \$19.19

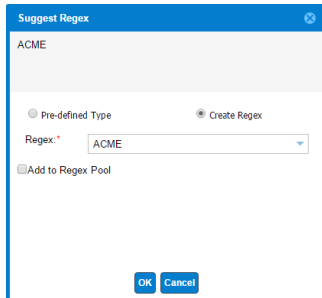
Powered by Ephasef

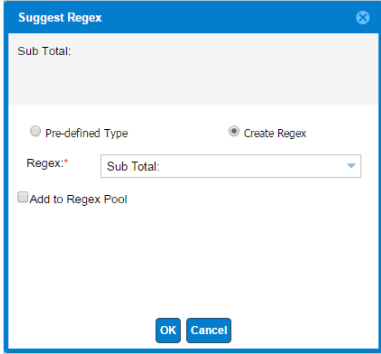
The below table provides the information related to various overlays and their usage –

Color	Description
	Start Pattern Using this overlay user can configure start pattern for any table.
	Column Header Using this overlay user can configure column header pattern and set the coordinates of any column.
	Right Pattern Using this overlay user can configure Pattern Right.
	Pattern Left Using this overlay user can configure Pattern Left.

Color	Description
	<p>Column Data</p> <p>Using this overlay user can configure Column Pattern. If user configures both Column Header and Column Data, Ephesoft system smartly identifies the type of column i.e. whether the column is horizontal or vertical. For example: If both overlays horizontally intersects each other the column is considered as horizontal. Further if both overlays are configured for vertical column the coordinates of column is calculated using the Max boundary that can be constructed using both overlays.</p>
	<p>End Pattern</p> <p>Using this overlay user can configure end pattern for any table.</p>

4. Enter the relevant information in the **Extraction Rules** section as described in the table below:

Item	Description
Rule Name	Enter a name for the extraction rule.
Start Pattern	<p>Enter a regex pattern in the Start Pattern text box to define the start point of the table</p> <p>OR,</p> <p>Use the Suggest Regex utility as described below:</p> <ol style="list-style-type: none"> Drag and place the Start Pattern overlay over an area of the image that you want to define as the beginning of the document. Double-click the overlay to display the Suggest Regex dialog box.  <ol style="list-style-type: none"> Use an existing regex from Pre-defined Type or create a new regex from Create Regex.


Item	Description	
	The Start Pattern text box in the Extraction Rule section is populated with the regex pattern.	
End Pattern	<p>Enter a regex pattern in the End Pattern text box to define the end point of the image.</p> <p>OR,</p> <p>Use the Suggest Regex utility as described below:</p> <ol style="list-style-type: none"> Drag and place the End Pattern overlay over an area of the image that you want to define as the beginning of the document. Double-click the overlay to display the Suggest Regex dialog box.  <ol style="list-style-type: none"> Use an existing regex from Pre-defined Type or create a new regex from Create Regex. <p>The End Pattern text box in the Extraction Rule section is populated with the regex pattern.</p>	
Extract Repeating Tables	Select the Extract Repeating Tables option if you want to extract table from multiple pages using the extraction rule.	
Table Extraction API	Column Coordinates	Select the Column Coordinates option to extract data based on defined column coordinates.
	Column Header	Select the Column Header option to extract data based on defined column headers.
	Regex Extraction	Select the Regex Extraction option to extract data based on defined regex patterns.
Operator	Select a rule operator to define the combination in which you want to use the extraction APIs.	
Overlapping Columns Table	This switch is used if the headers of column overlap each other. In other words, the different column headers are in multiple lines so the data is.	

5. After defining the table extraction rule, you must define the extraction rules for the table columns that constitute the table.

To configure a column extraction rule

1. Collapse the **Extraction Rule** section to have a better view of the **Column Configuration** section.

The **Extraction Rules/Column Configuration** screen is updated displaying the complete **Column Configuration** section.



Apply
Reset Columns
Reset All Columns
Cancel
Validate Regex
Test Table

Image US-Invoice.tif

Page No. 1

BCE :: MyDemoDocument :: US Invoice Table :: My Demo Rule

Extraction Rule

Column Configuration

Basic Configuration

Table Column: Part No

Column Header Pattern:

Column Pattern:

Pattern Left:

Pattern Right:

Start Coordinate:

End Coordinate:

Extract Data From Column:

Additional Configuration

Use Default Column Configuration: ☒

New Row Anchor: ☐ Required: ☐

Currency: ☐

INVOICE

ACME Company
 Belverly Hills Blvd
 Irvine, CA 90210
 Tel: 949-331-7500

Invoice No: 5432000
 Invoice Date: 04/06/08
 PO Number: 2005012345

Part No	Quantity	Unit Price	Description	Discount	Total
9981000000156	50	\$0.16	FS B&W Card Stock	0.00%	\$8.00
99810000001990	9	\$1.09	Cutting-Per Reim	0.00%	\$9.81

Sub Total:


Tax 7.75%

Total

\$17.81

\$1.38

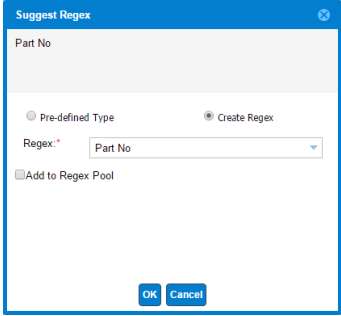
\$19.19



Powered by Ephesoft

5. Enter the relevant information for each column in the **Column Configuration** section as described in the table below:

Item	Description
Table Column	From the Table Column drop-down list, select the column for which you are defining the column extraction rule. The drop-down list displays all the table columns that you have configured.
Column Header Pattern	<p>Enter a regex pattern in the Column Header Pattern text box to define the location of the column header in the image.</p> <p>OR,</p> <p>Use the Suggest Regex utility as described below:</p> <ol style="list-style-type: none"> Drag and place the Column Header overlay over an area of the image that you want to define as column header.

Item	Description
	<p>b. Double-click the overlay to display the Suggest Regex dialog box.</p>  <p>c. Use an existing regex from Pre-defined Type or create a new regex from Create Regex.</p> <p>The Column Header Pattern text box in the Column Configuration section is populated with the regex pattern.</p>
Column Pattern	<p>Enter a regex pattern in the Column Pattern text box to define the location of the column data on the image.</p> <p>OR,</p> <p>Use the Column Data overlay to populate the Column Pattern text box with suggestions from the Suggest Regex utility as described above.</p>
Pattern Left	<p>Enter a regex pattern in the Pattern Left text box to define the area on the left side of the column data in the image.</p> <p>OR,</p> <p>Use the Pattern Left overlay to populate the Pattern Left text box with suggestions from the Suggest Regex utility.</p>
Pattern Right	<p>Enter a regex pattern in the Pattern Right text box to define the area on the right side of the column data in the image.</p> <p>OR,</p> <p>Use the Pattern Right overlay to populate the Pattern Right text box with suggestions from the Suggest Regex utility.</p>
Start Coordinate	<p>This field is used to extract data if extraction API is column coordinate. When you draw Column Header or combination of Column Header and Column Data the start coordinate of column is set.</p>
End Coordinate	<p>This field is used to extract data if extraction API is column coordinate. When you draw Column Header or combination of Column Header and Column Data the end coordinate of column is set.</p>

6. Similarly you can configure other columns of table by selecting column from Table Column Dropdown as shown –

INVOICE

ACME Company
Belverly Hills Blvd
Irvine, CA 90210
Tel: 949-331-7500

Invoice No: 5432000
Invoice Date: 04/06/08
PO Number: 2005012345

Part No	Quantity	Unit Price	Description	Discount	Total
9981000000156	50	\$0.16	FS B&W Card Stock	0.00%	\$8.00
9981000001990	9	\$1.09	Cutting-Per Reim	0.00%	\$9.81
Sub Total:					\$17.81
Tax 7.75%					\$1.38
Total					\$19.19

6. After configuring the column(s), click **TEST TABLE** from the toolbar on top of the **Extraction Rules/Column Configuration** screen.

The extracted results are displayed in the **Test Table Results** grid.

Belverly Hills Blvd
Irvine, CA 90210
Tel: 949-331-7500

Invoice Date: 04/06/08
PO Number: 2005012345

Part No	Quantity	Unit Price	Description	Discount	Total
9981000000156	50	\$0.16	FS B&W Card Stock	0.00%	\$8.00
9981000001990	9	\$1.09	Cutting-Per Reim	0.00%	\$9.81
Sub Total:					\$17.81
Tax 7.75%					\$1.38
Total					\$19.19

S.No.	Part No	Quantity	Unit Price	Description	Discount	Total
1	9981000000156	50	\$0.16	FS B&W Card Stock	0.00%	\$8.00
2	9981000001990	9	\$1.09	Cutting-Per Reim	0.00%	\$9.81

7. Click **APPLY** from the toolbar on top of the page to save the column extraction rule(s).

The **Table Column Extraction Rules** node page now displays the details of the newly added column extraction rule(s).

Column Name	Column Pattern	Between Left	Between Right	Column Header Pattern	Start Coordinate	End Coordinate	Extract Data From ...	Additional Configur...
Part No	((d(1,3)(.1)(0,4)...			Part No	163	341		
Quantity	+			Quantity	566	780		
Unit Price				Unit Price	866	1108		
Description				Description	1146	1463		
Discount				Discount	1653	1958		
Total				Total	1989	2167		

Page 1 of 1 | Displaying 1 - 6 of 6 | Showing 15 Records per page

8. Click **APPLY** and **DEPLOY** to save the changes.

Extract Repeating tables

Say a table is present in multiple pages of document and you want to extract the data of all table as a single result. Now this can be achieved using “Extract Repeating tables” switch.

Extraction Rule

Rule Name: My Demo Rule

Start Pattern: ACME

End Pattern: Sub Total:

Extract Repeating Tables: ☐

Overlapping Columns Table: ☐

Table Extraction API:

☐ Column Coordinates ☒ OR ☒ Column Header ☐ OR ☐ Regex Extraction

You can configure start pattern and end pattern for various tables in different pages. The resultant extracted data will be in a single table.

Configuring Horizontal Table

1. The horizontal table can be visualized as table in which header and corresponding value are in same line as shown below –

MONTHLY SERVICES Sep 1 - Sep 30	
Cable	
Cox Limited Basic	\$20.00
Cox Expanded Service	30.99
Digital Cable Service	14.95
Digital Receiver - HD	5.50
Total Cable	\$71.44
Internet	
Cox HSI Preferred w/ Powerboost	\$44.99
TOTAL MONTHLY SERVICES	\$116.43

2. Such tables do not require any additional configuration from the user. Ephesoft system is smart enough to recognize whether table is horizontal or vertical based on the relative position of **Column Header** overlay and **Column Data** overlay as shown below –

The screenshot displays the Ephesoft Transact software interface. On the left, the 'Extraction Rule' configuration panel is visible, showing the rule name 'My Horizontal Table', start pattern '< MONTHLY SERVICES', and end pattern 'TOTAL MONTHLY SER'. The 'Table Extraction API' section shows 'Column Coordinates' and 'Column Header' options. The 'Column Configuration' section shows 'Table Column' set to 'Monthly Service', 'Column Header Pattern' set to '< MONTHLY SERVICES', and 'Column Pattern' set to '[A-Za-z]{3}[0-9]{1,2}[0-9]'. The 'Pattern Left' and 'Pattern Right' fields are empty, and the 'Start Coordinate' is set to '159'.

The main window displays a sample bill from Cox Communications. The bill includes a 'SUMMARY OF CHARGES' section with the following data:

Monthly Services	\$116.43
Taxes, Fees and Surcharges	3.98
TOTAL NEW CHARGES	\$120.41

Below this, the 'MONTHLY SERVICES' section is highlighted, showing the same data as the table above. The 'TAXES, FEES AND SURCHARGES' section is also visible, showing 'Cable Fees' of \$0.07 and 'Franchise Fee' of 3.91.

On the right side of the bill, there is a 'News From Cox' section with a headline about Floyd Mayweather and a 'CUSTOMER INFORMATION' section with 'Advance Billing' details. A QR code is located at the bottom right of the bill.

3. In above image you can see the **Column Header** overlay and **Column Data** overlay horizontally intersect each other which is interpreted as horizontal header by Ephesoft.

- Once done with configuring column(s) click on Test Table button to see the extracted table as shown below –

Extraction Rule Configuration:

- Rule Name: My Horizontal Table
- Start Pattern: < MONTHLY SERVICES >
- End Pattern: TOTAL MONTHLY SER
- Table Extraction API: ☐ OR ☒ Column Header ☒ AND ☒ Regex Extraction

Column Configuration:

- Table Column: Monthly Service
- Column Header Pattern: < MONTHLY SERVICES >
- Column Pattern: [A-Za-z]{3}\s\d{1,2}\s-\s[0-9]
- Pattern Left:
- Pattern Right:
- Start Coordinate: 159

Test Table Results:

S.No.	Monthly Service	Cox Limited Basic	Cox Expanded Service	Digital Cable Service	Digital Receiver	Total Cable
1	Sep 1 - Sep 30	\$20.00	30.99	14.95	5.50	\$71.44

Configuring Table with Wrapped Header

- Ephesoft Transact provides to extract tables with wrapped headers using Column Header API or the combination with other Extraaction API.
- For this there is no any explicit configuration required by you. Ephesoft system smartly recognise whether the header is wrapped.
- Below image shows how you can configure wrapped column header –

Extraction Rule Configuration:

- Rule Name: Table With Wrapped Header
- Start Pattern: Qty
- End Pattern: Product Summary
- Table Extraction API: ☐ OR ☒ Column Header ☐ AND ☐ Regex Extraction

Column Configuration:

- Table Column: Unit Price
- Column Header Pattern:
- Column Pattern:
- Pattern Left:
- Pattern Right:
- Start Coordinate: 2583

Test Table Results:

Qty	UOM	Description	VND	CL	CD	Item Number	NDC/UPC	List	Unit Price	Extended Amount
1	EA	AZASAN 100 MG TAB 100	B-L	RX	P	10058087 935698	6 56490-241-41	667.13	530.48	530.48

- Click on Test Table button to check the extraction result –

Ephesoft
TRANSACTION

Apply Reset Column Reset All Columns Cancel Validate Regex Close

Image Demo.tif Page No. 1

Remi PO Box 100565 PASADENA CA 91189-0565 Sold 1606 W. WHISPERING WIND DR. 2ND FLO PHOENIX AZ 85085

Extraction Rule

Column Configuration

Basic Configuration

Table Column: Unit Price

Column Header Pattern: Unit Price

Column Pattern:

Pattern Left:

Pattern Right:

Start Coordinate: 2583

End Coordinate: 2813

Extract Data From Column:

Additional Configuration

Use Default Column Configuration: ☒

New Row Anchor: ☐ Required: ☐

Currency: ☐

Powered by Ephesoft

Qty	UOM Description	VND	CL	CD	Item Number	NDC/UPC	List	Unit Price	Extended Amount
1	EA AZASAN 100 MG TAB 100	B-L	RX	P	10058087 935698	6 56490-241-41	667.13	530.48	530.48

1 Total Number of Pieces for this Invoice

Product Summary

Supplies	RX Cost	OTC Cost	Retail	G.P. %
0.00	530.48	0.00	0.00	0.00

Test Table Results

S.No.	Qty	Description	UOM	VND	CL	CD	Item Number	NDC/UPC	List	Unit Price	Extended Amount
1	1	AZASAN 100 ...	EA	B-L	RX	P	10058087 935698	656490-241-41	667.13	530.48	530.48

Configuring Table with Horizontal + Vertical Columns

- Some table may have vertical columns and embeded horizontal columns as shown below –

AmerisourceBergen **INVOICE** Invoice Number: P.O. Number: 1 of 4
Invoice Date: 07/27/2015 Order Date: 07/24/2015 Ship Date: 07/26/2015

SELLER 020 **BUYER**

AMERISOURCEBERGEN DRUG CORP 1825 S. 43RD AVE. PHOENIX AZ 85009-6051

Telephone: 844-222-2273 State License: W000008 DEA: RA0290837

Ship To: [Empty Box] Acct: AZ 85085-0621

Remit To: 91189-0565 Sold To: 85085

Qty	UOM Description	VND	CL	CD	Item Number	NDC/UPC	List	Unit Price	Extended Amount
24	EA KENALOG 40 MG/ML VL 1 ML Carton Number : 00020677010102369549	BMS	RX		10006765 059634	0 0003-0293-05	10.22	8.16	195.84
12	EA FOLLISTIM AQ 900 IU CRT Carton Number : 00020677010102369563	MSD	RX	QP	10035473 545018	0 0052-0326-01	2,140.09	1,504.58	18,054.96
9	PK ENBREL SURCLK 50MG /ML KIT 4 Carton Number : 00020677010102369570	AMG	RX	Q	10003616 032918	5 84060-445-04	3,847.06	3,089.93	27,809.37
5	EA OVIDREL 250 MCG SYG 1 Carton Number : 00020677010102369594	SRO	RX	P	10057312 925221	4 40871-1500-1	150.87	119.98	599.90
5	EA GONAL-F REDIRECT 450 IU PEN Carton Number : 00020677010102369617	SRO	RX	P	10128948 253265	4 40871-1160-1	994.68	790.94	3,954.70
1	EA STELARA 45 MG SYG 0.5 ML Carton Number : 00020677010102369617	JOM	RX	P	10001734 016402	5 78940-060-03	9,846.33	7,814.69	7,814.69
12	EA CETROTIDE 0.25 MG KIT Carton Number : 00020677010102369624	SRO	RX	P	10033963 524603	4 40871-2250-1	210.56	167.44	2,009.28
3	EA COSENTYX SYG 150MG/ML 2 SENSOREADY PEN Carton Number : 00020677010102369631	NVR	RX	P	10148627 380584	0 0078-0639-41	4,387.19	3,488.54	10,465.62
3	EA COSENTYX SYG 150MG/ML 2 SENSOREADY PEN Carton Number : 00020677010102369631	NVR	RX	P	10148627 380584	0 0078-0639-41	4,387.19	3,488.54	10,465.62

This wholesale distributor, or a member of the affiliate of such wholesale distributor, purchased the product directly from the manufacturer, exclusive distributor of the manufacturer, or repackager that purchased the product directly from the manufacturer. Terms of sale and claims on reverse side.

2. To Extract Horizontal column embeded inside table vertical column no additional configuration is required. You need to draw the Column Header overlay and Column Data overlay properly to indicate the system that it is horizontal column as shown below –

Extraction Rule

Rule Name: Wrapped Header Tables

Start Pattern: Qty

End Pattern: This wholesale

Extract Repeating Tables: ☐

Overlapping Columns Table: ☐

Table Extraction API: ☐ Column Coordinates ☐ OR ☒ Column Header ☐ OR ☐ Regex Extraction

Column Configuration

Basic Configuration

Table Column: Carton Number

Column Header Pattern: Carton Number

Column Pattern: +

Pattern Left:

Pattern Right:

Start Coordinate: 206

Test Table

3. Once you are done with configuration click on Test table button to see the extracted table –

Test Table Results

S.No.	Qty	UOM	Description	EA	VND	CL	CD	Item Number	NDC/UPC	List	Unit Price	Extended Amount	Carton Number
1	24	EA	KENALOG 40 MG/ML VL 1 ML	BMS	RX			10006765 059634	0 0003-0293-05	10.22	8.16	195.84	00020877010102369549
2	12	EA	FOLLISTIM AQ 900 IU CRT	MSD	RX	QP		10035473 545018	0 0052-0326-01	2,140.09	1,504.58	18,054.96	00020877010102369563
3	9	EA	ENBREL SURCLK 50MG /ML KIT 4	AMG	RX	Q		10003616 032918	5 84060-445-04	3,847.06	3,089.93	27,809.37	00020877010102369570
4	5	EA	OVIDREL 250 MCG SYG 1	SRO	RX	P		10057312 922221	4 40871-1500-1	150.87	119.98	599.90	00020877010102369594
5	5	EA	GONAL-F REDI-JECT 450 IU PEN	SRO	RX	P		10128948 253265	4 40871-1160-1	994.68	790.94	3,954.70	00020877010102369594
6	1	EA	STELARA 45 MG SYG 0.5 ML	JOM	RX	P		10001734 016402	5 78940-060-03	9,846.33	7,814.69	7,814.69	00020877010102369594
7	12	EA	CETROTIDE 0.25 MG KIT	SRO	RX	P		10033963 524803	4 40871-2250-1	210.56	167.44	2,009.28	00020877010102369617
8	3	EA	COSENTYX SYG 150MG/ML 2 SENSOREADY PEN	NVR	RX	P		10148627 380584	0 0078-0639-41	4,387.19	3,488.54	10,465.62	00020877010102369624

Configuring Table with Overlapping Column and their data

There are certain tables in which columns are overlapping to each other and data is spanned across multiple lines are categorized as “Table with Overlapping columns” as shown in below image-

Item	New Part (Old Part) Qty	Description	\$ Price	Pricing unit	\$ Value
0010	1SXCS6465 () 3,168	Slate Moab C/T PC (37.27 SQ*)	@ 65.00 per	85 PC	2,422.59
0020	1RCCS6465C () 255	Ridge Moab C/T PC	@ 2.00 per	1 PC	510.00
0021	1002569 (1840) 13	Pallet Wood, 48" X 55" 5 Stringer EA	@ 16.00 per	1 EA	208.00
Final amount (Currency:USD)					3,140.59
Amount qualifying for cash discount		2,932.59			

To extract such table, you should enable a switch named as “**Overlapping Columns Table**” on Extraction Rule as shown below –

Extraction Rule

Rule Name:

Start Pattern:

End Pattern:

Extract Repeating Tables: ☐

Overlapping Columns Table: ☒

Table Extraction API:

☐ Column Coordinates

☒ OR

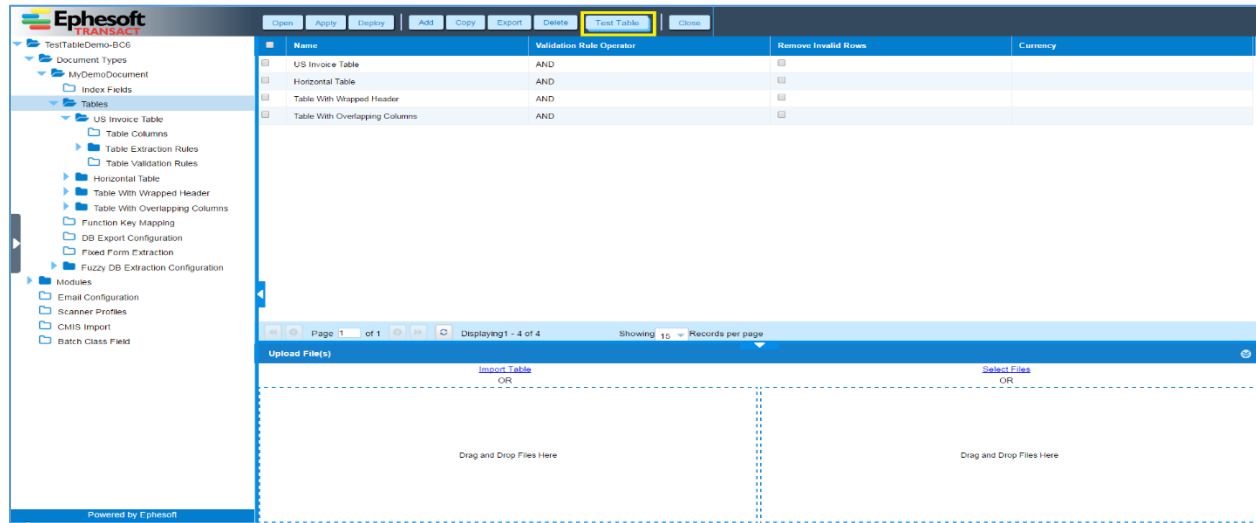
☒ Column Header

☒ AND

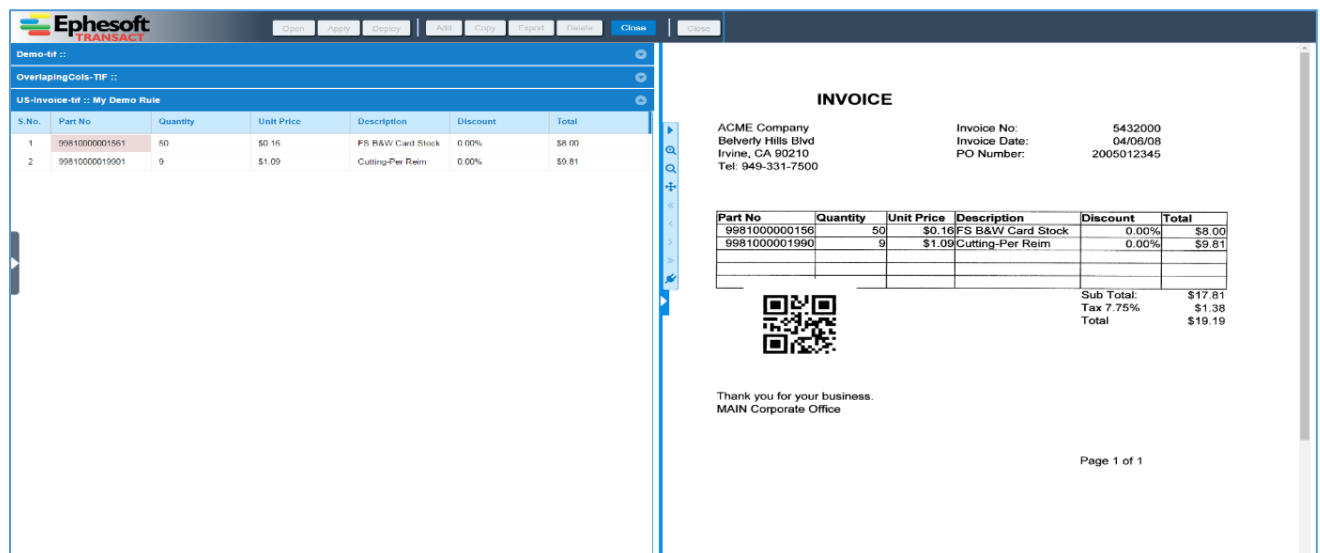
☒ Regex Extraction

Once all columns are configured user can test the result by clicking “Test Table” button –

Following screenshot displays table listing page:



Here user needs to upload images from “select file” link at the bottom for which table needs to be tested. After uploading images, user can select the table and click on “Test Table” button to see the result.



The **Test Table Extraction** screen has two sections.

- Results grid on the left
- Image view pane on the right

The results are displayed row-wise in the results grid on the left and the corresponding image is displayed in the image view pane on the right. Images are displayed in accordion layout in left hand panel where user can select a particular image from to see the extracted results.

Overlays will be displayed on image upon selecting data from the table. Apart from overlay of the selected data, start pattern, end pattern and header pattern (in case of header based extraction) are also displayed in the image to give user a feedback of configuration made.



In case of multipage documents or multiple files, by default, the first row from the results grid is selected and the image view pane displays the image corresponding to the first result.

The results grid displays tooltip messages related to low OCR confidence and validation errors for the extracted data.

The screenshot displays the Ephesoft TRANSCRIPT application. On the left, a results grid is visible under the 'US-Invoice-01 :: My Demo Rule' tab. It contains two rows of data with a tooltip for the second row indicating 'OCR Confidence 88 is Less than the threshold value 90'. On the right, a preview pane shows an 'INVOICE' document. The invoice includes company details for ACME Company, invoice number 5432000, date 04/06/08, and PO number 2005012345. It also features a table of items, a QR code, and a total amount of \$19.19.

S.No.	Part No	Quantity	Unit Price	Description	Discount	Total
1	9981000001561	50	\$0.16	US B&W Card Stock	0.00%	\$8.00
2	9981000001990	9	\$1.09	Cutting-Per Reim	0.00%	\$9.81

Part No	Quantity	Unit Price	Description	Discount	Total
9981000001561	50	\$0.16	US B&W Card Stock	0.00%	\$8.00
9981000001990	9	\$1.09	Cutting-Per Reim	0.00%	\$9.81
Sub Total:					\$17.81
Tax 7.75%					\$1.38
Total					\$19.19

Plugin Configuration

TABLE_EXTRACTION Plugin governs the extraction of data from tables.

This plugin has only one configuration, a switch named as "Table Extraction switch". By default, the value of this switch is "OFF" and in this case the table from the document will not get extracted if you run the batch. To extract table from the document you have to change the value of Table Extraction switch to "ON".

Ephesoft

TRANSACTION

Apply

Deploy

Close

Tables

US Invoice Table

Table Columns

Table Extraction Rules

Table Validation Rules

Horizontal Table

Table With Wrapped Header

Table With Overlapping Columns

Function Key Mapping

DB Export Configuration

Fixed Form Extraction

Fuzzy DB Extraction Configuration

Modules

Folder Import

Page Process

Document Assembly

Review Document

Extraction

RECOSTAR_EXTRACTION

REGULAR_REGEX_EXTRACTION

KEY_VALUE_EXTRACTION

TABLE_EXTRACTION

EXTRACTION_SCRIPTING_PLUGIN

ADVANCED_BARCODE_EXTRACTION

Automated Validation

Validate Document

Export

Email Configuration

Scanner Profiles

CMIS Import

Batch Class Field

Powered by Ephesoft

Plugin Configuration

Table Extraction switch:

OFF

ON

OFF