

Cross Section Extraction

Cross Section Extraction is a type of extraction rule used where for a value a single key pattern is not sufficient and it lies at the cross-section of two keys.

The user can use this extraction rule to extract values present at a cross section of two matching regex patterns (using simple coordinate geometry).

Cross section is identified on the basis of the following conditions:

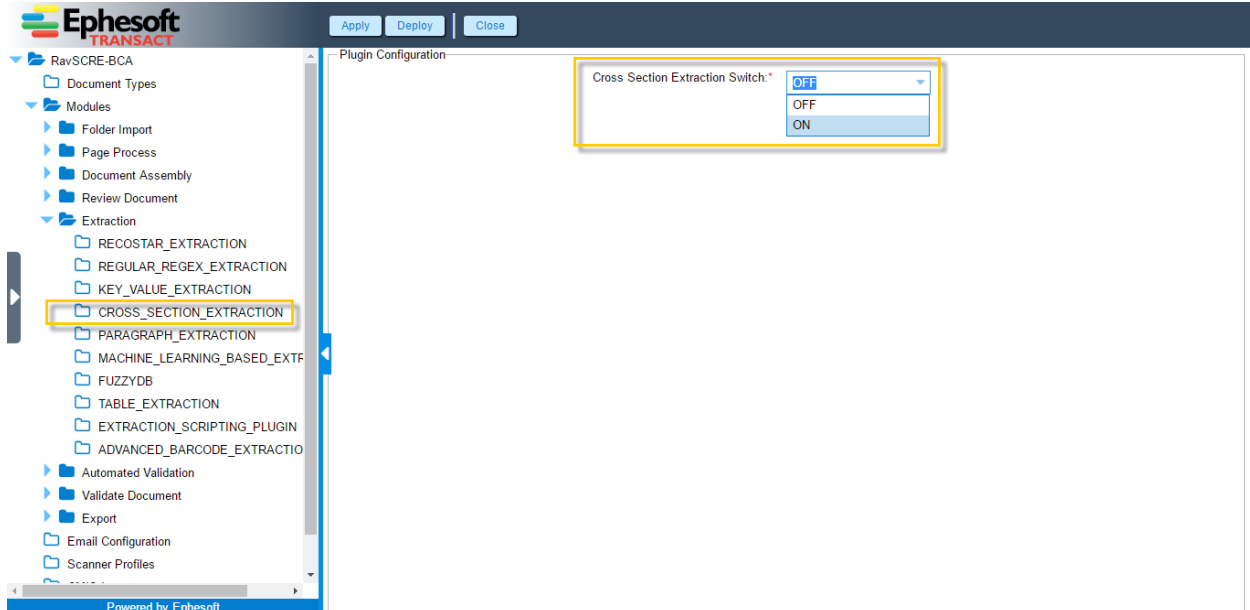
- Regex match for horizontal pattern is treated as the horizontal plane in which the value is searched. Similarly, for vertical pattern, the same set of rules is applied
- Taking the point of intersection between these two pattern matches, the value is extracted.

This functionality enables the user, as an administrator of batch classes, to configure extraction rules for index fields.

Configuration

The index field values for which **Cross Section Extraction** is configured are extracted using a plugin.

CROSS_SECTION_EXTRATION Plugin governs the extraction of configured index field while using **Cross Section Extraction**.



This plugin has only one configuration which is a switch. If the value of the switch is set to **ON**, the configured index field is extracted. By default, the switch is set to **OFF**.

Configuring Cross Section Extraction

Configuration for Cross Section Extraction is like any other extraction rule such as the KV Extraction Rule.

To configure Cross Section Extraction

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

The Ephesoft Enterprise **Login** page displays.

2. Enter valid credentials to log in.

The **Batch Class Management** screen displays.

3. Select the batch class from the list in the **Batch Class Management** screen and click **OPEN**.

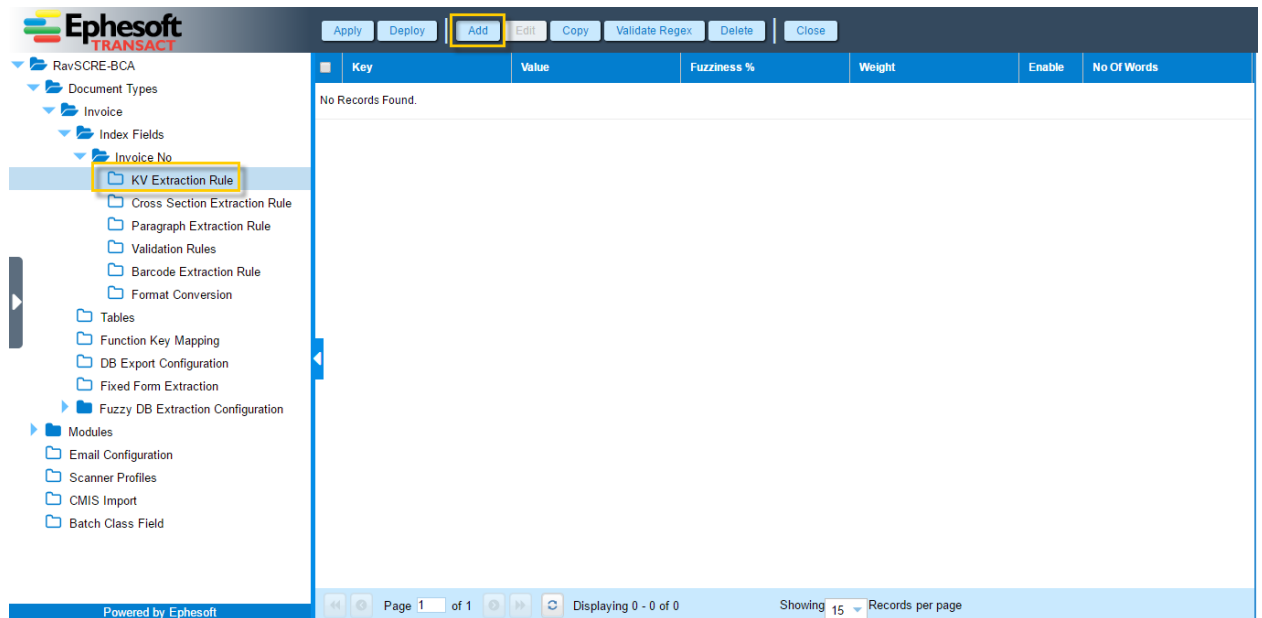
The batch class opens with **Document Types** node selected by default.

4. Select the document type from the list and click **OPEN**.

The document type node expands displaying a list of index fields.

5. Select the index field from the list and click **OPEN**.

The index field node expands displaying all the available extraction rules in the left navigation pane and **KV Extraction Rule** selected by default as shown in the image below.



6. Select **Cross Section Extraction Rule** from the navigation pane and click **ADD**.

The following screen displays.

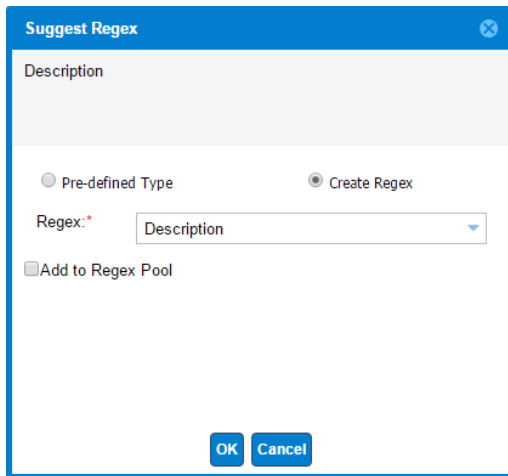
8. Drag and position the **Header** overlay over an area of the image such that horizontal plane below the overlay covers the area to be considered for extraction.



Please note that overlays in **Cross Section** extraction are just for helping the user with regex pattern creation and do not play any role in value extraction.

9. Click on the **Header** overlay in the image view pane.

The **Suggest Regex** dialog box appears with **Create Regex** option selected by default.



10. Enter a regular expression in the **Regex** text box for the **Header Pattern** or select Regex Builder from the **Regex** drop-down list to create a regular expression for the **Header Pattern** and click **OK** on the **Suggest Regex** dialog box.

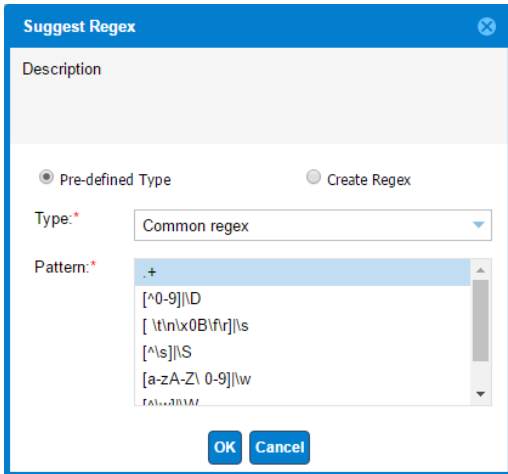


Select Add to Regex Pool if the new regex created by the user is required for future use.

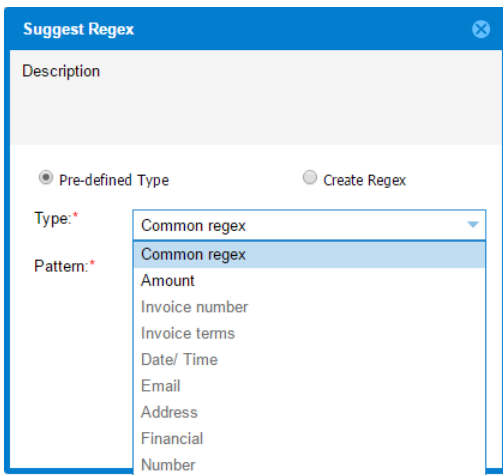
OR,

11. Select **Predefined Type** option on the **Suggest Regex** dialog box.

The **Suggest Regex** dialog box is updated.



12. Select a regex type from the Type drop-down list.



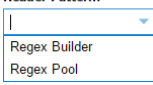
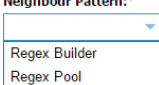
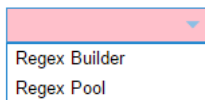
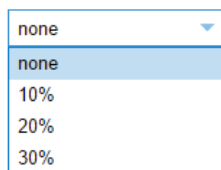
13. Select a regular expression from the Pattern drop-down list.

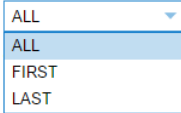
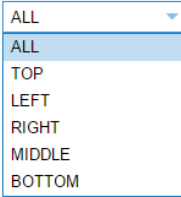
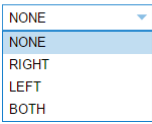
14. Click **OK** on the **Suggest Regex** dialog box.

15. Repeat the Steps 8 through 14 for the **Neighbour** and **value** overlays.

16. Enter the relevant configuration details as described in the table below:

Component	Description
Header Pattern	The User can configure a header pattern. The position of the spans that match this pattern are used. Values within the horizontal plane of this pattern are considered for extraction.

Component	Description
	<p>Header Pattern:*</p>  <p>The User can enter a regular expression.</p> <p>OR,</p> <p>Select Regex Builder to create a new regular expression to be used as the search pattern.</p> <p>OR,</p> <p>Select Regex Pool to select an existing regular expression to be used as the search pattern.</p>
Neighbor Pattern	<p>The User can configure a neighbour pattern. The position of the spans that match this pattern are used. Values within the vertical plane of this pattern are considered for extraction.</p> <p>Neighbour Pattern:*</p>  <p>The User can enter a regular expression or use Regex Builder/Regex Pool options to enter a search pattern.</p>
Value	<p>The User can use this parameter to filter out the extraction results. This parameter is used after extraction is done. All values which do not match the value pattern are discarded.</p> <p>Value:</p>  <p>The User can enter a regular expression or use Regex Builder/Regex Pool options to enter a search pattern.</p>
Fuzzy %	<p>The User can use this parameter to do a fuzzy search while searching for Regex Patterns.</p> <p>Fuzzy %:</p> 
Page	<p>The User has three options available to choose from for this parameter: ALL, FIRST, and LAST.</p>

Component	Description
	<p>Page:</p>  <p>Depending on the selected value, the extraction algorithm runs on ALL/FIRST/LAST Page of the document.</p>
Zone	<p>Every page is divided into 5 zones: TOP, MIDDLE, BOTTOM, LEFT and RIGHT along with the default option of ALL.</p> <p>Zone:</p>  <p>The User can use this parameter to specify the portion of the page where the algorithm searches for start value of paragraph to extract it.</p> <p>For example, if the user configures this parameter value as BOTTOM, the start pattern of the paragraph is searched only in the BOTTOM zone.</p>
Weight	<p>The user can use this parameter to implement weighted confidence values.</p> <p>Weight:*</p> <input data-bbox="480 1094 716 1136" type="text" value="1"/> <p>This is used to give bias/weight to a particular Extraction rule.</p>
Search Direction	<p>The User can use this option in case the cross-section point does not lie on any of the configured spans.</p> <p>By default, RIGHT is the selected Search Direction.</p> <p>The possible values for this option are NONE, RIGHT, LEFT, and BOTH.</p> <p>Search Direction:</p>  <p>When the user selects from LEFT/RIGHT, the first found span value to the Left/Right of the intersection point is extracted as the value.</p>
Is Multiword	<p>The User can use this option to enable the multiword algorithm.</p> <p>Is Multiword:</p> <input checked="" data-bbox="480 1717 509 1745" type="checkbox"/> <p>If Multiword is checked then the user can fetch all the multiple words as per algorithm of Cross Section extraction. All spans with spacing less than 1.5 times the normal space character are included in the value.</p>

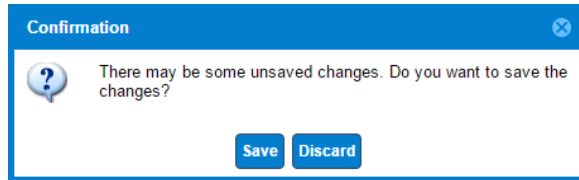
Users can test their configuration by clicking **Test Cross Section** from the toolbar on top of the page. The extraction results are displayed in the **Advance KV Test** grid as shown in the image below:



Advance KV Test							
Header Pattern	Neighbour Pattern	Value	Confidence%	Header Coordinates	Neighbour Coordinates	Value Coordinates	Colorcode
Borrower's	Our service charge	43 B	23	(1686,112)(1821,135)	(218,355)(408,384)	(1684,359)(1723,378)	

17. Click **Apply Cross Section** to apply the rule.

If the user clicks **Cancel** without saving changes, the following confirmation message displays.



Click **Save** to save changes or click **Discard** to discard any configuration changes and navigate to the **Cross Section Extraction Rule** screen.
