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**.

📒 Ep	hesoft	Apply Deploy Close			
🤝 左 RavSCI	RE-BCA	Plugin Configuration	1		
🗅 Doci	ument Types		Cross Section Extraction Switch:*	OFF 👻	
🤝 🔁 🗸	ules			OFF	
🕨 🕨 Fi	older Import			ON	
) 🖿 P	age Process				
🕨 🖿 D	ocument Assembly				
🕨 🖿 R	eview Document				
🥆 🍃 E	xtraction				
_ 🗅	RECOSTAR_EXTRACTION				
	REGULAR_REGEX_EXTRACTION				
	KEY_VALUE_EXTRACTION				
	CROSS_SECTION_EXTRACTION				
	PARAGRAPH_EXTRACTION				
C	MACHINE_LEARNING_BASED_EXTF				
	FUZZYDB	Г			
6	TABLE_EXTRACTION				
6	EXTRACTION_SCRIPTING_PLUGIN				
6	ADVANCED_BARCODE_EXTRACTIO				
🕨 🖿 A	utomated Validation				
🕨 🖿 V	alidate Document				
) 🕨 🖿 E	xport				
🗀 Ema	il Configuration				
🗀 Scar	nner Profiles				
< <u> </u>					
	Powered by Ephesoft				

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.

Ephesoft	Apply Deploy Add	Edit Copy Validate Reg	ex Delete Close			
🕆 🗁 RavSCRE-BCA	🔲 Кеу	Value	Fuzziness %	Weight	Enable	No Of Words
🕆 🝃 Document Types	No Records Found					
🕆 📂 Invoice	no no control i control.					
V Index Fields						
 Invoice No KV Extraction Rule Cross Section Extraction Rule Paragraph Extraction Rule Validation Rules Barcode Extraction Rule Format Conversion Tables Function Key Mapping DB Export Configuration Fixed Form Extraction Fixed Form Extraction Modules Email Configuration Scanner Profiles CMIS Import 	4					
Batch Class Field Pawared by Enhandler	Page 1 of 1	Displaying 0 - 0 of 0	Showing 15			

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

The following screen displays.

Ephesoft	Apply Cross Section Test Cross Section Validate Regex Back Cancel View OCR Data Image V Page No. V
Contract Con	
Fuzzy %: none Page: ALL Zone: ALL Weight:* 1	
Search Direction:	Cross Section Extraction Test 😵
Is Multiword:	Select Files OR
Powered by Enhanoff	Drag and Drop Files Here

7. Click Select Files link from Advance Cross Section Test section and upload an image file.

The uploaded image is displayed in the image view pane.

Ephesoft	Apply Cross Section Test Cross Section Validate Regex Back C	Cancel View OCR Data mage US-Invoice.tif Page No.	1 💌
BCA :: Invoice :: Invoice No Header Pattern:* Neighbour Pattern:* Value: Fuzzy %: none Page:	INVOICE ACME Company In Belverly Hills Blvd In Irvine, CA 90210 Pr Tel: 949-331-7500	voice No: 5432000 voice Date: 04/06/08 O Number: 2005012345	
ALL Zone: ALL Veight: 1 Search Direction: RIGHT I Is Multiword:	Part No Quantity Unit Price D 9981000000156 50 \$0.16 F3 9981000001990 9 \$1.09 C	Discount Total S B&W Card Stock 0.00% \$8.00 utting-Per Reim 0.00% \$9.8 Sub Total: \$17.8 Tax 7.75% \$1.36 Total \$19.15	
Powered by Enhesoft	A nank you for your business. MAIN Corporate Office		

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

1	- Altr
I	
I	_
I	

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.

Suggest Regex			8	
Description				
O Pre-defined	Туре	Oreate Regex		
Regex:*	Description		•	
Add to Regex	Pool			
OK Cancel				

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.

Suggest Rege	x	⊗		
Description				
Pre-define	d Type 🔍 Create Regex			
Type:*	Common regex			
Pattern:*	.+			
	[^0-9]\D			
	[\t\n\x0B\f\r] \s			
	[^\s] \S			
	[a-zA-Z\ 0-9] \w	_		
	TALLERINA	*		
	OK Cancel			

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

Suggest Regex						
Description						
Pre-define	d Type	Create Regex				
Type:*	Common regex		-			
Dettern:*	Common regex					
Pattern.	Amount					
	Invoice number					
	Invoice terms					
	Date/ Time					
	Email					
	Address					
	Financial					
	Number					

- 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 **Neighour** 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
	Header Pattern:*
	OR, Select Regex Pool to select an existing regular expression to be used as the search pattern.
Neighbor Pattern	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. Neighbour Pattern:* Regex Builder Regex Pool The User can enter a regular expression or use Regex Builder/Regex Pool options to enter a search pattern.
Value	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. Value: Regex Builder Regex Pool The User can enter a regular expression or use Regex Builder/Regex Pool options to enter a search pattern.
Fuzzy %	The User can use this parameter to do a fuzzy search while searching for Regex Patterns. Fuzzy %: none 10% 20% 30%
Page	The User has three options available to choose from for this parameter: ALL, FIRST, and LAST.

Component	Description
	Page: ALL ALL FIRST LAST Depending on the selected value, the extraction algorithm runs on ALL/FIRST/LAST Page of the document.
Zone	Every page in divided into 5 zones: TOP, MIDDLE, BOTTOM, LEFT and RIGHT along with the default option of ALL. Zone: ALL TOP LEFT RIGHT MIDDLE BOTTOM 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. For example, if the user configures this parameter value as BOTTOM, the start pattern of the paragraph is searched only in the BOTTOM zone.
Weight	The user can use this parameter to implement weighted confidence values. Weight:* 1 This is used to give bias/weight to a particular Extraction rule.
Search Direction	The User can use this option in case the cross-section point does not lie on any of the configured spans. By default, RIGHT is the selected Search Direction . The possible values for this option are NONE, RIGHT, LFET, and BOTH. Search Direction: NONE RIGHT LEFT BOTH 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.
Is Multiword	The User can use this option to enable the multiword algorithm. Is Multiword: 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.

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:

دياري	The extraction	he extraction results are displayed in the Advance KV Test grid as shown in the image below:							
	Advance KV Test	Advance KV Test							۲
	Header Pattern	Neighbour Pattern	Value	Confidence%	Header Coordinates	Neighbour Coordinates	Value-Coordinates	Colorcode	
	Borrower's	Our service charge	43 S	23	(1686,112)(1821,135)	(218,355)(498,384)	(1684,359)(1723,376)		

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.