

ADVANCED LINE RECOGNITION

Documentation of concept and setup

Version: ALR013
Created: 29.06.2021
Author: Sebastian Radloff

TABLE OF CONTENT

1.	About the solution	3
1.1.	Understand functionality of basic line recognition	3
1.1.1.	Requirements of basic line recognition.....	3
1.1.2.	How to capture lines with Document Capture	4
2.	Functional scope of the Advanced Line Recognition	5
2.1.	Concept of Advanced line recognition	5
2.2.	Before using the Advanced line recognition	5
2.3.	Actions on Document Card lines	6
2.3.1.	Find value by linked field	6
2.3.2.	Find value by column heading.....	8
2.3.3.	Find value by caption.....	10
2.3.4.	Reset field	11
2.3.5.	Version.....	11
2.4.	Addition Functionality.....	12
2.4.1.	Substitution field	12
2.4.2.	Copy value from previous value.....	13

1. ABOUT THE SOLUTION

The solution is extending Document Capture to recognize documents with a complex line structure that cannot be processed by Document Capture.

This documentation has the purpose to explain the differences between the basic (default) line recognition and the advanced line recognition and how to set it up. This document will not describe the way how to setup the basic line recognition for a document.

1.1. Understand functionality of basic line recognition

Document Capture is able to recognize documents with lines by default but with some limitations an user should be aware of. To get the advanced line recognition working it's important to understand the basic functionality first.

1.1.1. Requirements of basic line recognition

For Document Capture to identify line-based data in the document, the following requirements must be met:

- the values must be placed in separate columns in tabular form. Each column can then be read out via its own template fields and configured as required via rules, field translations, etc.
- Line based data are all in one line and not staggered over several lines:

If the requirements are met Document Capture can easily find all information in the document like in the following example:

INVOICE I-77777		N16 34Z London Bank 01-05-4455-8789 VAT No. 99-999-999		
Date Invoiced	14.04.2019			
Date Due	21.04.2019			
Item #	Text	Quantity	Price	Amount
IPAD	iPad 64GB	1	1,155.00	1,155.00
MOUSE	Apple Mouse	1	89.00	89.00
HEADSET	Apple Bluetooth Headset	1	199.00	199.00
AMOUNT EXCL. VAT:				1,443.00
VAT 25%:				360.75
AMOUNT INCL. VAT:				1,803.75

Figure 1 - Simple document for basic line recognition

If the requirements are not met and the information of each invoice/document line is staggered over different lines you won't be able to get a result.

The following figure shows one document line of an invoice that cannot be processed with basic line recognition:

Position	Item No.	Description	Qty./Unit	Unit Price (EUR)	Amount (EUR)
100	2101548	Order No	1114452		
		HP 312X Black Toner			
		OEM No.	CF380X		
		Your Item No.	3068050		
		COO/Tax Code	JP / 84439990		
		EAN/UPC	887111367754		
	Serial No..	(100061677249)			
		2 ST		325.45	650,90
		Order No	1114483		

Figure 2 - Documentation cannot be recognized by basic line recognition

1.1.2. How to capture lines with Document Capture

This documentation is not intended to explain the training of basic line recognition.

As a Continia partner you can use Continia learn to find a guide that describes the concept:

<https://continia.docebosaas.com/learn/course/41/play/439:392/basic-line-recognition>

Or visit the Continia youtube channel:

[Recognize lines in PDF invoices - YouTube](#)

2. FUNCTIONAL SCOPE OF THE ADVANCED LINE RECOGNITION

The Advanced line recognition adds different functionality to Document Capture that is explained in this article.

2.1. Concept of Advanced line recognition

The main difference to the basic line recognition is, that the Advanced line recognition will automatically try to find the area of the beginning and ending of an e.g. invoice line with multiple lines:

100	2101548	Order No	1114452	1	
		HP 312X Black Toner			
		OEM No.	CF380X		
		Your Item No.	3068050		
		COO/Tax Code	JP / 84439990		
		EAN/UPC	887111367754		
		Serial No..	(100061677249)		
		2	ST	325,45	650,90
200	2101898	Order No	1114483	2	
		HP 312A Cyan Toner			
		OEM No.	CF381A		
		Your Item No.	3068080		
		COO/Tax Code	JP / 84439990		
		EAN/UPC	887111367761		
		Serial No..	(100087885330)		
		1	ST	718,30	718,30
300	2101549	Order No	1114483	3	
		HP 312A Yellow Toner			
		OEM No.	CF382A		
		Your Item No.	3068090		
		COO/Tax Code	JP / 84439990		
		EAN/UPC	887111367778		
		Serial No..	(100065998523)		
		1	ST	718,30	718,30

Figure 3 - Complex invoice with the marked areas of three invoice lines

All methods, that will be described on the next pages will use/need this identified area to find the field values.

2.2. Before using the Advanced line recognition

Before you can start training line fields to use Advanced line recognition functionality you must make sure to **identify at least one field** per invoice line with the basic line recognition.

Item #	Text	Quantity/Quantity	Amount
IPAD	iPad 32GB		11,000.00
	Price	1,100.00	
	Quantity	10	
MOUSE	Apple Mouse		100.00
	Price	10.00	
	Quantity	10	

2.3. Actions on Document Card lines

The actions explained below are available in the Document Card - Lines section:

Figure 4 - Available actions of the advanced line recognition on the document card

2.3.1. Find value by linked field

This action can identify field values by an offset that is automatically calculated to a reference field that has been found before. The following screenshot shows that the quantity (10) should be identified by an offset from the item no. (e.g. IPAD)

Item #	Text	Quantity/Quantity	Amount
IPAD	iPad 32GB	1,100.00	11,000.00
	Price		
	Quantity	10	
MOUSE	Apple Mouse	10.00	100.00
	Price		
	Quantity	10	

Figure 5 - Find quantity by an offset to reference field NO

Instruction to identify the field value

This instruction should guide you to identify the field value quantity based on the offset to the reference field Item# (NO).

1. Capture Document with basic line recognition
2. Capture the field value of the Quantity in the preview window

3. Use "Find value by linked field" action to setup the field capturing

4. Select the reference field No.

i Please select first the source field on the basis of which position the value should be found.

OK

FIELD NAME
No.
Description
Quantity
Unit cost
Line amount

5. Select the field you want to identify

i Please choose the field, which should be linked with the source field "No.".

OK

FIELD NAME
Description
Quantity
Unit cost
Line amount

6. Feedback message after successful setup of the field

If you have selected every correctly you the following message should be shown to you:

i The field "Quantity" is now linked with field "No.".

OK

7. Test the training with by using the running "Recognize fields" action

The system should now show the correct field values for the field quantity:

Item #	Text	Quantity/Quantity
IPAD	IPad 32GB	1,100.00
	Price	10
	Quantity	
MOUSE	Apple Mouse	10.00
	Price	10
	Quantity	

Payment Terms: 7 days – payment due by 21.04.2010
Please pay by direct credit to our bank account 01-777-8888-9999

QUANTITY	UNIT COST	LINE AMOUNT
10	1,100.00	11,000.00
10	10.00	100.00

AMOUNT EXCL. VAT
AMOUNT INCL.

2.3.2. Find value by column heading

The purpose of this action is to identify the field value by search for a headline/caption at the beginning of the table, like the concept of the basic line recognition.

Item #	Text	Quantity/Quantity
IPAD	iPad 32GB	
	Price	1,100.00
	Quantity	10
MOUSE	Apple Mouse	
	Price	10.00
	Quantity	10

Instruction to identify the field value

1. Capture Document with basic line recognition if not done before
2. Capture the headline/caption of the column with the right mouse button
3. Capture the field value in the first invoice line with the left mouse button

Item #	Text	Quantity/Quantity
IPAD	iPad 32GB	
	Price	1,100.00
	Quantity	10
MOUSE	Apple Mouse	
	Price	10.00
	Quantity	10

UNIT COST	LINE AMOUNT
1,100,00	11,000,00
	100,00

4. Use "Find value by column heading" action to setup the field capturing
5. Select the field you want to identify

i Select the field that should be found by the column heading.

OK

FIELD NAME

No.

Description

Quantity

Unit cost

Line amount

8. Feedback message after successful setup of the field

i The field "Unit cost" will now be searched via the column heading "Quantity/Quantity".

OK

9. Test with Recognize Field action

Item #	Text	Quantity/Quantity
IPAD	IPad 32GB	1,100.00
	Price	10
	Quantity	
MOUSE	Apple Mouse	10.00
	Price	10
	Quantity	

UNIT COST	LINE AMOUNT
1,100.00	11,000.00
10.00	100.00

Payment Terms: 7 days – payment due by 21.04.2010
Please pay by direct credit to our bank account 01-777-8888-9999

AMOUNT EXCL. V.
VAT 25
AMOUNT INCL. V.

2.3.3. Find value by caption

The purpose of this action is to identify field values with a related caption/search text which follows the same logic/concept of identifying header fields.

Item #	Text	Quantity/Quantity
IPAD	iPad 32GB	
	Price	1.100.00
	Quantity	10
MOUSE	Apple Mouse	
	Price	10.00
	Quantity	10

Instruction to identify the field value

1. Capture Document with basic line recognition if not done before
2. Capture the caption "Quantity" and the related field value of the first invoice line

3. Use "Find value by caption" action to setup the field capturing
4. Select the field you want to identify

i Choose the field whose value should be found by the caption in the current line.

OK

FIELD NAME

No.

Description

Quantity

Unit cost

Line amount

5. Feedback message after successful setup of the field

i The value of field "Quantity" will now be searched via the caption in the current line.

OK

6. Test with Recognize Field action

QUANTITY	UNIT COST	LINE AMOUNT
10	1.100.00	11.000.00
10	10.00	100.00

IPAD iPad 32GB 1.100.00
Price
Quantity 10

MOUSE Apple Mouse 10.00
Price
Quantity 10

Payment Terms: 7 days – payment due by 21.04.2010
Please pay by direct credit to our bank account 01-777-8888-9999

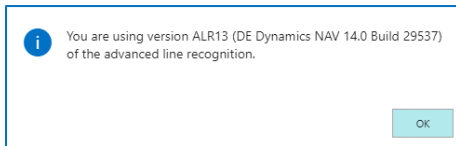
AMOUNT EXCL. V
VAT 2
AMOUNT INCL. V

2.3.4. Reset field

This action can reset the field setup of a field that has been trained to use the Advanced line recognition.

2.3.5. Version

This field is for support reason and displays the current version of the Advanced line recognition.



2.4. Addition Functionality

There are a few other options integrated in the Advanced line recognition that can help in certain situations to capture a document. These options can be configured on the template field card of the desired in the Advanced line recognition area, where you can also find the "hidden" fields that are setup while configuring a field for using the Advanced line recognition.

Manage Captions Rules Codeunits Show Attached More options

EDIT - TEMPLATE FIELD CARD - T000006 - LINE - QUANTITY - QUANTITY

General >

Advanced Line Recognition Show more

Advanced Line Recognition... Field search with caption Substitution Field

Anchor Field

Field position Below standard line

Copy value from previous... ☒

Field Translations Manage

2.4.1. Substitution field

The idea of the substitution field is, that the system will use the field value of another field of the same invoice line if the system couldn't identify a field value.

In the following example you would normally try to capture the item no. (NO) via the caption "Your Item No.". Unfortunately this field doesn't exist in the invoice line with position no. 500. Here you could setup the Vendor **Item No.** (column "Item No." in the screenshot) as a substitute field for the field NO.

Position	Item No.	Description	Qty./Unit	Unit Price (EUR)	Amount (EUR)
400	2101550	HP 312A Magenta Toner			
		OEM No.	CF383A		
		Your Item No.	3068110		
		COO/Tax Code	JP / 84439990		
		EAN/UPC	887111367785		
	Serial No..	(100063396392)			
			1 ST	718,30	718,30
500	7000061	Freight	1 ST	40,00	40,00

Payment Terms: 14 days net

The result would be that the 7000061 would be used as "Item no." in the last invoice line.

2.4.2. Copy value from previous value

The purpose of this field is to inherit the field value of the current field from the previous invoice line, if it has been found in this there.

This is often useful for collective invoices where the vendor is printing the reference order no. only once in the beginning for certain invoice lines, like in the following example, where the third invoice line (Position 300) doesn't have the order no. information:

ORDER NO.	QUANTITY	Position	Item No.	Description	Qty./Unit	Order No.
1114452						1114452
1114483		100	2101548	HP 312X Black Toner	2 ST	1114483
				OEM No. CF380X		
				Your Item No. 3068050		
				COO/Tax Code JP / 84439990		
				EAN/UPC 887111367754		
				Serial No. (100061677249)		
		200	2101898	HP 312A Cyan Toner	1 ST	
				OEM No. CF381A		
				Your Item No. 3068080		
				COO/Tax Code JP / 84439990		
				EAN/UPC 887111367781		
				Serial No. (100087885330)		
		300	2101549	HP 312A Yellow Toner		
				OEM No. CF382A		
				Your Item No. 3068090		
				COO/Tax Code JP / 84439990		
				EAN/UPC 887111367778		
				Serial No. (100065998523)		

If you setup this field to copy the value from the previous line it will inherit it's value to all following lines until a new order no. could be found.

DESCRIPTION	ORDER NO.
P 312X Black Toner	1114452
P 312A Cyan Toner	1114483
P 312A Yellow Toner	1114483
P 312A Magenta	1114483
eight	1114483