Continia Software GmbH
Mittelweg 144
D-20148 Hamburg
Germany

Tel +49 40 696 324 180
sales-dach@continia.com
www.continia.com

ADVANCED LINE RECOGNITION

Documentation of concept and setup

Version: ALR013
Created: 29.06.2021
Author: Sebastian Radloff

TABLE OF CONTENT

1. ABOUT THE SOLUTION

The solution is extending Document Capture to recognize documents with a complex line structure that cannot been 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:

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:

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:

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.

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)

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

Continia

4. Select the reference field No.

5. Select the field you want to identify

6. Feedback message after successful setup of the field
   If you have selected every correctly you the following message should be shown to you:

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:

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.

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

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

8. Feedback message after successful setup of the field

9. Test with Recognize Field action

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.

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

5. Feedback message after successful setup of the field

6. Test with Recognize Field action
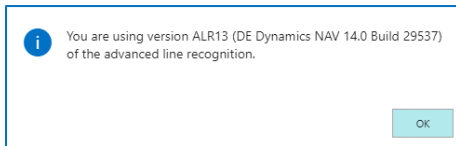
2.3.4. Reset field

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.

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.

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:

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.